AgriSen-COG, a Multicountry, Multitemporal Large-Scale Sentinel-2 Benchmark Dataset for Crop Mapping Using Deep Learning

Teodora Selea

Faculty of Mathematics and Informatics, West University of Timisoara, 300223 Timisoara, Romania;
teodora.selea@e-uvt.ro

Abstract: With the increasing volume of collected Earth observation (EO) data, artificial intelligence (AI) methods have become state-of-the-art in processing and analyzing them. However, there is still a lack of high-quality, large-scale EO datasets for training robust networks. This paper presents AgriSen-COG, a large-scale benchmark dataset for crop type mapping based on Sentinel-2 data. AgriSen-COG deals with the challenges of remote sensing (RS) datasets. First, it includes data from five different European countries (Austria, Belgium, Spain, Denmark, and the Netherlands), targeting the problem of domain adaptation. Second, it is multitemporal and multiyear (2019–2020), therefore enabling analysis based on the growth of crops in time and yearly variability. Third, AgriSen-COG includes an anomaly detection preprocessing step, which reduces the amount of mislabeled information. AgriSen-COG comprises 6,972,485 parcels, making it the most extensive available dataset for crop type mapping. It includes two types of data: pixel-level data and parcel aggregated information. By carrying this out, we target two computer vision (CV) problems: semantic segmentation and classification. To establish the validity of the proposed dataset, we conducted several experiments using state-of-the-art deep-learning models for temporal semantic segmentation with pixel-level data (U-Net and ConvStar networks) and time-series classification with parcel aggregated information (LSTM, Transformer, TempCNN networks). The most popular models (U-Net and LSTM) achieve the best performance in the Belgium region, with a weighted F1 score of 0.956 (U-Net) and 0.918 (LSTM).The proposed data are distributed as a cloud-optimized GeoTIFF (COG), together with a SpatioTemporal Asset Catalog (STAC), which makes AgriSen-COG a findable, accessible, interoperable, and reusable (FAIR) dataset.

Keywords: benchmark dataset; crop monitoring; crop detection; deep learning; image segmentation; multitemporal analysis; crop classification; agricultural application; Sentinel-2; common agricultural policy (CAP)

1. Introduction

Artificial intelligence (AI) has become a hot topic in the past decade, and since 2015, it has been an important method used in the Earth observation (EO) community [1]. The adoption of AI techniques—machine learning (ML) and deep learning (DL)—for EO-related use cases is a consequence of the increasing volume of publicly available satellite data (e.g., Copernicus Sentinels), and thus, a need to process it. Possible use cases are land cover and land use, deforestation, urban mapping, and agriculture. In this paper, we focus on the latter, particularly on the task of crop type mapping.

Crop type mapping supports the crop monitoring task, which is helpful for agricultural insurance or implementing common agricultural policies (CAP). Applying crop monitoring at a global scale can be conducted with the help of satellite data, which offers global coverage. Monitoring crops is a crucial factor for further agricultural development, as we need a production increase of 60% to serve the demand of current population growth

(Food and Agriculture Organization (FAO) of the United Nations study [2]). However, the increase in yield needs to be sustainable without damaging natural resources or the environment (Climate Sustainable Development Goals (SGD) [3]). To properly put these two objectives together, global and joint monitoring must be established.

In the context of AI methods, the crop type mapping use case is assimilated with the semantic segmentation task from computer vision (CV), described as attaching a class label to every pixel from the image. It can also be incorporated into a classification problem by aggregating parcel data and offering one label for each parcel. However, as opposed to standard 3-channel (red, green, blue) images, satellite data enrich the input by including multitemporal and multichannel information. For the selected use case, the temporal data are beneficial for identifying crop phenology, and therefore, for better classifying it. Furthermore, the multichannel satellite data refine the crop information, as the various spectra capture different characteristics.

The advancement of AI methods is correlated to the training datasets' availability, size, and quality. While other domains (e.g., computer vision) already have several benchmarking datasets (e.g., ImageNet [4], CityScape [5]), there is still a lack of large-scale, high-quality EO datasets. Even though large volumes of satellite data are publicly available, the challenge comes from creating ground truth (GT) data. For crop type mapping, GT information for dataset creation may be extracted from land parcel identification systems (LPIS). Several European countries have made their LPIS information public in the past few years, and therefore, there has been an increase in crop-related datasets for the European area. However, since the LPIS is created individually by each country with data received from farmers, it raises the following problems: no crop naming conventions, different languages used, and errors in collecting the crop type for each parcel. Existing datasets include information from only one or two EU countries without a standard way of storing and distributing the datasets. Datasets that combine knowledge from several areas would improve the state-of-the-art methods related to domain adaptation. A unified way of accessing the dataset leads to ease of usage and faster integration with DL techniques. In addition, there is also a lack of a methodology for LPIS processing so that the existing training crop-type datasets may be easily extended once new data become available. Moreover, as LPIS is created based on farmer data, it is prone to human error, which may lead to incorrect labels affecting the performance of the ML/DL model; therefore, the data need to be curated.

In this paper, we propose AgriSen-COG, a new large-scale crop-type mapping dataset with the following characteristics: (1) it is based on publicly available data only (Sentinel-2 and LPIS), making it easily extensible; (2) it includes data from five different European countries (Austria, Belgium, Spain, Denmark, and the Netherlands), targeting the problem of domain adaptation; (3) it incorporates an anomaly detection based on autoencoder as preprocessing step that lowers the amount of mislabeled information for GT; (4) it is multitemporal and multiannual, incorporating crop phenology and seasonal variability; (5) it includes pixel-based data to account for the crop parcel's spatiality; (6) it incorporates time-series data for crop classification (when the parcel geometry is known). AgriSen-COG is distributed using accessible formats such as COGs, Zarr, and Parquet and is indexed under SpatioTemporal Assets Catalogues (STAC). By using COGs, we ensure an easy way of accessing the data without the need to download the entire dataset. STAC enables a standard way to discover and describe our dataset, making it reusable and interoperable. Our main contributions are as follows:

1. We created AgriSen-COG, a large-scale benchmark dataset for crop type mapping, including the largest number of different European countries (five), designed for AI applications.
2. We introduce a methodology for LPIS processing to obtain GT for a crop-type dataset, useful for further extensions of the current dataset.

3. We incorporate an anomaly detection method based on autoencoders and dynamic time warping (DTW) distance as a preprocessing step to identify mislabeled data from LPIS.
4. We experiment with popular DL models and provide a baseline, showing the generalization capabilities of the models on the proposed dataset across space (multicountry) and time (multitemporal).
5. We provide the LPIS preprocessing, anomaly detection, training/testing code, and our trained models to enable further development.

The remainder of this article starts with a review of other datasets for crop type mapping. Next, we present popular ML/DL techniques for our selected use case. Furthermore, we continue with the proposed methodology for dataset creation and describe how we applied anomaly detection to curate the LPIS information. Finally, we present our experimental results, followed by a discussion, and summarize our conclusions.

2. Crop Datasets for ML/DL Applications

Benchmark datasets play a significant role in developing ML/DL methods. New methods need to be assessed using the same input dataset to eliminate the bias given by learning from different data. The progress in computer vision deep learning methods is partially due to numerous large benchmarking datasets (e.g., ImageNet [4], MS COCO [6], Cityscape [5]) that offer access to many annotated samples. As the interest in ML/DL applied to remote sensing data is increasing, so is the demand for large-scale datasets processed to fit various use cases (e.g., land cover, flooding, building detection, agriculture). Figure 1 presents the publication date of popular RS datasets, showing a recently increased interest in large-scale crop type mapping datasets. Each dataset is characterized by the following: (1) the area covered; (2) the time extent for the collected data; (3) the source for both input and GT data; (4) the intended use case; (5) the number and size of provided patches (the original image size might be too large to fit into hardware memory; therefore, the images are cropped in smaller, ready-to-use patches).
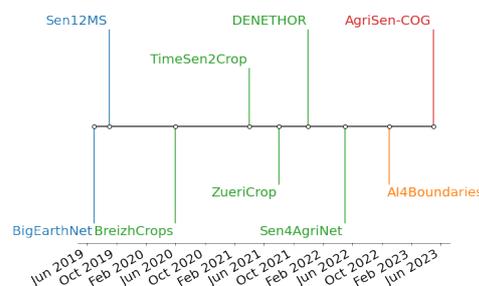
Figure 1. Crop-related datasets' publication timeline.

The first attempts to create large-scale datasets for RS scenarios targeted the land cover problem. BigEarthNet [7] represents one of the first large-scale RS image archives. It targets land-cover multi-classification and uses Sentinel-2 data as input, matched with Corine Land Cover (CLC) information for ground truth. It includes data from June 2017 to May 2018, covering ten different European regions: Austria, Belgium, Finland, Ireland, Kosovo, Lithuania, Luxembourg, Portugal, Serbia, and Switzerland. Since it is a dataset designed to work with computer vision ML/DL models, the authors distribute the dataset in patches, resulting in a total of 590,326 nonoverlapping patches, with different sizes for each resolution: 120×120 pixels for 10m band, 60×60 pixels for 20 m, and 20×20 for 60 m bands.

Also targeting land cover, but from a semantic segmentation perspective, is the Sen12MS [8] dataset. Like BigEarthNet, it uses Sentinel-2 data, but adds Sentinel-1 and extracts its ground truth information from MODIS land cover maps. Sen12MS is distributed in overlapping patches of 256×256, with a stride of 128. The dataset includes 180,662 patch triplets (Sentinel-1, Sentinel-2, GT) covering a global area and including all meteorological

seasons. Sen12MS distributes its patches at a resolution of 10m, with upsampled data for the 60 m and 20 m resolution bands.

As large-scale remote sensing land cover datasets appeared, the need to target more specific use cases also emerged. This paper focuses on the particular scenario of crop-type mapping. Crop type mapping datasets are based on land parcel identification system [9] (LPIS) information to generate ground truth data. Therefore, datasets are developed together with the release of open access to regions' parcel information. Compared to a land cover dataset, a crop type dataset includes only one category of labels—agricultural fields—but with increased granularity. Therefore, each crop parcel is delimited and labeled with the corresponding crop type.

BreizhCrops [10] is the first large-scale dataset for crop classification, covering the Brittany area of France (27,200 km^2), spanning over the entire years of 2017 and 2018. The proposed dataset provides nine broad categories of crops and uses both Sentinel-2 Level-1C and Level-2A data to gain better regional coverage. The GT is created from the France's publicly available LPIS database (Registre Parcellaire Graphique—RPG). The authors provide mean-aggregated values per band/timeframe over each field parcel, not image patches at a pixel level.

ZueriCrop [11] is a dataset for crop type mapping based on Sentinel-2 Level-2A bottom-of-atmosphere images from the Switzerland area (50 km $\times$ 48 km area). The GT data were extracted from Switzerland's LPIS (Swiss Federal Office for Agriculture), which is not publicly available. The dataset comprises 48 different crop classes at the lowest hierarchical level and 5 categories at most at the top level. The crops are observed for 2019, resulting in 28,000 patches of 24 $\times$ 24 pixels, including 116,000 crop parcels.

The Austrian region for crop type mapping is covered in the dataset proposed by [12] (TimeSen2Crop). The dataset uses Sentinel-2 bottom-of-atmosphere images spanning between September 2017 and August 2018. For GT, the authors used the publicly available Austrian LPIS, offering 16 different labels in the end. TimeSen2Crop is distributed as a monthly median composite for each tile, and comprises around 1 million labeled samples in total.

DENETHOR [13] is the first dataset for crop type mapping that includes commercial data, namely Planet. It consists of three types of input data: Planet (3 m resolution), Sentinel-1, and Sentinel-2. DENETHOR covers the area of Northern Germany over two years: 2018 and 2019. Like the datasets mentioned above, DENETHOR uses the LPIS from Germany; however, it is not a publicly available database. The dataset includes nine crop type labels, with 4500 crop fields.

Sen4AgriNet [14] is also a crop classification dataset based on Sentinel-2 Level-1C and LPIS data. The dataset uses the FAO ICC [15] classification for aggregating the LPIS information from France and Catalonia. It is the first multicountry and multiyear dataset, comprising two regions: France and Catalonia. Sen4AgriNet comprises aggregated data (crop parcel average) and pixel-based data (image patches), making it a valuable dataset for crop classification and crop segmentation. Sen4AgriNet offers the full spectrum of Sentinel-2 bands, preserving the initial spatial resolution. The data are divided into smaller patches (366 $\times$ 366 pixels for 10m resolution, 183 $\times$ 183 pixels for 20 m resolution, and 61 $\times$ 61 pixels for 60m resolution). The temporal extent includes the years 2019 and 2020, resulting in patches with 168 class labels. Sen4AgriNet is distributed using the NetCDF format, making it compatible with modern self-describing tools such as Xarray [16,17].

AI4Boundaries [18] is the first multicountry crop dataset intended for field boundary detection, including 14,8 million parcels. It covers 2019 and incorporates seven different regions: Austria, Catalonia, France, Luxembourg, the Netherlands, Slovenia, and Sweden. AI4Boundaries offers two complementary datasets. First, it uses Sentinel-2 cloud-free monthly composites, with tiles 256 pixels in size, including the four 10 m resolution bands. Second, AI4Boundaries provides three channels at 1m resolution orthophotos of 512 $\times$ 512 pixels.

We described several crop-type mapping datasets that include the European region, as they provide particularities in creating the ground truth data from LPIS databases. Most previously mentioned datasets use the available LPIS information for the selected country. However, they lack a methodology for the steps needed to process the data to obtain high-quality GT. Extending current datasets is tedious and prone to error tasks without a standard way of processing. Furthermore, although several regions are covered (Austria, Northern Germany, France, Catalonia, Switzerland), there is no standard in class naming and grouping from the initial LPIS. Besides Sen4AgriNet (covering France and Catalonia), each dataset only covers one type of data: pixel or parcel aggregate, making it difficult to benchmark other results. Additionally, there is no mention of possible mislabeled parcels in the original LPIS, which may cause training errors later.

AgriSen-COG is designed to extend existing datasets (Table 1), complementing both temporal and spatial perspectives. Like Sen4AgriNet, we use the FAO ICC crop naming conventions to provide a standard and extensible way of labeling the dataset. AgriSen-COG includes pixel-level and object-aggregated data for two years (2019 and 2020), enriching BreizhCrops, TimeSen2Crop, and DENETHOR. Compared to Sen4AgriNet, for the Catalonia region, our proposed dataset provides additional granularity in crop type label selection paired with additional anomalous label removal; it is also based on Sentinel-2 Level-2A data, as opposed to Level-1C as in Sen4AgriNet, including more patches from the Catalonia area (5168 patches vs. 4638 patches), and incorporates a different parcel aggregation method (barycenter vs mean and standard deviation). AgriSen-COG is the first dataset to include information from five EU countries. It is also easily integrated with AI4Boundaries for parcel boundary refinement or for discovering and creating new crop type datasets.

Table 1. Summary of the main characteristics of popular large-scale datasets for land cover and crop type mapping.

Dataset	Nr. of Samples	Sample Size	Data Source	Short Summary
BigEarthNet	590,326 patches	up to 120 × 120	Sentinel-2 L2A, CLC	land cover multiclass classification; nonoverlapping patches; 10 EU regions; June 2017–May 2018
SEN12MS	541,986 patches	256 × 256	Sentinel-1, Sentinel-2, MODIS Land Cover	land cover semantic segmentation; patches overlap with a stride of 128; global coverage; December 2016–November 2017
BreizhCrops	768,000 fields	mean for each parcel band/timeframe	Sentinel-2, LPIS France	crop classification; covers Brittany region (France); January 2017–December 2018
TimeSen2Crop	1,200,000 fields	monthly medians	Sentinel-2 L2A, LPIS Austria	crop type mapping; covers Austria; September 2017–August 2018
ZueriCrop	28,000 patches	24 × 24	Sentinel-2 L2A, LPIS Switzerland	crop type mapping; covers Swiss Cantons of Zurich and Thurgau ; January 2019–December 2019
DENETHOR	4500 fields	2400 × 2400	Planet, Sentinel-1, Sentinel-2, LPIS Germany	pixel level; crop type classification; covers Northern Germany; January 2018–December 2019
Sen4AgriNet	5000 patches	up to 366 × 366	Sentinel-2 L1C, LPIS France and Catalonia	crop type mapping; covers Catalonia and France for two years (2019 and 2020); ICC labels; pixel and parcel aggregated time series (mean and standard deviation)
AI4Boundaries	7831 fields	256 × 256 (S2), 512 × 512 (aerial)	Sentinel-2 L1C, aerial orthophoto at 1 m	crop field boundaries; covers seven regions (Austria, Catalonia, France, Luxembourg, the Netherlands, Slovenia, and Sweden); monthly composite March 2019–August 2019
AgriSen-COG	6,972,485 fields 41,100 patches	366 × 366	Sentinel-2 L2A, LPIS of 5 EU countries	crop type mapping; crop classification; parcel-based time series aggregated using barycenters; covers five regions (Austria, Belgium, Catalonia, Denmark, the Netherlands); 2019 and 2020

3. Popular DL Methods for Crop Type Mapping

As the crop type mapping problem is frequently assimilated with the semantic segmentation task from computer vision (CV), the first DL approaches were heavily based on CV models based on classic convolutional neural networks (CNNs). However, satellite data pose additional characteristics to the three-channel (red, green, blue—RGB) CV image. These distinct properties include larger image size, increased number of channels, and the temporal dimension of the data. For the specific task of crop type mapping, the near-infrared band proves particularly useful in combination with the RGB channels, as it captures vegetation characteristics. The temporal dimension is essential for classifying crops because it provides information about their growing cycles. The larger size issue is solved by tiling the initial image (e.g., 10,980 × 10,980 pixels) into smaller patches (e.g., 366 × 366 pixels) that fit the network and hardware limitations.

Regarding the previously mentioned factors, crop type mapping is implemented using two approaches: a semantic segmentation problem or a time-series classification task. The first case uses DL networks composed entirely of CNNs or in combination with recurrent neural networks (RNNs) [19,20]. The second approach benefits from simple RNNs or transformer [21] networks. However, a plain CNN network can also be used for time-series classification.

U-Net [22] is a popular CNN topology with good results on semantic segmentation. The main characteristic of U-Net is the "U"-shaped architecture, where information from intermediate layers of the encoder part of the network is transferred to the decoder. Even though it was built for biomedical images, U-Net has been successfully applied on EO-related use cases such as land cover ([23–25]), cloud masking ([26–28]), building detection ([29–31]), crop type mapping ([32–34]), and others.

ConvLSTM [35] uses a combination of CNN layers with a particular type of RNN, the LSTM cell. The proposed network was designed for spatiotemporal inputs, in particular for the use case of precipitation nowcasting. ConvLSTM enhances the simple LSTM [36] cell by applying a convolutional layer over the input data. Therefore, the network exploits the data's temporal (LSTM) and spatial (convolution) dimensions. This topology is also popular among remote sensing data, in particular for a use case that benefits from both spatial and temporal dimensions such as land cover ([37–39]), soil moisture ([40,41]), solar radiation ([42]), air quality [43], and others, including crop type mapping ([44–46]). In [47], the authors propose a more efficient version of ConvLSTM, named **ConvSTAR**. It eliminates several operations inside the LSTM cell (input and output gates), which results in a faster and more stable training process.

Crop type mapping may be viewed as a time-series classification problem if the spatial dimension is discarded and information is aggregated at the parcel level, usually with a mean over a parcel's pixels. The result is a sequence for each crop parcel, with the length equal to the number of sensing times for each polygon. Therefore, simple RNNs such as LSTMs have been used, for example, with datasets such as BreizhCrop for crop classification.

The Transformer model [21] is another popular way to deal with sequence data. It uses the attention mechanism, which enables the network to access any particular step. Transformers are the new state of the art in natural language processing, as they require fewer parameters than LSTMs, resulting in faster training and more accurate results.

TempCNN [48] is a CNN network capable of handling sequence data. As opposed to the LSTM and transformer, it was designed specifically for crop classification in France. It uses one-dimensional convolutional layers applied to satellite image time series (SITS).

To prove the validity of the proposed dataset, we experiment with the two methods of crop detection: similar to semantic segmentation and time-series classification. Therefore, we present how AgriSen-COG is a promising dataset for new developments, regardless of the approach. To carry this out, we use popular deep-learning models from each category. We aim to provide a benchmark and starting point for each technique to help them further progress.

4. Anomaly Detection

Dataset quality is strongly related to the performance of an ML/DL algorithm and directly influences a model's ability to predict the desired result accurately. The garbage in, garbage out principle particularly applies when learning from data. Bad-quality data affect the training process, leading to longer training time and poor performance. However, when trained on well-labeled data, the algorithms are faster and better at discovering the patterns in the data.

As previously seen, the LPIS requires information directly from the farmers. Therefore, it might include errors affecting the quality of the generated GT. To assess the quality of the labels in the proposed dataset, we treat it as a problem of anomaly detection applied to time series. Since we only have a set of labels and we aim to identify outliers, we apply an unsupervised anomaly detection method.

Dynamic time warping (DTW) is a popular distance used for measuring similarities between time series. As opposed to the Euclidian distance, it enables the comparison of shifted time series with heterogenous lengths. In the case of crop phenology, it is essential to be able to identify similar growing cycles, even if they are shifted. DTW has already been applied on several use cases that aimed to find outliers or to align a time series: magnetic data [49], electric grid [50], livestock activity [51].

DTW is also implemented in combination with ML techniques, such as k-nearest neighbors (KNN) [52] or the DL method, with autoencoders [53]. In [54], the authors propose a way to compute a barycenter, a cluster center that minimizes the DTW distance.

DTW has already been used to identify crop similarity, mostly for the crop classification task. In [55], the authors used the DTW distance to identify outliers in the LPIS data for the area of Italy, converting the problem of crop classification to an anomaly detection task. The proposed method used the normalized difference vegetation index (NDVI) [56] and computes a reference feature trend from the histogram of the available parcels. Afterward, several thresholds were calculated based on the true and false positive rates. Anomaly detection on crop sites was also performed by [57], where the authors used a histogram to identify the anomalous pixels. In [58], the authors also integrated an outlier detection step during their dataset preparation stage to target the use case of grassland mowing detection over Estonia. For each grassland field, the authors computed the NDVI using Sentinel-2 data and applied an anomaly detection step to distinguish the outliers introduced by clouds. The outliers were identified by considering triplets of consecutive NDVI measurements and identifying the triplet that forms a rhombus shape. Moreover, the authors proposed to exclude several prediction results with low confidence from the employed deep learning network.

The idea of anomaly detection in agriculture was also explored in [59]. The authors performed outlier detection on IoT data using deep learning methods. The DTW distance is used as a measure of performance. DTW distance was also used in [60] to detect anomalies in satellite sensor time series data acquisition. The authors argue that a data-driven method has the advantage of not requiring expert knowledge, as in the case of establishing specific thresholds. Another plus would also be against model-based approaches, which require an accurate mathematical model that is time-consuming to achieve. However, the data-driven method is based on the similarity between the time series. The authors combined the DTW distance with KNN algorithm to better identify the outliers. The DTW was also applied by [61], where it was used to detect outliers in plane traffic control. However, the authors use the DTW algorithm to obtain the best sequence alignment, and afterward to apply a Euclidian distance to measure the distance between two instances.

Autoencoders were applied in [62] for the specific task of identifying mislabels on LPIS data over the Sevilla (Spain) region. The authors proposed a method using Sentinel-1 as input data and an autoencoder with 1D convolutional layers in its encoder part. The proposed work analysed the LPIS labels from two perspectives: at the parcel level and at the class level. Therefore, the authors identified possible mislabeled pixels from one parcel, such as entire polygons, which might have received the wrong crop label. In the paper,

there was no aggregation performed at the parcel level, but the convolutional layers were used to extract the temporal feature. In order to separate anomalies, the dynamic Otsu [63] thresholding was applied.

For our proposed dataset, we apply an anomaly detection method using an LSTM-based autoencoder as we work with variable-length sequences. Furthermore, as our dataset is formed from six regions, we need to reduce the dimensionality of our data. Therefore, we apply it at the class level, using our Sentinel-2 data, aiming to identify only mislabeled parcels.

5. The AgriSen-COG Dataset

The area reflected in the proposed dataset covers regions from five EU countries (Austria, Belgium, Spain, Denmark, and Netherlands). AgriSen-COG comprises 6,972,485 parcel observations, grouped in 41,100 patches of size 366×366 pixels. Each observation is described by the parcel's spatial and temporal characteristics or as a univariate time series, with an aggregated version for each polygon.

5.1. Input Satellite Data

The dataset is based on ESA Sentinel-2 [64] data, as they have the largest spectral and spatial resolution out of all free optical satellite optical data, which helps for a more precise segmentation. The Sentinel-2 mission is a constellation of two satellites in the same orbit, with a phase of 180° from each other. It has a high revisit time (5 days) and offers 13 spectral bands at 3 spatial resolutions: 10 m, 20 m, and 60 m. The Sentinel-2 product is available to users under two processing levels: Level 1C—top of atmosphere, and Level 2A—bottom of atmoshepere. The latter is derived from Level 1C; as the name suggests, it contains an atmospheric corrected image. In addition, Level 2A is delivered with cloud probability data and a scene classification mask (SCL), incorporating land cover, cloud, and snow labels. A Sentinel-2 orthoimage (tile) corresponds to an area of 100×100 km^2, distributed in UTM/WGS84 projection.

5.2. LPIS—Crop Type Labels

Land parcel identification systems [9] (LPIS) are designed to record all of the EU's crop parcel information. Common agricultural policy (CAP) uses them to verify agricultural subsidies and environmental obligations. LPISs are an essential component of IACS, the primary system CAP uses for handling subsidiaries. The system serves several objectives: validating the parcel identification information, assessing the eligibility area, and aiding in on-the-spot controls conducted for administrative purposes or by CAP. LPISs are databases created using information from the farmers and are under the governmental administration of each country. Therefore, no naming convention makes combining LPIS datasets from different countries difficult. Several EU administrations have published their LPIS information in the past few years, accelerating the creation of crop-type-mapping-related datasets.

5.3. Dataset Creation Methodology

Crop type mapping datasets are built upon publicly available data. Several countries have already published their LPIS information for the EU region, making it possible to create large-scale crop type mapping datasets for ML/DL. In this paper, we analyze the available parcel information from all open-source EU crops' data, resulting in the collection, processing, and analysis of five different areas (Austria, Belgium, Catalonia, Denmark, and Netherlands). Since there is no standard regarding LPIS data, creating a dataset raises several challenges due to the nonuniformity of the data. We also propose a detailed and reproducible methodology based on the LPIS input's heterogeneity, which may be applied further to extend the current dataset with additional areas. Multiple regions included in an ML/DL dataset for crop type mapping can increase a model's generalization capability or help in the domain adaptation method. In this context, a similar processing method is crucial for providing consistency among the data. Our dataset creation process

is divided into three stages: (1) LPIS processing, to obtain a standard crop description for each polygon; (2) preparing the rasterized data, to obtain pairs of input data; and GT (3) improving dataset quality using anomaly detection. In Figures 2 and 3, we present the workflow of our methodology and the challenges they handle. The final dataset and several intermediate preprocessing steps are available for download from our repository. The entire processing code may be accessed as Python scripts on the project's GitHub, and serves as a quickstart in modifying or extending the proposed dataset.
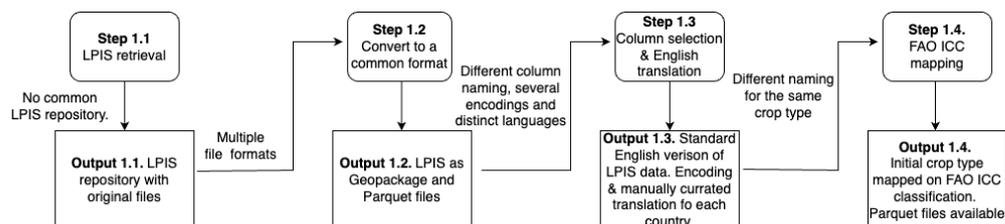
Figure 2. Overview of the LPIS processing workflow.

Figure 3. Overview of the rasterization workflow.

(1) LPIS processing

Step 1.1 is represented by LPIS data collection. In the absence of a shared database, retrieving each piece of LPIS information is time-consuming and performed manually. Most of the LPIS datasets are available on the Ministry of Agriculture of each country's webpage. However, navigation is hampered, as most websites use specific acronyms in the original language to denote the files. We collected and published the original files involved in the proposed dataset for easier access (Output 1.1). The original files helped to test the proposed workflow or conduct different analyses/processing.

Step 1.2 consists of converting the LPIS data to a unified format, which helps with further processing. The lack of a standard is visible first in the output format used to deliver the LPIS data. We encountered Shapefiles [65], Geopackages [66], Geodatabases [67], and GeoJSON [68] files. The output format consistency does not apply even for the same area, as there is a distinct format for different years. Therefore, we unified the file types, providing the data in two formats (Output 1.2): Geopackage and Parquet. Geopackage was chosen due to its popularity among the geocommunity and the integration with geotools (e.g., QGIS [69]). We also provided the data in a partitioned Parquet [70] format that allows for further distributed processing, which is needed when handling a large number of polygons. For our end goal, we require a list of geometrical shapes mapped with a label for each polygon. Step 1.2 is crucial in standardizing the LPIS data to apply uniform processing algorithms later.

Step 1.3 continues our standardization by selecting a set of columns of interest, renaming them based on a chosen convention way, and translating the corresponding values to English. For each LPIS, we considered the crop type, crop group, area, and geometries-related columns. We believe crop type and geometries are mandatory fields, as they include the preliminary information we need for the proposed dataset. Crop group and area information are supplementary materials useful for statistical analysis. If the area for each polygon was not provided, we automatically computed it. Next, we proceeded with the English translation of the unique labels from the original LPIS. For a proper translation, we identified the correct encoding for each original LPIS file. The translations were manually corrected for each country to remove errors. This was a time-consuming manual process, but it was necessary to align the labels among various LPIS systems. Output 1.3 includes the English version of the LPIS data, with the same column naming for each country, together with a list of the corresponding encodings and translation of each label. Both

types of information serve as a starting point for further extension of the current dataset. Sen4AgriNet also offers translation data for France and Catalonia. However, there are differences between our proposed translation and theirs, probably due to our translation revision process, which improves the quality of the final output.

Step 1.4 addresses another issue created by the absence of a standard: distinct names for the same crop label class. This problem exists even for the same region when analyzing data from different years. We used the FAO Indicative Crop Classification (ICC) [15] categories to solve this and map each crop type to a new label. We followed the same naming convention as Sen4AgriNet, as we created AgriSen-COG to integrate with existing datasets. Furthermore, we believe that using a clear standard for crop labeling is helpful in further dataset extensions. FAO ICC uses a taxonomy that divides the crop types based on group > class > subclass > order. For AgriSen-COG, we chose the innermost ICC label and attached a number code to each crop type. There are 168 classes and subclasses in all, making up the custom FAO/CLC classification scheme. We incorporated two supplementary classes, Fallow land and background, in our final GT. In addition, each region's resulting file (Output 1.4) contains all the upper levels from FAO ICC, including group, class, subclass, and order.

(2) Preparing rasterized data

Step 2.1 starts our rasterization process for converting the LPIS data into actual raster data. It consists of finding the exact boundaries of each area of interest (AOI). The limits are needed in the next step to finding the intersecting Sentinel-2 tiles. One may choose between a region/country's actual border coordinates to retrieve the boundaries or extract them from the LPIS file. We tested both approaches and decided to follow the latter. Even though the first version is faster, as the border files are publicly available, we are only interested in the region with agricultural representation. Therefore, we computed the boundaries from the previously generated LPIS. In this way, we eliminated from the start all of the Sentinel-2 tiles that do not intersect with any crop polygons.

Step 2.2 continues with the discovery of the Sentinel-2 tiles that will serve as input data for the proposed dataset. To ease the searching process, we used the S2 Amazon STAC catalogue (AWS S2 COGs STAC: https://registry.opendata.aws/sentinel-2-l2a-cogs/, accessed on 4 June 2023). We conducted STAC searching queries based on each AOI's boundary, cloud percentage, and our dates of interest.

Step 2.3 starts the actual rasterization of our LPIS information. The advantage of using S2 COGs is that we can study a Sentinel-2 tile without needing to download it. We took the unique tile regions identified in the previous step and used their bounding boxes to map the LPIS polygon on a new raster for each tile. The result (Output 2) was a georeferenced array for each tile. These constitute the ground truth data of the proposed AgriSen-COG. We used the previously mapped FAO ICC code to generate the pixel values for each geometry. The raster images were generated by matching each Sentinel-2 tile's coordinate reference system (CRS). The GT raster was released under the following formats: Geotiff [71], as it is a popular geo-format; and Zarr [72], to enable distributed processing; and COGs (cloud-optimized Geotiff: https://www.cogeo.org, accessed on 4 June 2023), to allow image access without downloading the data.

(3) Improving dataset quality with Anomaly Detection

We integrated an identification of mislabeled GT as a preprocessing step for our dataset creation. To our knowledge, AgriSen-COG is the first crop type dataset to incorporate an anomaly detection step to curate the data. Our goal is to identify the mislabeled crop parcels. Therefore, we need to prepare aggregated information at the polygon level. The workflow for our data preparation and anomaly detection process is described in Figures 4 and 5.

Step 3.1 starts our data preparation for the anomaly detection task. First, we computed the NDVI index ($\frac{NIR-Red}{NIR+Red}$) to capture the characteristics of our crop vegetation while

reducing the multichannel structure to a one-channel image. The NDVI was computed for each tile at a pixel level to preserve the temporal and spatial dimensions.

Step 3.2 continues with cloud masking the NDVI image. As the NDVI was later used for anomaly detection, clouds would alter the process and include bias in a polygon's time series. Therefore, we applied a cloud mask on each image. We decided to use the SCL mask, already delivered with the Sentinel-2 product, eliminating the need for another cloud processing algorithm to be added to our workflow. The SCL mask offers comparable results to top cloud masking methods [73]. From the 12 labels present in the SCL mask, we implemented the cloud- and snow-related pixel classes, namely saturated or defective, cast shadows, cloud shadows, cloud medium probability, cloud high probability, thin cirrus, and snow or ice.

Figure 4. Overview of the anomaly detection preprocessing workflow.

Figure 5. Overview of the anomaly detection workflow.

Step 3.3 assembles our NDVI time series. Each pixel from our dataset has the following properties: (1) a sequence of NDVI values, representing the vegetation characteristics captured at different moments; (2) a crop label, describing the corresponding crop class (from LPIS); (3) a polygon identifier, assimilated to a number given to each polygon from the LPIS data. Initially, the information is stored as multiple matrixes of pixels, which are transformed into sequences having the values mentioned earlier.

Step 3.4 is the final data processing step before applying the LSTMAutoencoder for anomaly detection. As we are only interested in detecting anomalies at the polygon level, we aggregated the time series corresponding to pixels from the same polygon. Possible aggregations include polygon median, mean, or computing the barycenter. In our time series, we might have missing data for the same polygon due to cloud masking or just missing data from the original Sentinel-2 image. The median and mean are more sensitive to the missing data situations, as mentioned earlier. Therefore, we computed the barycenter for each polygon, capturing the time-related variability of each polygon. The barycenter (Equation (1)) was computed using the DTW barycenter averaging (DBA) [54] algorithm. The barycenter is a sequence for each polygon that reflects a crop's growing cycle from the respective parcel.

$$barycenter(D) = \min_{\mu} \sum_{x \in D} DTW(\mu, x)^2 \tag{1}$$

Step 3.5 starts our anomaly detection process by grouping the barycenter time series for each crop type. As in [55,62], we expected most crop labels from the same class to be correct, and aimed to identify the outliers only. As we had a large variability regarding the number of time-series for each category (from a few hundred to ten thousand), we chose an autoencoder network instead of a KNN with DTW. We eliminated the dates without input for each time series and used interpolation to fill in missing values.

Step 3.6 corresponds to the actual model training, as we trained a LSTM Autoencoder for each category. The architecture of our models is described in Table 2 and is based on the LSTM autoencoders from here (LSTM autoencoders: https://github.com/shobrook/sequitur, accessed on 4 June 2023). The proposed network follows a classic autoencoder structure composed of encoder and decoder parts. In our case, the encoder and decoder use two LSTM layers, followed by a fully connected layer at the end of the decoder.

Table 2. Architecture of the LSTM autoencoder.

Layer	Input Size	Hidden Size	Number of Recurrent Layers	Output Size
		Encoder		
Input time series	[batch_size, seq_len, n_features]	-	-	-
LSTM Layer	1	256	1	[batch_size, seq_len, 256]
LSTM Layer	256	128	1	[batch_size, seq_len, 128]
		Decoder		
LSTM Layer	128	128	1	[batch_size, seq_len, 128]
LSTM Layer	128	256	1	[batch_size, seq_len, 256]
FC Layer	256	-	-	[batch_size, seq_len, 1]

Step 3.7 consists of passing again through the trained autoencoder model in prediction mode to record the prediction error. The autoencoder tries to reconstruct the input by minimizing the reconstruction loss. We chose the mean squared error (MSE—Equation (2)) loss for our model. We saved the value of the MSE for each sequence and used a threshold to identify the anomalies based on it.

$$MSE = \frac{1}{M} \sum_{i=1}^{M} (y_i - \hat{y}_i)^2 \tag{2}$$

Step 3.8 identifies the outliers based on the MSE loss values determined in the previous step. We have an array of prediction loss for each crop type label, on which we apply a threshold to separate the regular class from the possible abnormalities. Even though the threshold might be chosen by a visual analysis of the distribution, in our case, we have more than 50 label types for each country. Therefore, we proceeded with a dynamic threshold, as in [62], the Otsu thresholding. This technology, which was initially developed to convert gray-level photos into black-and-white images, enables the separation of a histogram with two spikes. It looks for a binary threshold that yields the least intraclass variance when the two groups are averaged. We computed the Otsu thresholding for each category and eliminated the crop parcel with higher values than the corresponding threshold for each class. The Otsu thresholding is defined in Equation (3), where $\omega_1(t)$ and $\omega_2(t)$ is the empirical probability that the loss is equal or below t, respectively above. The variance of normal/abnormal values is reflected in $\sigma_1^2(t)$ and $\sigma_2^2(t)$.

$$\text{find } t \text{ that minimizes } \sigma^2(t) = \omega_1(t) * \sigma_1^2(t) + \omega_2(t) * \sigma_2^2(t) \tag{3}$$

5.4. Dataset Description

The resulting AgriSen-COG is a multiyear, multicountry dataset for crop type mapping. It includes 2019 and 2020 data covering the following five areas: Austria, Belgium, Catalonia, Denmark, and the Netherlands. We used the corresponding LPIS information for each region, distributed under the Open Data Commons Attributions Licence. We selected the years 2019 and 2020, summing up to 10.2 M parcels (the detailed distribution of polygons is presented in Table 3). Each original AOI includes a large and varied number of unique labels (Table 3), mapped to FAO ICC standard, resulting in, at most, 102 common crop classes, including the additional *Fallow* category. The noncrop pixels are marked with the *background* label. In Figure 6, we present a sample for the proposed dataset from all six regions, including both years for the same area. As depicted, we highlight the spatial variability and temporal changes included in the AgriSen-COG dataset. Our GitHub repository provides a more thorough explanation, examples of data loading functions, and graphic demonstrations. Additionally, code samples are offered to help people write the

logic presented in the creation methodology, regarding both LPIS processing and data rasterization.

Table 3. LPIS original information for 2019-2020 used in AgriSen-COG.

AOI	Source	Number of Parcels	Number of Unique Labels
Austria	LPIS Austria [74]	5,144,532	220
Belgium	LPIS Belgium [75]	1,046,725	300
Catalonia	LPIS Catalonia [76]	1,283,820	176
Denmark	LPIS Denmark [77]	1,171,409	320
Netherlands	LPIS Netherlands [78]	1,592,285	377
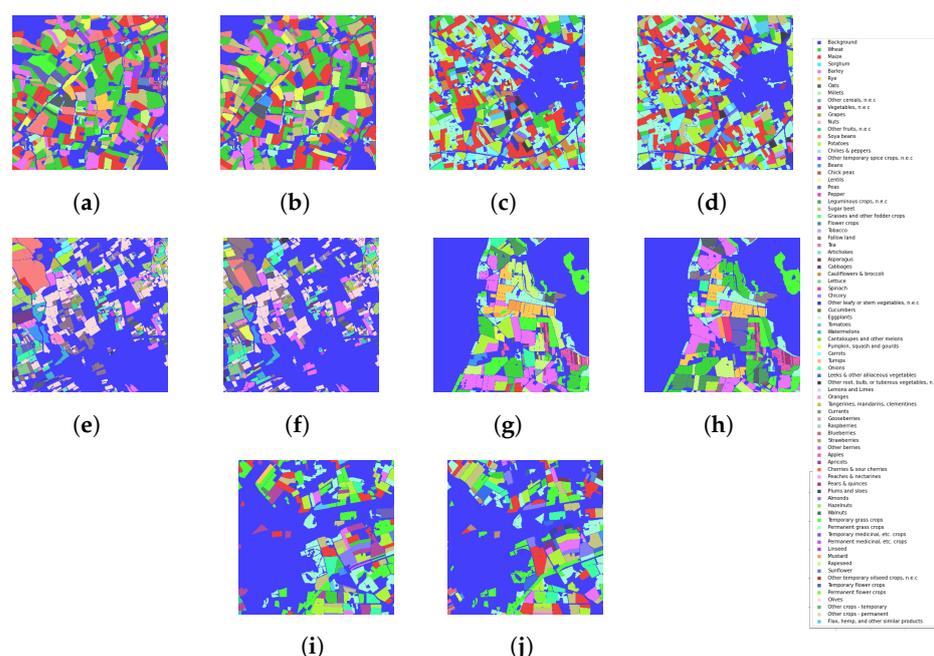
Figure 6. Sample patches from the proposed dataset (AgriSen-COG): (**a**) Austria 2019 33UVP; (**b**) Austria 2020 33UVP; (**c**) Belgium 2019 31UDS; (**d**) Belgium 2020 31UDS; (**e**) Catalonia 2019 31TCF; (**f**) Catalonia 2020 31TCF; (**g**) Denmark 2019 32UNG; (**h**) Denmark 2020 32UNG; (**i**) Netherlands 2019 31UET; (**j**) Netherlands 2020 31UET.

The proposed dataset contains two subsets created to match the two approaches in crop type mapping: pixel-level patch subset (for temporal semantic segmentation) and parcel-level aggregated subset (for time-series classification). Both subsets contain data from all five regions, covering two years (2019, 2020). Therefore, we enable further research focused on a single area or studying how models handle different geographical characteristics.

We followed the methodology mentioned earlier in creating AgriSen-COG. It relies on a total of 62 Sentinel-2 tiles. We selected the tiles that intersected with the LPIS bounds of each region and retrieved the Sentinel-2 Level-2A tiles with less than 30% cloud percentage. Next, we rasterized the LPIS polygons, following the bounding boxes of each tile, but we discarded the parcels with less than 0.1 ha area. After rasterization, we applied our anomaly detection algorithm and identified possible anomalous fields. Therefore, we eliminated the corresponding polygons by labeling them as the background class. Ultimately, we divided each tile into patches so that our data fit the hardware restrictions.

The proposed dataset includes the 10m resolution bands only (red, green, blue, and near-infrared), as they provide most of the vegetation-related information and they do not require further upsampling to use. Therefore, we chose size 366 × 366 for each patch, an integer division with the initial tile size for the 10 m resolution bands. The patch size, as mentioned earlier, also makes AgriSen-COG compatible with other datasets, such as Sen4AgriNet. From each tile, there is a total of 900 resulting patches. However, we

discarded the patches that did not include any crop-related polygon, summing up to 41,100 patches in AgriSen-COG. Table 4 presents a detailed description of the eliminated polygons and patches during each stage.

Table 4. AgriSen-COG data summary.

AOI	S2 Tiles	Number of Parcels with Area >= 0.1 ha (% Kept from the Original Data)	Number of Parcels after Anomaly Detection (% Kept from the Data with Area >= 0.1 ha)	Final Number of Patches
Austria	20	4,115,899 (80%)	3,788,326 (92%)	16,563
Belgium	5	898,542 (85.8%)	611,743 (68%)	2607
Catalonia	10	1,006,573 (78.4%)	719,631 (71.5%)	5166
Denmark	17	1,061,070 (90.6%)	924,914 (87%)	9685
Netherlands	10	1,457,810 (91.5%)	927,871 (63.64%)	7079
Total	62	8,539,894 (83.47%)	6,972,485 (81.65%)	41,100

The patches are saved as Zarr arrays in a format offering a self-describing design compatible with Xarray. The selected format is also compatible with a distributed processor (like Dask) and is the preferred format for cloud-stored data. The ground truth (LPIS masks) data files are stored using the COG format and are available on a public S3 bucket. This way, we enrich the existing COGs databases and make the proposed data easily accessible (no download needed) and findable (STAC catalogues indexes). The aggregated data for each field (the barycenters) constitute the proposed time-series dataset, and are distributed in Parquet format to ensure a smaller size and distributed processing if needed.

The five AOIs comprise around 62 Sentinel-2 tiles, with the patches dataset summing up to 6,972,485 fields, for 2019–2020, with 41,100 patches.

6. Experimental Results

This section presents our experiments performed on the proposed AgriSen-COG dataset, for 2019 and 2020, including all six regions. We created the training, validation, and testing datasets using label stratification, with a ratio of 60%-20%-20% for each set. The input data were aggregated into a monthly median to lower the number of input timesteps. The experiments were based on the four 10 m resolution bands (red, green, blue, and nir) within a period of 6 months for each year, from month 4 (April) to 9 (September).

Each AOI has a different number of labels, ranging between 60–80 of other classes. However, to help the training process, we reduced the number of crops and chose 11 common categories: wheat, maize, barley, oats, rapeseed, potatoes, peas, rye, sunflower, sorghum, and grapes. The distribution of the selected classes is depicted in Figure 7) for both of the proposed datasets. We notice that wheat, maize and barley are the dominant crop categories. However, there is a major difference given by the dataset type. For example, the number of wheat and maize field crops in Austria is larger by more than 50k than in Denmark. However, when compared to the actual number of pixels from each parcel, Denmark comprises more than 30 M pixels for the respective categories.

Based on the barycenters computed from NDVI in the anomaly detection step, we depict (Figure 8) how the barycenter changes over time for each selected AOIs, for two representative crops: wheat and maize. We observe similar variations in the crops' growing cycles for each region, even though they are shifted from one year to another. We also encounter differences between the AOIs for the same culture. Therefore, the proposed dataset incorporates the real-world challenges of crop detection: spatial and temporal variation.
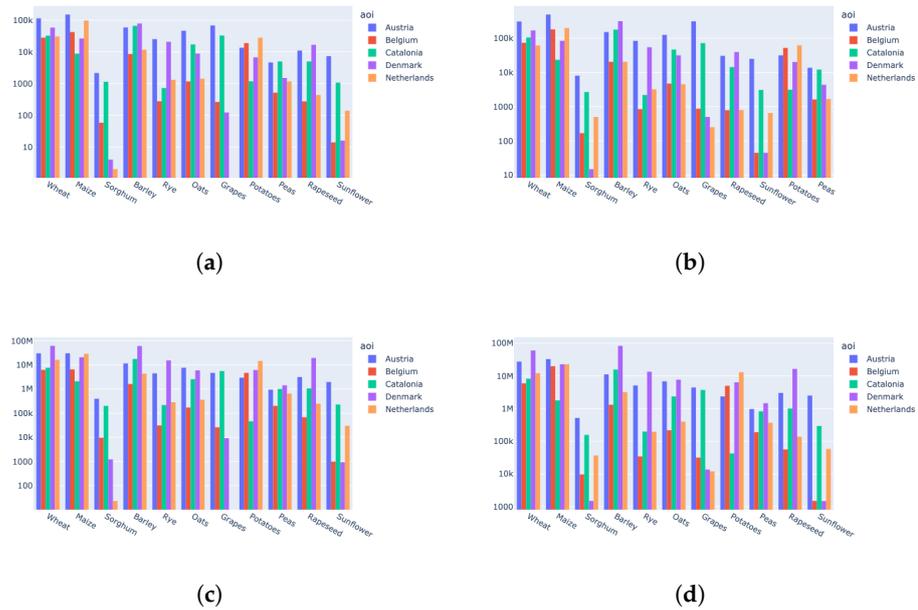
Figure 7. Selected crop distribution for the barycenter time-series and pixel-wise patch datasets. Vertical axis is in logarithmic scale. (**a**) Crop time-series distribution for 2019, (**b**) Crop time-series distribution for 2020, (**c**) Crop pixel distribution for 2019, (**d**) Crop pixel distribution for 2020.
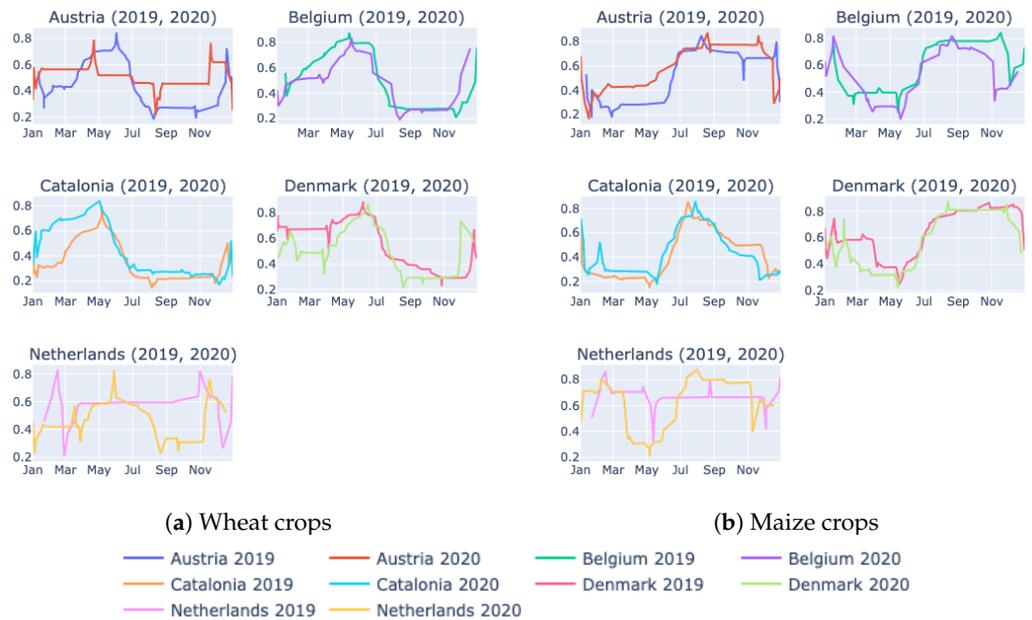
Figure 8. Barycenters over NDVI for different crops.

This work aims to train DL models for two crop-related tasks: crop type classification and crop type mapping. Performing this shows how the proposed dataset helps handle both scenarios while covering crops' spatial and temporal variability. Therefore, we present the following experimental strategies:

1. Experiment Type 1 (anomalies variation): We conduct individual experiments on each AOI for a single year, with one model, to highlight the importance of curated data labels.

2. Experiment Type 2 (temporal generalization): We conduct individual experiments on each AOI using the model trained in Experiment Type 1 and predict the instances for 2020.

3. Experiment Type 3 (spatial generalization): We conduct several experiments, using data from one year and splitting our data based on regions' similarity in crop patterns.
4. Experiment Type 4 (overall generalization): We train on two AOIs for 2019, with different models (LSTM, Transformer, TempCNN, U-Net, ConvStar) to see the behavior of the proposed dataset.

The crop type mapping use case was assimilated to a semantic segmentation problem, and therefore, we employed the popular U-Net model. In addition to this, we also used a temporal-designed model, the ConvStar, for comparison. The crop type classification approach is based on aggregated time series at the parcel level. Previous work included mean-aggregated time series; however, we also propose, for comparison, the barycenter series computed for each crop polygon. This approach was tested with three types of models for time series: LSTM, transformer, and TempCNN. We evaluated our experiments using a weighted F1 score as we encounter high-class imbalance. Moreover, we included several normalized confusion matrices for further visual analysis.

6.1. Crop Type Classification Experiments—Time-Series Classification

In this paper, we used the barycenters computed for each crop parcel for crop type classification. The barycenters were previously also used to eliminate possible anomalous polygons. This set of experiments serves as a proof-of-concept on using a different metric for aggregation crop parcel information instead of the classic mean, median, or standard deviation.

We resampled our time series in the proposed experiments to obtain a weekly value. By carrying this out, we preserved the crop growth information while still obtaining a regular interval for our time-series data. However, the experiments presented in this paper used information from month 4 (April) to month 9 (September), as it captured the growing cycles of the analyzed cultures.

LSTM is the first architecture used to assess our barycenter-based dataset. LSTMs are a popular choice in dealing with time-series data; we used them for most experiments due to their simplicity. The proposed network consists of three bidirectional LSTM layers, with one input feature and a hidden size of 128. The result is then passed through two fully connected layers, with a ReLU function in between. A final Softmax function is applied for classification. The model includes 349 k trainable parameters. The experiments using LSTM employ cross entropy as a loss function, the Adam optimizer, with a starting learning rate of 0.001 and a learning rate scheduler decrease of 0.1 every five epochs.

Experiment Type 1 (anomalies variation): The first set of experiments aims to highlight how curated data labels impact a model's performance. Using the barycenters for each polygon, we removed the influence given by larger parcels from a category and focused only on detecting crop growing patterns. We conducted ten experiments, two for each included AOI (Austria, Belgium, Catalonia, Denmark, Netherlands), for 2019, using the initial time-series data and the curated data after the anomaly detection process. The results of the experiments are presented in Table 5, showing the score obtained using the initial version of the dataset (v0) and the time series after the anomaly detection process (v1). We observe that the model trained with the curated dataset outperforms in the overall score for each selected AOI, showing the benefit of training with more trustworthy labels.

Experiment Type 2 (temporal generalization): On the time-series data, the second set of experiments aims to show how the dataset scores for the same AOI while being trained on the 2019 time series and using the 2020 data for testing. We used the same LSTM models trained in the previous experiments; however, we considered only the curated version of our proposed dataset. Table 5 reflects the results obtained for each region using our dataset (v1 2020). For most of our areas, we observed a considerable decrease in the overall score, which shows the variability of the crops in time, even for the same AOI. Corresponding confusion matrices (Figure 9a–j) illustrate a detailed view of the eleven labels considered. For Austria, we observed a decrease in correctly identifying crops for all the labels except maize. However, for Belgium, we noticed consistency for the wheat category and a lack of

recognition of the maize crop from 2020. Sorghum raises challenges for Belgium in both years. Catalonia results highlight a significant difference between the 2019 and 2020 crops, with consistency only for the maize and grapes categories. The Denmark area provides the closest results between the two years, indicating a low shift between the crop-growing cycles. Finally, for the Netherlands, we obtained similar results for the wheat and potatoes labels, with a score decrease for the maize crop. The results show that the proposed dataset incorporates the challenges of the time variability of crop growing cycles, making it suitable for further developing stable crop monitoring methods.

(a)   (b)   (c)   (d)   (e)

(f)   (g)   (h)   (i)   (j)

(k)   (l)   (m)   (n)   (o)
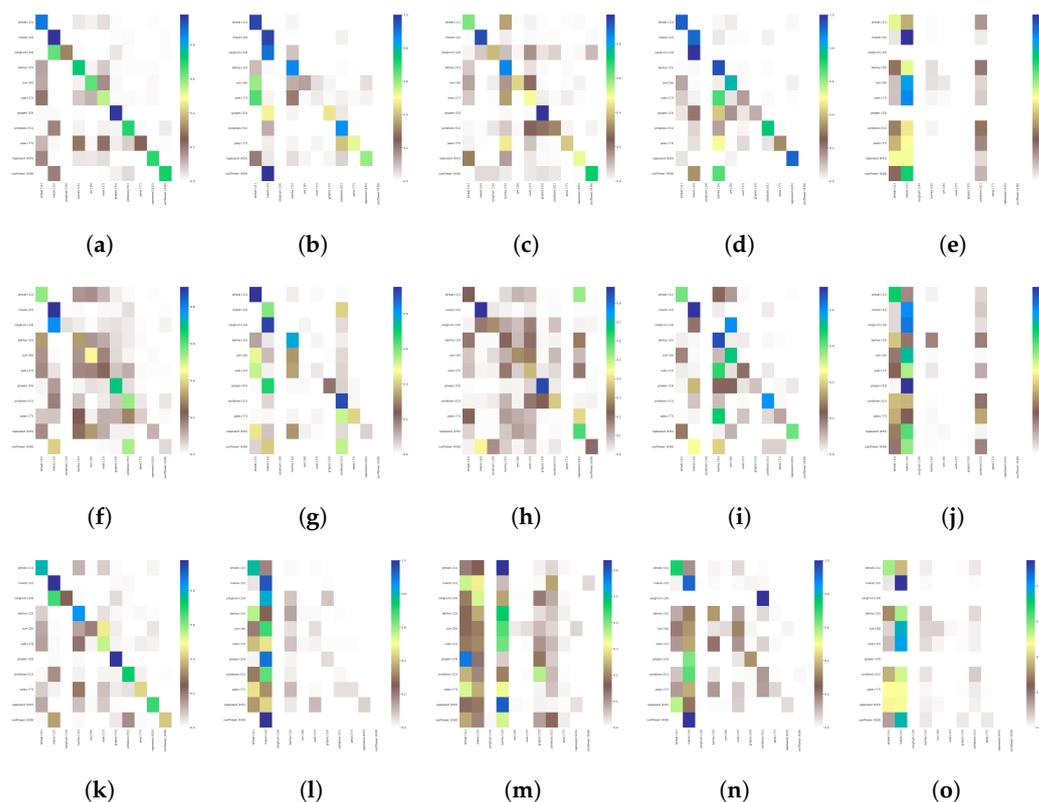
Figure 9. Confusion matrixes for Experiment Type 2 (temporal generalization) and results. Experiment Type 3 (spatial generalization). LSTM model trained with 2019 time series-dataset and tested with both 2019 and 2020 time series: (**a**) predicted labels Austria 2019; (**b**) predicted labels Belgium 2019; (**c**) predicted labels Catalonia 2019; (**d**) predicted labels Denmark 2019; (**e**) predicted labels Netherlands 2019; (**f**) predicted labels Austria 2020; (**g**) predicted labels Belgium 2020; (**h**) predicted labels Catalonia 2020; (**i**) predicted labels Denmark 2020; (**j**) predicted labels Netherlands 2020; (**k**) predicted labels Austria 2019 (Type 3); (**l**) predicted labels Belgium 2019 (Type 3); (**m**) predicted labels Catalonia 2019 (Type 3); (**n**) predicted labels Denmark 2019 (Type 3); (**o**) predicted labels Netherlands 2019 (Type 3).

Experiment Type 3 (spatial generalization): The third set of experiments aims to show how the proposed time series can capture crop-growing cycle similarities between different regions. We only used the 2019 year for training and testing. An LSTM model was trained using three AOIs (Austria, Catalonia, and the Netherlands) and tested using the five areas from the proposed dataset. The similarity between the Austrian and Dutch crops' time series, as opposed to Catalonia, is reflected in the results presented in Table 5 (v1 2019 Type 3 rows). We observed a significant decrease in the scores obtained for the Catalan region. A lower score was obtained for the two AOIs (Belgium and Denmark). However, they still performed better than Catalonia due to their similarity to Austria and Netherlands. From the confusion matrixes (Figure 9k–o), we observe that maize and wheat crops are strongly

identified in most of the regions. In contrast, for Belgium and Denmark, rapeseed and barley are classified with lower accuracy.

Experiment Type 4 (model behavior): The last experiments present how different models perform when trained with the proposed dataset. We experimented with two popular time-series models for the proposed barycenter data: the transformer model and the TempCNN. We chose two of our AOIs, Denmark and Catalonia, as they present different crop markers. We trained the models using the 2019 data and predicted using 2019 and 2020. Based on the results shown in Table 5 (v1 Type 4), we observed that TempCNN outperforms the LSTM and transformer models for both regions. However, when tested on the 2020 data, the TempCNN achieved the lowest performance in Catalonia but an increased score for Denmark.

Table 5. Results for the crop type classification experiments. The best results are in bold for each AOI.

Score	Version	Exp.	Model	Austria	Belgium	Catalonia	Denmark	Netherlands
	v0 2019	Type 1	LSTM	0.819	0.917	0.747	0.862	0.642
	v1 2019	Type 1	LSTM	**0.841**	**0.924**	**0.762**	**0.885**	**0.659**
	v1 2020	Type 2	LSTM	0.614	0.719	0.369	0.752	0.657
Acc. W. (%)	v1 2019	Type 3	LSTM	0.805	0.637	0.263	0.454	0.652
	v1 2019	Type 4	Transformer	-	-	0.667	0.791	-
	v1 2020	Type 4	Transformer	-	-	0.402	0.688	-
	v1 2019	Type 4	TempCNN	-	-	0.782	0.899	-
	v1 2020	Type 4	TempCNN	-	-	0.331	0.713	-
	v0 2019	Type 1	LSTM	0.815	0.907	0.739	0.851	0.580
	v1 2019	Type 1	LSTM	**0.837**	**0.915**	**0.758**	**0.873**	**0.600**
	v1 2020	Type 2	LSTM	0.640	0.830	0.549	0.755	0.624
Prec. W (%)	v1 2019	Type 3	LSTM	0.813	0.604	0.350	0.632	0.592
	v1 2019	Type 4	Transformer	-	-	0.640	0.836	-
	v1 2020	Type 4	Transformer	-	-	0.489	0.727	-
	v1 2019	Type 4	TempCNN	-	-	0.776	0.894	-
	v1 2020	Type 4	TempCNN	-	-	0.489	0.757	-
	v0 2019	Type 1	LSTM	0.816	0.911	0.739	0.851	0.593
	v1 2019	Type 1	LSTM	**0.838**	**0.918**	**0.755**	**0.874**	**0.613**
	v1 2020	Type 2	LSTM	0.618	0.733	0.401	0.745	0.625
F1 W(%)	v1 2019	Type 3	LSTM	0.796	0.548	0.267	0.439	0.586
	v1 2019	Type 4	Transformer	-	-	0.626	0.779	-
	v1 2020	Type 4	Transformer	-	-	0.420	0.664	-
	v1 2019	Type 4	TempCNN	-	-	**0.777**	**0.894**	-
	v1 2020	Type 4	TempCNN	-	-	0.349	0.716	-

6.2. Crop Mapping Experiments—Semantic Segmentation

The crop mapping experiments present the second use case of the proposed dataset that uses the actual parcel's spatial and temporal information. As in the previous case, we only used data from month 4 (April) to month 9 (September). The crop mapping application is practical when a crop's geometry is unknown. Therefore, the task is to discover a crop's type and the parcel's geometry. As it resembles the problem of semantic segmentation, we used the popular U-Net model to test the utility of the proposed dataset. Our experiments were based on the four 10m resolution bands, resulting in four input channels. In addition to this, we also incorporated the crop's temporal information, using a monthly median, with the final input data in the shape of [T, C, H, W] (T: timesteps, C: channels, H: height, W: width). As U-Net is not designed to support the temporal size, we concatenated the first two dimensions, achieving an input image shape of [T×C, H, W]. In the latter part of the proposed experiment, we also experimented with two models for temporal semantic segmentation: U-Net and ConvStar.

We created smaller nonoverlapping chips of 61 × 61 (H×W) for faster training from our initial patches of 366 × 366 pixels. Therefore, our input for training was in the shape of 6 × 4 × 61 × 61 (T×C×H×W). Due to the high-class imbalance, given not only by the presence of dominating crops (as with the time series) but also due to the size of the pixels from specific parcels, we use a weighted negative log-likelihood loss. During the training

and testing stages, we masked all the pixels that did not correspond to the 11 selected classes as background. The Adam optimizer was employed for training, starting with a learning rate of 0.001, with a decrease strategy applied to validation loss plateaus.

Experiment Type 1 (anomaly variation): In the first series of experiments, we analyzed how curated data labels affect the performance of a model. As in the previous case, we conducted ten experiments in 2019, two for each included AOI (Austria, Belgium, Catalonia, Denmark, and the Netherlands), using the initial parcel's dataset and the one that was curated. The results of the experiments are shown in Table 6, which displays the score derived with the initial version of the dataset (v0) and the time series after the anomaly detection process (v1). As opposed to the experiments based on time series, the current results are influenced by the number of pixels from each parcel. However, we also observe an improvement in the curated dataset for each AOI, even for the proposed pixel-based segmentation. Therefore, it shows the value of training with more reliable labels.

Experiment Type 2 (temporal generalization): The second set of experiments on the time-series data seeks to demonstrate how the dataset performs for the same AOI when trained on the 2019 time series and tested with the 2020 data. We employed the same U-Net models trained in previous experiments, but we only considered the curated version of our proposed dataset. Table 6 reflects the results obtained for each region using our dataset (v1 2020). Except for the Netherlands, there was a decrease in the overall F1 score. Even though we obtained good accuracy and precision on the 2020 dataset, we achieved poor recall, as the models struggled to identify actual pixels from each category. The previously mentioned tradeoff between precision and recall is visible for each crop category for all AOIs. Corresponding confusion matrices (Figure 10) illustrate a detailed view of the eleven labels considered. By incorporating the pixel-based characteristic for crop mapping, we observe that the models can better classify crops from one year to another. The findings demonstrate that the proposed dataset addressed the difficulties provided by the temporal unpredictability of crop growth cycles, making it appropriate for the continuing development of reliable crop monitoring techniques.

Experiment Type 3 (spatial generalization): The third set of experiments exemplifies how the proposed pixel-based dataset captures regional variations in crop-growing cycles. Similar to the time-series experiments, we only used the year 2019 for training and testing. We trained a U-Net model using three AOIs (Austria, Catalonia, and the Netherlands) and tested it using the five areas from the proposed dataset. As seen in Table 6), the resemblance between the Austrian and Dutch crops is highlighted, with an actual improvement in the overall score for the Netherlands. As seen in the corresponding confusion matrixes (Figure 10k–o), the Catalonia region obtains a lower score. Even though maize, wheat, and barley crops are identified, the sorghum class is mislabeled as wheat and barley. Maize and wheat are also classified with a good score for Belgium and Denmark; however, barley, oats, and grapes are not recognized in either of the two test areas.
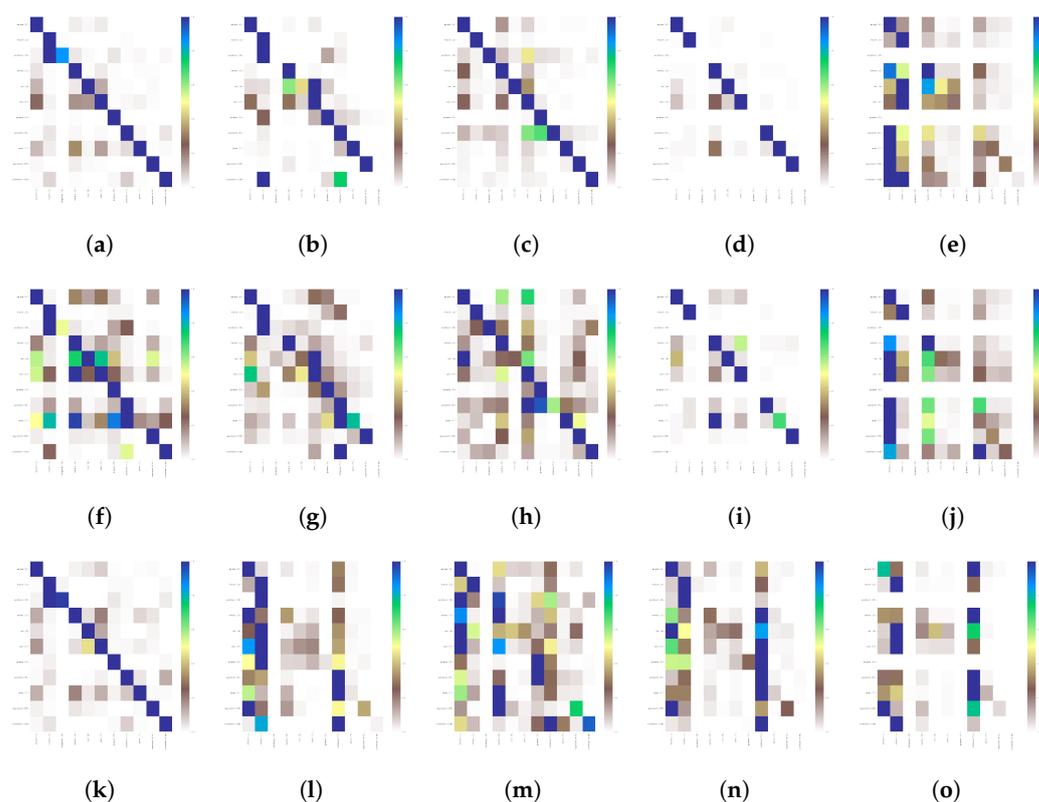
Figure 10. Confusion matrixes for Experiment Type 2 (temporal generalization) and results. Experiment Type 3 (spatial generalization). U-Net model trained with 2019 time-series dataset and tested with both 2019 and 2020 pixel-based patches: (**a**) predicted labels Austria 2019; (**b**) predicted labels Belgium 2019; (**c**) predicted labels Catalonia 2019; (**d**) predicted labels Denmark 2019; (**e**) predicted labels Netherlands 2019; (**f**) predicted labels Austria 2020; (**g**) predicted labels Belgium 2020; (**h**) predicted labels Catalonia 2020; (**i**) predicted labels Denmark 2020; (**j**) predicted labels Netherlands 2020; (**k**) predicted labels Austria 2019 (Type 3); (**l**) predicted labels Belgium 2019 (Type 3); (**m**) predicted labels Catalonia 2019 (Type 3); (**n**) predicted labels Denmark 2019 (Type 3); (**o**) predicted labels Netherlands 2019 (Type 3).

Experiment Type 4 (model behavior): Our final experiments aim to illustrate how various models perform when trained using the dataset. Alongside the popular U-Net for semantic segmentation, we also experimented with ConvStar, a model created for temporal semantic segmentation. We chose two of our AOIs, Denmark and Catalonia, as they present different crop markers. We trained the models using the 2019 data and predicted using 2019 and 2020. Based on the results shown in Table 6 (v1 Type 4), the temporal U-Net still obtains a better overall score than the ConvStar network. In Figure 11, we illustrate two test patches from Catalonia and Denmark, representing the patch for both 2019 and 2020. We observe crop variability in the two samples, where the same parcels shifted their crops from rye to wheat (Catalonia) and maize to barley (Denmark). For Catalonia, the ConvStar model fails to identify the oats' culture, mislabeling it with rye. However, the yearly changes from rye to wheat for the same parcels are better captured for the patch by the ConvStar model, as U-Net wrongly classifies rye as wheat.
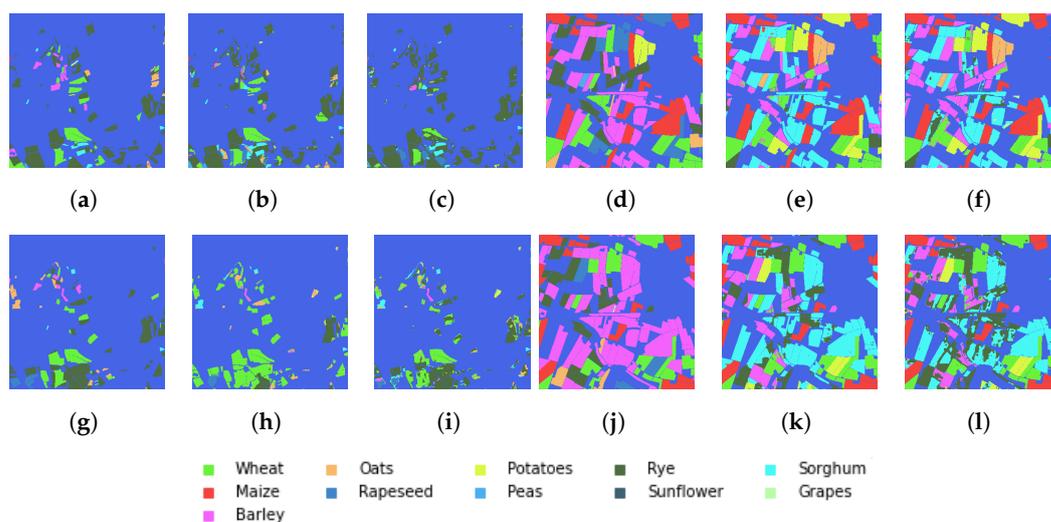
Figure 11. Predicted samples for Catalonia and Denmark. U-Net and ConvStar models trained with 2019 data. Tested using 2019 and 2020 data: (**a**) Catalonia generated ground truth (GT) 2019; (**b**) Catalonia U-Net prediction 2019; (**c**) Catalonia ConvStar prediction 2019; (**d**) Denmark generated ground truth (GT) 2019; (**e**) Denmark U-Net prediction 2019; (**f**) Denmark ConvStar prediction 2019; (**g**) Catalonia generated ground truth (GT) 2020; (**h**) Catalonia U-Net prediction 2020; (**i**) Catalonia ConvStar prediction 2020; (**j**) Denmark generated ground truth (GT) 2020; (**k**) Denmark U-Net prediction 2020; (**l**) Denmark ConvStar prediction 2020.

Both models can identify the wheat and maize class for the Denmark patch but struggle with differentiating between the sorghum and barley crops. As of 2020, parcels with maize have been replaced with barley, and the models encounter the same problem, therefore misclassifying barley as sorghum. However, for the 2020 sample, we also observe that the ConvStar model struggles with differentiating between the rye and sorghum categories.

Table 6. Results for the crop type mapping experiments. The best results are in bold for each AOI.

Score	Version	Exp.	Model	Austria	Belgium	Catalonia	Denmark	Netherlands
Acc. W. (%)	v0 2019	Type 1	U-Net	0.967	0.979	0.887	0.974	0.771
	v1 2019	Type 1	U-Net	**0.969**	**0.981**	**0.894**	**0.982**	**0.783**
	v1 2020	Type 2	U-Net	0.904	0.883	0.799	0.901	0.786
	v1 2019	Type 3	U-Net	0.954	0.801	0.756	0.761	0.775
	v1 2019	Type 4	ConvStar	-	-	0.882	0.973	-
	v1 2020	Type 4	ConvStar	-	-	0.804	0.893	-
Prec. W (%)	v0 2019	Type 1	U-Net	0.879	0.956	0.787	0.938	0.581
	v1 2019	Type 1	U-Net	**0.888**	**0.961**	**0.814**	**0.951**	**0.585**
	v1 2020	Type 2	U-Net	0.702	0.857	0.658	0.823	0.655
	v1 2019	Type 3	U-Net	0.851	0.618	0.520	0.526	0.588
	v1 2019	Type 4	ConvStar	-	-	0.790	0.939	-
	v1 2020	Type 4	ConvStar	-	-	0.687	0.871	-
F1 W(%)	v0 2019	Type 1	U-Net	0.880	0.952	0.767	0.936	0.523
	v1 2019	Type 1	U-Net	**0.8886**	**0.956**	**0.793**	**0.951**	**0.541**
	v1 2020	Type 2	U-Net	0.664	0.762	0.584	0.768	0.556
	v1 2019	Type 3	U-Net	0.835	0.607	0.444	0.329	0.561
	v1 2019	Type 4	ConvStar	-	-	0.758	0.928	-
	v1 2020	Type 4	ConvStar	-	-	0.550	0.779	-

7. Discussion

Agricultural monitoring implies several crop-related tasks: crop type classification, crop mapping, parcel extraction, and crop phenology. The availability of Sentinel-2 satellite data enables further development in agricultural monitoring due to the global coverage, multispectral bands at different resolutions, and temporal dimensionality. As official LPIS

data from multiple regions became available, we created the AgriSen-COG dataset to fill in current demands for DL datasets for agricultural monitoring: multiregion, multiyear, trustworthy labels, and methodology to enable extension as new data became available. The first challenge in creating AgriSen-COG was label harmonization, as each country uses different naming conventions for their crops, and they even differ from year to year. As in [14], we followed the official FAO crop naming conventions. The second challenge involved the discovery of potential anomalous labels, as the LPIS data are created based on farmers' declarations. The anomaly detection process required the computation of a time series for each crop parcel and identifying the crops that follow a significantly different crop-growing pattern. As this process is performed on all the data, we used Sentinel-2 AWS COGs (Sentinel-2 AWS COGs repository: https://registry.opendata.aws/sentinel-2/, accessed on 4 June 2023), which enabled us to conduct the required computations without downloading a large amount of data.

The proposed dataset meets two use-case scenarios: crop type classification and crop mapping. In the first scenario, we offer the time-series dataset computed during the anomaly detection stage. It comprises a barycenter time series for each polygon from all five AOIs: Austria, Belgium, Catalonia, Denmark, and the Netherlands, for two years: 2019 and 2020. The second scenario is met by our pixel-based patches, of size 366×366, which incorporate the four Sentinel-2 10 m resolution bands and a ground truth patch with all the labels.

To show the validity of the proposed dataset, we designed four types of experiments, which also aim to highlight the potential of further research conducted using AgriSen-COG. *Experiment Type 1* compared the performance of a popular DL model trained using the original dataset and a model trained using the curated dataset. We observed improved performance for both scenarios, showing that AgriSen-COG is a trustworthy dataset for crop detection. The second experiment (*Experiment Type 2*) illustrated the importance of having a multiyear dataset for crops due to the yearly variability. Our results showed a considerable decrease in the performance of the models trained for 2019 and tested with 2020 data. Therefore, AgriSen-COG is a dataset that could be used to develop better classification methods that better incorporate the yearly temporal shift. The spatial generalization experiment (*Experiment Type 3*) revealed the need for conducting targeted research for domain adaptation to enable crop classification for a larger number of regions. AgriSen-COG incorporates the most different areas (5) as opposed to existing datasets, making it a good benchmark for this task. The final *Experiment Type 4* showed that AgriSen-COG is not only a dataset that works with simple yet popular DL models (LSTM and U-Net). Good performance was also obtained with three other types: transformer, TempCNN, and ConvStar, showing the potential of the proposed dataset to be used in the development of better DL models.

8. Conclusions

This paper proposes AgriSen-COG, a new crop type mapping dataset that gathers information from five different areas (Austria, Belgium, Catalonia, Denmark, and the Netherlands). To this moment, AgriSen-COG is the only dataset that includes information from more than two areas, comprising the largest number of tiles and crop polygons (61 tiles and almost 7 M fields). We also detail the steps taken to create the dataset (together with code for reproducibility) and provide several intermediate results. Both output assets are crucial components for a consistent future extension with new regions as the LPIS data become available. In addition to this, by offering the output both in geopopular formats (Geopackage and Geotiff) and compliant distributed storage formats (Parquet and Zarr), we support further advancement in several domains: geoanalysis on crop trends, deep learning methods for agricultural monitoring, and big data distributed processing adapted to georeferenced data. We also distribute the dataset using COGs, enabling the data's usage without prior download. In addition to this, we describe the dataset using a STAC catalogue, making it easier to discover and use.

AgriSen-COG is the only crop-related dataset that offers a DTW—based barycenter time series for each parcel, as opposed to mean or median. By undergoing the anomaly detection process, AgriSen-COG is also the only crop-related dataset with curated labels for five different regions.

Based on these characteristics, the proposed dataset may be regarded as a benchmark dataset for further development on agricultural monitoring, serving various CAP-related use cases: agricultural insurance, early warning risk assessment, or crop yield management.

Funding: This work was primarily funded by a grant from the Romanian Ministry of Research and Innovation, CNCS—UEFISCDI, project number PN-III-P4-ID-PCE-2016-0842, within PNCDI III, COCO (PN III-P4-ID-PCE-2020-0407) and the HARMONIA project, from the EU Horizon 2020 research and innovation programme under agreement No. 101003517. This work has partially supported by the European Space Agency through the ML4EO Project (contract number 4000125049/18/NL/CBI), Romanian UEFISCD FUSE4DL and by project POC/163/1/3/ "Advanced computa- tional statistics for planning and monitoring production environments" (2022–2023).

Data Availability Statement: The data presented in this study are openly available in Zenodo at https://doi.org/10.5281/zenodo.7892012 (accessed on 4 June 2023).To ease data access for large files, we also provide a Dropbox folder https://www.dropbox.com/sh/5bc55skio0o5xd7/AAAQVG3 ZmVGFNvPiltQ9Esqma?dl=0 (accessed on 4 June 2023) and a public Minio S3 bucket, name: agrisen-cog-v1, endpoint: https://s3-3.services.tselea.info.uvt.ro (Set anonymous access for download. Accessed on 4 June 2023). Upon request, for further development on Alibaba Cloud, access to private Alibaba S3 Bucket can also be granted.The code for preprocessing and training is available here: https://github.com/tselea/agrisen-cog.git (accessed on 4 June 2023)

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177.
2. Sustainable Agriculture|Sustainable Development Goals | Food and Agriculture Organization of the United Nations. Available online: https://www.fao.org/sustainable-development-goals/overview/fao-and-the-2030-agenda-for-sustainable-development/sustainable-agriculture/en/ (accessed on 20 October 2022).
3. THE 17 GOALS | Sustainable Development. Available online: https://sdgs.un.org/goals#icons (accessed on 20 October 2022).
4. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
5. Cordts, M.; Omran, M.; Ramos, S.; Scharwächter, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; Schiele, B. The cityscapes dataset. In Proceedings of the CVPR Workshop on the Future of Datasets in Vision, Boston, MA, USA, 7–12 June 2015; Volume 2.
6. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common objects in context. In Proceedings of the Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 740–755.
7. Sumbul, G.; Charfuelan, M.; Demir, B.; Markl, V. Bigearthnet: A large-scale benchmark archive for remote sensing image understanding. In Proceedings of the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 5901–5904. Reply: confirmed.
8. Schmitt, M.; Hughes, L.H.; Qiu, C.; Zhu, X.X. SEN12MS–A Curated Dataset of Georeferenced Multi-Spectral Sentinel-1/2 Imagery for Deep Learning and Data Fusion. *arXiv* **2019**. arXiv:1906.07789.
9. European Court of Auditors. *The Land Parcel Identification System: A Useful Tool to Determine the Eligibility of Agricultural Land—But Its Management Could Be Further Improved*; Special Report No 25; Publications Office: Luxembourg, 2016. https://doi.org/10.2865/39118.
10. Rußwurm, M.; Pelletier, C.; Zollner, M.; Lefèvre, S.; Körner, M. BreizhCrops: A Time Series Dataset for Crop Type Mapping. *ISPRS Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci.* **2020**, *XLIII-B2-2020*, 1545–1551.
11. Turkoglu, M.O.; D'Aronco, S.; Perich, G.; Liebisch, F.; Streit, C.; Schindler, K.; Wegner, J.D. Crop mapping from image time series: Deep learning with multi-scale label hierarchies. *arXiv* **2021**. arXiv:2102.08820.
12. Weikmann, G.; Paris, C.; Bruzzone, L. TimeSen2Crop: A Million Labeled Samples Dataset of Sentinel 2 Image Time Series for Crop-Type Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 4699–4708. https://doi.org/10.1109/JSTARS.2021.3073965.

13. Kondmann, L.; Toker, A.; Rußwurm, M.; Camero, A.; Peressuti, D.; Milcinski, G.; Mathieu, P.P.; Longépé, N.; Davis, T.; Marchisio, G.; et al. DENETHOR: The DynamicEarthNET dataset for Harmonized, inter-Operable, analysis-Ready, daily crop monitoring from space. In Proceedings of the Thirty-Fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2), Virtual, 6–14 December 2021.

14. Sykas, D.; Sdraka, M.; Zografakis, D.; Papoutsis, I. A Sentinel-2 Multiyear, Multicountry Benchmark Dataset for Crop Classification and Segmentation With Deep Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 3323–3339.

15. Food and Agriculture Organization of the United Nations. *A System of Integrated Agricultural Censuses and Surveys: World Programme for the Census of Agriculture 2010*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2005; Volume 1.

16. Hoyer, S.; Hamman, J. xarray: N-D labeled arrays and datasets in Python. *J. Open Res. Softw.* **2017**, *5*, 10. https://doi.org/10.5334/jors.148.

17. Hoyer, S.; Fitzgerald, C.; Hamman, J.; Akleeman; Kluyver, T.; Roos, M.; Helmus, J.J.; Markel; Cable, P.; Maussion, F.; et al. xarray: V0.8.0. 2016. Available online: https://doi.org/10.5281/zenodo.59499 (accessed on 4 June 2023).

18. d'Andrimont, R.; Claverie, M.; Kempeneers, P.; Muraro, D.; Yordanov, M.; Peressutti, D.; Batič, M.; Waldner, F. AI4Boundaries: An open AI-ready dataset to map field boundaries with Sentinel-2 and aerial photography. *Earth Syst. Sci. Data Discuss.* **2022**, *15*, 317–329.

19. Jordan, M.I. *Serial Order: A Parallel Distributed Processing Approach*; Technical Report, June 1985–March 1986; 1986. Available online: https://doi.org/10.1016/S0166-4115(97)80111-2 (accessed on 4 June 2023).

20. Rumelhart, D.E.; McClelland, J.L., Learning Internal Representations by Error Propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*; U.S. Department of Energy Office of Scientific and Technical Information: MIT Press: Cambridge, MA, USA, 1987; pp. 318–362.

21. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**. Available online: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf (accessed on 4 June 2023).

22. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; pp. 234–241.

23. Rakhlin, A.; Davydow, A.; Nikolenko, S. Land cover classification from satellite imagery with u-net and lovász-softmax loss. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 262–266.

24. Solórzano, J.V.; Mas, J.F.; Gao, Y.; Gallardo-Cruz, J.A. Land use land cover classification with U-net: Advantages of combining sentinel-1 and sentinel-2 imagery. *Remote Sens.* **2021**, *13*, 3600.

25. Wang, J.; Yang, M.; Chen, Z.; Lu, J.; Zhang, L. An MLC and U-Net Integrated Method for Land Use/Land Cover Change Detection Based on Time Series NDVI-Composed Image from PlanetScope Satellite. *Water* **2022**, *14*, 3363.

26. Zhang, Z.; Iwasaki, A.; Xu, G.; Song, J. Cloud detection on small satellites based on lightweight U-net and image compression. *J. Appl. Remote Sens.* **2019**, *13*, 026502.

27. Guo, Y.; Cao, X.; Liu, B.; Gao, M. Cloud detection for satellite imagery using attention-based U-Net convolutional neural network. *Symmetry* **2020**, *12*, 1056.

28. Xing, D.; Hou, J.; Huang, C.; Zhang, W. Spatiotemporal Reconstruction of MODIS Normalized Difference Snow Index Products Using U-Net with Partial Convolutions. *Remote Sens.* **2022**, *14*, 1795.

29. Ivanovsky, L.; Khryashchev, V.; Pavlov, V.; Ostrovskaya, A. Building detection on aerial images using U-NET neural networks. In Proceedings of the 2019 24th Conference of Open Innovations Association (FRUCT), Moscow, Russia, 8–12 April 2019; pp. 116–122.

30. Irwansyah, E.; Heryadi, Y.; Gunawan, A.A.S. Semantic image segmentation for building detection in urban area with aerial photograph image using U-Net models. In Proceedings of the 2020 IEEE Asia-Pacific Conference on Geoscience, Electronics and Remote Sensing Technology (AGERS), Jakarta, Indonesia, 7–8 December 2020; pp. 48–51.

31. Wu, C.; Zhang, F.; Xia, J.; Xu, Y.; Li, G.; Xie, J.; Du, Z.; Liu, R. Building damage detection using U-Net with attention mechanism from pre-and post-disaster remote sensing datasets. *Remote Sens.* **2021**, *13*, 905.

32. Wei, S.; Zhang, H.; Wang, C.; Wang, Y.; Xu, L. Multi-temporal SAR data large-scale crop mapping based on U-Net model. *Remote Sens.* **2019**, *11*, 68.

33. Fan, X.; Yan, C.; Fan, J.; Wang, N. Improved U-Net Remote Sensing Classification Algorithm Fusing Attention and Multiscale Features. *Remote Sens.* **2022**, *14*, 3591.

34. Li, G.; Cui, J.; Han, W.; Zhang, H.; Huang, S.; Chen, H.; Ao, J. Crop type mapping using time-series Sentinel-2 imagery and U-Net in early growth periods in the Hetao irrigation district in China. *Comput. Electron. Agric.* **2022**, *203*, 107478.

35. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.C. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Adv. Neural Inf. Process. Syst.* **2015**, *2015*, 802–810.

36. Hochreiter, S.; Schmidhuber, J. Long Short-term Memory. *Neural Comput.* **1997**, *9*, 1735–80. https://doi.org/10.1162/neco.1997.9.8.1735.

37. Farooque, G.; Xiao, L.; Yang, J.; Sargano, A.B. Hyperspectral image classification via a novel spectral–spatial 3D ConvLSTM-CNN. *Remote Sens.* **2021**, *13*, 4348.

38. Cherif, E.; Hell, M.; Brandmeier, M. DeepForest: Novel Deep Learning Models for Land Use and Land Cover Classification Using Multi-Temporal and-Modal Sentinel Data of the Amazon Basin. *Remote Sens.* **2022**, *14*, 5000.

39. Meng, X.; Liu, Q.; Shao, F.; Li, S. Spatio–Temporal–Spectral Collaborative Learning for Spatio–Temporal Fusion with Land Cover Changes. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5704116.

40. Habiboullah, A.; Louly, M.A. Soil Moisture Prediction Using NDVI and NSMI Satellite Data: ViT-Based Models and ConvLSTM-Based Model. *SN Comput. Sci.* **2023**, *4*, 140.

41. Park, S.; Im, J.; Han, D.; Rhee, J. Short-term forecasting of satellite-based drought indices using their temporal patterns and numerical model output. *Remote Sens.* **2020**, *12*, 3499.

42. Yeom, J.M.; Deo, R.C.; Adamowski, J.F.; Park, S.; Lee, C.S. Spatial mapping of short-term solar radiation prediction incorporating geostationary satellite images coupled with deep convolutional LSTM networks for South Korea. *Environ. Res. Lett.* **2020**, *15*, 094025.

43. Muthukumar, P.; Cocom, E.; Nagrecha, K.; Comer, D.; Burga, I.; Taub, J.; Calvert, C.F.; Holm, J.; Pourhomayoun, M. Predicting PM2. 5 atmospheric air pollution using deep learning with meteorological data and ground-based observations and remote-sensing satellite big data. *Air Qual. Atmos. Health* **2021**, *15*, 1221–1234.

44. Yaramasu, R.; Bandaru, V.; Pnvr, K. Pre-season crop type mapping using deep neural networks. *Comput. Electron. Agric.* **2020**, *176*, 105664.

45. Chang, Y.L.; Tan, T.H.; Chen, T.H.; Chuah, J.H.; Chang, L.; Wu, M.C.; Tatini, N.B.; Ma, S.C.; Alkhaleefah, M. Spatial-temporal neural network for rice field classification from SAR images. *Remote Sens.* **2022**, *14*, 1929.

46. Ienco, D.; Interdonato, R.; Gaetano, R.; Minh, D.H.T. Combining Sentinel-1 and Sentinel-2 Satellite Image Time Series for land cover mapping via a multi-source deep learning architecture. *ISPRS J. Photogramm. Remote Sens.* **2019**, *158*, 11–22.

47. Turkoglu, M.O.; D'Aronco, S.; Wegner, J.D.; Schindler, K. Gating revisited: Deep multi-layer rnns that can be trained. *arXiv* **2019**. arXiv:1911.11033.

48. Pelletier, C.; Webb, G.I.; Petitjean, F. Temporal convolutional neural network for the classification of satellite image time series. *Remote Sens.* **2019**, *11*, 523.

49. Mitra, P.; Akhiyarov, D.; Araya-Polo, M.; Byrd, D. Machine Learning-based Anomaly Detection with Magnetic Data. *Preprints.org* **2020**. https://doi.org/10.20944/preprints202012.0092.v1.

50. Sontowski, S.; Lawrence, N.; Deka, D.; Gupta, M. Detecting Anomalies using Overlapping Electrical Measurements in Smart Power Grids. In Proceedings of the 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 15–18 December 2021, pp. 2434–2441.

51. Wagner, N.; Antoine, V.; Koko, J.; Mialon, M.M.; Lardy, R.; Veissier, I. Comparison of machine learning methods to detect anomalies in the activity of dairy cows. In Proceedings of the International Symposium on Methodologies for Intelligent Systems, Graz, Austria, 23–25 September 2020; pp. 342–351.

52. Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27.

53. Ballard, D.H. Modular learning in neural networks. In Proceedings of the AAAI, Seattle, WA, USA, 13 July 1987; Volume 647, pp. 279–284.

54. Petitjean, F.; Ketterlin, A.; Gançarski, P. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognit.* **2011**, *44*, 678–693.

55. Avolio, C.; Tricomi, A.; Zavagli, M.; De Vendictis, L.; Volpe, F.; Costantini, M. Automatic Detection of Anomalous Time Trends from Satellite Image Series to Support Agricultural Monitoring. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 6524–6527.

56. Huang, S.; Tang, L.; Hupy, J.P.; Wang, Y.; Shao, G. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *J. For. Res.* **2021**, *32*, 1–6.

57. Castillo-Villamor, L.; Hardy, A.; Bunting, P.; Llanos-Peralta, W.; Zamora, M.; Rodriguez, Y.; Gomez-Latorre, D.A. The Earth Observation-based Anomaly Detection (EOAD) system: A simple, scalable approach to mapping in-field and farm-scale anomalies using widely available satellite imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *104*, 102535.

58. Komisarenko, V.; Voormansik, K.; Elshawi, R.; Sakr, S. Exploiting time series of Sentinel-1 and Sentinel-2 to detect grassland mowing events using deep learning with reject region. *Sci. Rep.* **2022**, *12*, 983.

59. Cheng, W.; Ma, T.; Wang, X.; Wang, G. Anomaly Detection for Internet of Things Time Series Data Using Generative Adversarial Networks With Attention Mechanism in Smart Agriculture. *Front. Plant Sci.* **2022**, *13*, 890563.

60. Cui, L.; Zhang, Q.; Shi, Y.; Yang, L.; Wang, Y.; Wang, J.; Bai, C. A method for satellite time series anomaly detection based on fast-DTW and improved-KNN. *Chin. J. Aeronaut.* **2022**, *36*, 149–159.

61. Diab, D.M.; AsSadhan, B.; Binsalleeh, H.; Lambotharan, S.; Kyriakopoulos, K.G.; Ghafir, I. Anomaly detection using dynamic time warping. In Proceedings of the 2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), New York, NY, USA,1–3 August 2019; pp. 193–198.

62. Di Martino, T.; Guinvarc'h, R.; Thirion-Lefevre, L.; Colin, E. FARMSAR: Fixing AgRicultural Mislabels Using Sentinel-1 Time Series and AutoencodeRs. *Remote Sens.* **2022**, *15*, 35.

63. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man, Cybern.* **1979**, *9*, 62–66.

64. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* **2012**, *120*, 25–36.

65. PaperdJuly, W. ESRI shapefile technical description. *Comput. Stat.* **1998**, *16*, 370–371.
66. Yutzler, J. OGC® GeoPackage Encoding Standard-with Corrigendum, Version 1.2. 175. 2018. Available online: https://www.geopackage.org/spec121/ (accessed on 4 June 2023).
67. Zeiler, M. *Modeling Our World: The ESRI Guide to Geodatabase Design*; ESRI, Inc.: Redlands, CA, USA, 1999; Volume 40.
68. Butler, H.; Daly, M.; Doyle, A.; Gillies, S.; Hagen, S.; Schaub, T. *The Geojson Format*; Technical Report; 2016. Available online: https://www.rfc-editor.org/rfc/rfc7946 (accessed on 4 June 2023).
69. Moyroud, N.; Portet, F. Introduction to QGIS. *QGIS Generic Tools* **2018**, *1*, 1–17.
70. Vohra, D.; Vohra, D. Apache parquet. In *Practical Hadoop Ecosystem: A Definitive Guide to Hadoop-Related Frameworks and Tools*; Apress, Berkley, CA, USA, 2016; pp. 325–335.
71. Trakas, A.; McKee, L. OGC standards and the space community—Processes, application and value. In Proceedings of the 2011 2nd International Conference on Space Technology, Athens, Greece, 15–17 September 2011; pp. 1–5. https://doi.org/10.1109/ICSpT.2011.6064683.
72. Durbin, C.; Quinn, P.; Shum, D. *Task 51-Cloud-Optimized Format Study*; Technical Report; NTRS: Chicago, IL, USA, 2020.
73. Sanchez, A.H.; Picoli, M.C.A.; Camara, G.; Andrade, P.R.; Chaves, M.E.D.; Lechler, S.; Soares, A.R.; Marujo, R.F.; Simões, R.E.O.; Ferreira, K.R.; et al. Comparison of Cloud cover detection algorithms on sentinel–2 images of the amazon tropical forest. *Remote Sens.* **2020**, *12*, 1284.
74. AgrarMarkt Austria InVeKoS Strikes Austria. Available online: https://www.data.gv.at/ (accessed on 9 February 2023).
75. Department of Agriculture and Fisheries Flemish Government. Available online: https://data.gov.be/en (accessed on 9 February 2023).
76. Government of Catalonia Department of Agriculture Livestock Fisheries and Food. Available online: https://analisi.transparenciacatalunya.cat (accessed on 9 February 2023).
77. The Danish Agency for Agriculture. Available online: https://lbst.dk/landbrug/ (accessed on 9 February 2023).
78. Netherlands Enterprise Agency. Available online: https://nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/home (accessed on 9 February 2023).