The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-M-7-2025
44th EARSeL Symposium, 26–29 May 2025, Prague, Czech Republic

SenForFlood: A New Global Dataset for Flooded Area Detection

Bruno Menini Matosak[1], Getachew Workineh Gella[1], Stefan Lang[1]

[1] Christian Doppler laboratory for geospatial and EO-based humanitarian technologies, Department of Geoinformatics, University of Salzburg, Schillerstrasse 30, 5020 Salzburg, Austria

Keywords: Dataset, Flood Mapping, Multi-Component Samples, Change Detection

Abstract

Floods are devastating hazards that cause human displacement, loss of life and damage of properties. Getting accurate information about the extent and severity of floods is essential for planning proper humanitarian emergency assistance. Though integrating Earth observation with deep learning models supports rapid information extraction, mapping floods accurately is still a challenging task, because of the necessity of extensive, representative datasets with high quality labels to train models. While there exist some datasets that focus on providing satellite imagery for flood events, these are typically limited to data either from few floods or for specific regions. Moreover, the majority of these datasets provide images captured only during the flood event, which hinders methods that rely on detecting change. Therefore, in this work, we created a global dataset for mapping flood extent (SentForFlood), including images before and during flood from Sentinel-1 and -2, terrain elevation and slope, Land Use and Land Cover (LULC), and flood masks. The samples included in each flood event were selected by analysts considering quality of flood mask and completeness of the available satellite imagery. The dataset incorporated data from over 350 distinct flood events, encompassing all continents except Antarctica. The dataset was tested by training a convolutional neural network for detecting floods without permanent water bodies and the results are discussed. We expect that the dataset will facilitate the development of robust, transferable models for automatic flood mapping, thereby contributing to the humanitarian emergency response in crisis situations. Dataset download instructions, as well as code for easy usage is available at `https://github.com/menimato/SenForFlood`.

1. Introduction

Advances in remote sensing have led to an immense availability of Earth observation (EO) imagery with a range of sensing modality, fine spatio-temporal granularity and scope. In the era of big geodata, machine learning has emerged as an important tool to efficiently extract information from Earth observation. In particular, deep learning models have been increasingly employed, often with superior accuracy. However, training such models is challenging and requires a large amount of data (Adegun et al., 2023).

In this context, open-access datasets are important tools to enable rapid model workflow prototyping, training, evaluation and inference. Open datasets help to foster collaboration and innovation in the Earth observation and machine learning communities. Applications such as flood mapping can greatly benefit from the widespread use of these datasets, as flood mapping is considered to be affected by the small data problem due to floods rarity and brevity, resulting in fewer available data and reference (Amitrano et al., 2024).

In this regard, there are a few openly available flood datasets. Sen1Floods11 (Bonafilia et al., 2020) includes data from 11 flood events and about 90% of the labels are automatically generated. Other recent flood dataset is S1S2-Water (Wieland et al., 2024), that compiles data from Sentinel-1 and -2 focusing on water bodies and some instances of seasonal floods, providing each sample aligned with the Sentinel-2 tiling system ($100\,km \times 100\,km$ images). SEN12-Flood (Rambour et al., 2020) is another high quality dataset that instead of providing images only during the flood event, provides entire image time series for these events. It includes radar and optical samples, allowing the training of robust multi-source models, but on the other hand, it includes only a few floods limited to parts of Africa, the Middle East, and Australia. An important data source is the Copernicus Emergency Management Service (CEMS) (Salamon et al., 2021), which provides flood masks via an ensemble of three independent well-evaluated methods working on Sentinel-1 image time series, but is not provided directly as a dataset for easy training of deep learning models. Other important datasets are described in (Amitrano et al., 2024), most of which provide data at a $10\,m$ spatial resolution (from Sentinel-1 and -2) and other coarser spatial resolution orbital platforms.

A common shortcoming of existing flood datasets is that they only provide data for the time during the flood, making them difficult to use for methods that compare images pre-event and during flood to detect changes in water extent or flooded surfaces. In addition, most datasets contain only a few flood events, resulting in a lack of geographic and temporal representation. This imparts high spatial autocorrelation to the samples, which can lead to overfitting and poor model spatio-temporal transferability, constraining the development of robust and scalable flood extraction models. To improve these aspects, we developed a new dataset with a wide geographic coverage and temporal span, with the inclusion of pairs of optical and SAR images before and during the flood with other thematic layers.

2. Dataset Design

The process for designing and creating the dataset will be described in the following sections.

2.1 Flood Data and Sample Selection

Flood occurrence data is produced by many different flood monitoring systems. For indicating where and when floods occurred around the globe, we used the database made available by the Dirtmouth Flood Observatory (DFO) (Brakenridge, 2024) and CEMS (Salamon et al., 2021). In the DFO database, flood events are comprised of the region affected and the approximated start and end times. The database documents flood occurrence using information gathered from news, instrumental, governmental, and remote sensing sources. On the other hand, some data in CEMS is based on user activation, where mapping requests for a specific flood event can be submitted through a dedicated system. In the construction of our dataset, we used data from flood-related activations since open, high-quality flood mask products could be obtained from previous activations.

Combining information from DFO and CEMS regarding the occurrence time and duration, geographic location and extent of flooded areas, we created a pipeline to hand-select appropriate samples to include in the dataset. First, images from Sentinel-1 and Sentinel-2 archives pre- and post-event were obtained to be shown in a Geographical Information System (GIS). Then, automatic flood masks were created based on the difference between pre- and during flood VH images. Then, a grid with the size of the desired samples were overlaid on the images and analysts were responsible to hand select which samples of the grid would be included in the final dataset, considering the quality of flood mask and availability of satellite imagery. This stage was devised in order to overcome possible inconsistencies in the flood databases used as reference for the location and time of flood, as well as data related issues, improving the overall quality of the dataset.

2.2 Data Retrieval and Dataset Characteristics

With a collection of intended sample locations and flood occurrence times, we started the procedure of data retrieval and processing. The Sentinel-1 and -2 image archives were searched for images that matched the flood events geographically and temporally. Each sample has two reference periods for the satellite imagery, one before and one during the flood. Therefore, images were then separated into two groups: before and during flood. These groups of images were then ordered by their temporal proximity to the beginning of their respective time intervals, and then mosaicked and downloaded matching the extent of each sample patch. During this process, terrain, land use and cover (LULC) and flood mask data were also obtained.

Regarding the data layers included for each sample in the dataset, Table 1 and Figure 1 depict the seven image / information layers that compose a single sample patch. Synthetic Aperture Radar (SAR) and optical images were obtained before and during the flood event. The included radar data was obtained from Sentinel-1 Ground Range Detected (GRD) images and is composed of VV and VH polarizations, as well as the ratio VV/VH. Optical data was included from Sentinel-2 archives, available in the bands 2, 3, 4, 8, 11, and 12 top-of-atmosphere reflectance (Level-1C). Due to the shortcomings often found in the Sentinel-2 cloud masks generated for Level-1C products (Skakun et al., 2022), we included the cloud mask information from Cloud Score+ (V1) (Pasquarella et al., 2023). Moreover, for radar and optical data, a band indicating the sensing date of each pixel is available as well because different dates could have been mosaicked together to compose one single sample at the border of Sentinel's acquisition strips.

As indicated in section 2.1, the location of the samples was defined manually for each flood event considering the satellite data availability and the quality of the flood mask at hand, which was created in one of the following ways:

Type	Reference Time	Bands
Sentinel-1 Sentinel-1	Before Flood During Flood	VV VH VV/VH Pixel Date
Sentinel-2 Sentinel-2	Before Flood During Flood	Band 2 Band 3 Band 4 Band 8 Band 11 Band 12 Cloud Score+ Pixel Date
Flood Mask	During Flood	Computed/CEMS
Copernicus DEM GLO-30	2010 - 2015	Elevation Terrain Slope
ESA World Cover V200	2021	LULC

Table 1. Components of a sample in the dataset.

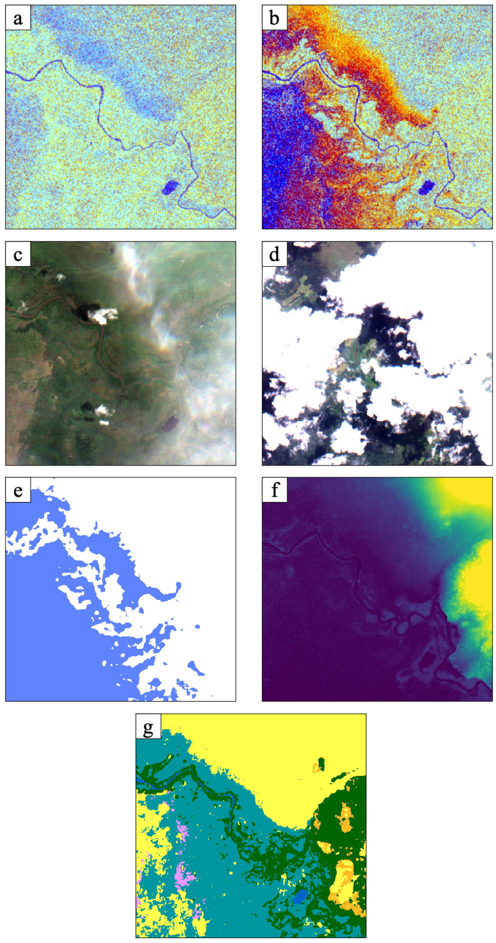

Figure 1. Images that compose a single sample in the dataset. Sentinel-1 image before (a) and during (b) flood; Sentinel-2 image before (c) and during (d) flood; Computed flood extent with band differencing (e); Terrain elevation (f); and LULC (g).

1. Automatically generated via thresholding of Sentinel-1 VH difference between images pre- and during flood, re-

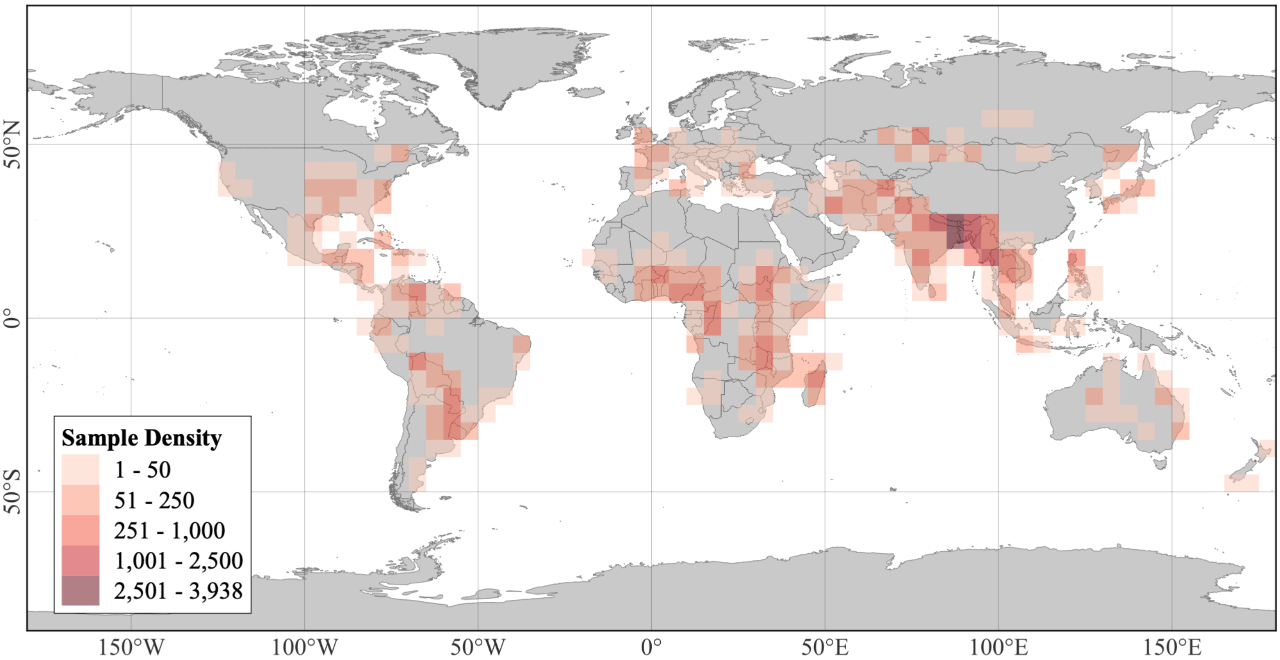

Figure 2. Density of samples (512 × 512 pixels) in the dataset per geographical location.

fined with terrain slope and permanent water bodies data; and

2. Directly from manually checked, high quality flood masks obtained from previous flood mapping CEMS activations;

Topography data from the Copernicus DEM GLO-30 (Copernicus DEM, 2022) has been included in the form of elevation and terrain slope. Moreover, LULC information obtained from ESA World Cover V2 (Zanaga et al., 2022) was added to the dataset. As its reference time is the year 2021, the land cover information may not be a highly accurate representation of LULC during the flood event. We assume that major land cover and use classes would remain stable that this layer could serve as additional contextual information to detect flood events.

Overall, the dataset created includes data for more than 350 flood events that happened from the year 2016 to 2025. It contains 40,770 georeferenced samples with dimensions of 512 × 512 pixels, that because of the seven information layers amount to a total of 285,390 image patches.

The number of samples with precise flood masks is smaller than samples with automatically generated ones. More specifically, 39,431 samples obtained from DFO documented floods have automatic flood masks, while 1,339 sample patches had their flood mask obtained directly through CEMS's flood activation archives. A map indicated in Figure 2 shows the geographic distribution and density of included samples. The samples are globally distributed across all continents where regions with frequent flood incidence exhibit dense sample inclusion. Accordingly, high density sample regions are falling in Asia that includes countries like Bangladesh, India, and Myanmar, where floods happen with a high periodicity. In South America, in comparison to other regions in the continent, the basin of the Uruguay river stands out with a high density of samples. It should be noted that the density of samples in each region has a direct relation with the floods recorded in DFO and CEMS databases, some regions that have lower density are either not

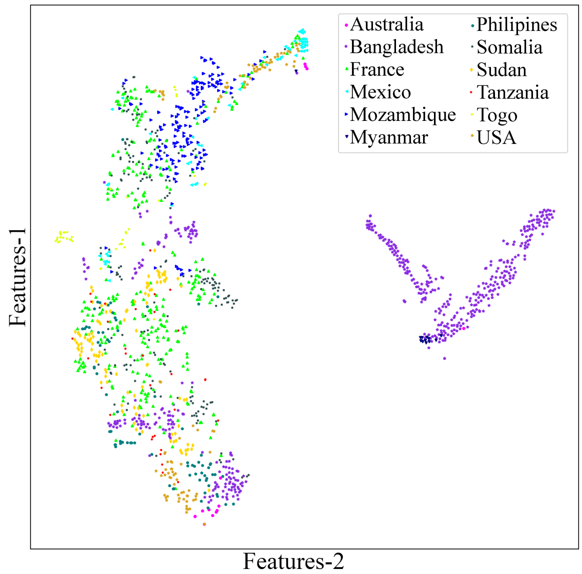

Figure 3. Latent space distribution of images from selected flood events per country. Due to the considerable size of the dataset and computational demand, we did not to include all samples in the latent space.

flood prone or have their flood events not recorded by these databases. This explains the apparent lower amount of samples in countries like China and Brazil.

Due to the variable nature of seasons and geography, there is a considerable alteration in the inherent data distribution of the dataset, akin to a domain shift. This phenomenon can be attributed to the changing responses of the target to electromagnetic radiation. This can be easily evidenced by examining the compressed latent space plot (Figure 3) of images captured before and during flood events, which are taken from selected floods where the latent representations of each one barely overlap. This limits the transferability of models trained exclusively

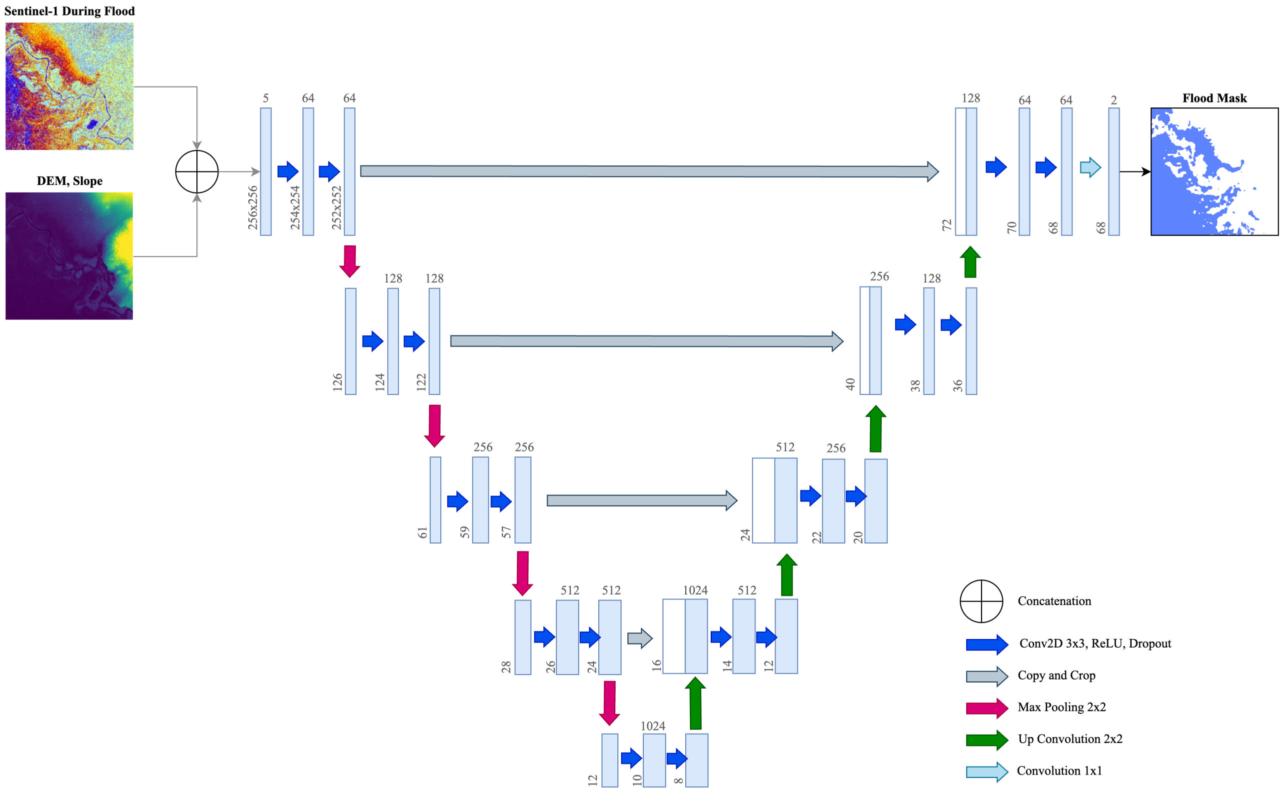

Figure 4. U-Net model architecture for the dataset usage example.

on data from a specific region to previously unseen locations. Consequently, one of the main advantages of this dataset is its covering of multi-temporal flood events globally, enabling the training of models with high transferability.

The dataset is made publicly available and further information regarding its access and usage can be found at `https://github.com/menimato/SenForFlood`. With this dataset, we aim at providing the community with an extensive collection of multi-temporal, multi-layer, multi-source and multi-modal samples for training deep learning models focused on flood mapping. To this end, the aforementioned repository also includes a dataset exploitation code based on PyTorch (Paszke et al., 2019), which facilitates the integration of the dataset into training and validation workflows. The dataset exploitation scripts allows the user to use the samples according to specific countries, as well as sub-tiling the 512×512 pixel images into smaller parts (32×32, 64×64, $128x \times 128$, and 256×256 pixels).

3. Applications

In order to test the application of our dataset, we trained a deep learning model for automatic flood mapping. More specifically, we created a U-Net model (Ronneberger et al., 2015) that takes the SAR images during flood and the terrain information as input, yielding flood masks once trained. Figure 4 details the layers of this model, as well as its inputs and output.

To train this model, we used only a part of the dataset, corresponding to all samples in South Asia. We sub-tiled the samples into 256×256 image patches and trained the model for 50 epochs, with cross-entropy loss function, Adam optimizer, and learning rate of 0.001. During training, we used a batch size of 96, from which 64 samples were used for training and 32 for validation. During training, the U-Net achieved an overall accuracy of around 85% taking the flood masks included in the dataset as reference. After the model was trained, we used it to create the flood mask for a region in Bangladesh with data never seen by the model, where an extensive flood event took place around the end of May 2024. The results are illustrated in Figure 5.

In Figure 5a, it can be seen that the flooded areas present a darker colour, making it challenging to see the difference between perennial water bodies and the non-permanent flood. However, in Figure 5b, one can notice that the model successfully avoids classifying rivers as flooded areas, even though only SAR data obtained during-flood was used. This happened because the flood masks available cover only temporary floods, due to the change-detection nature of the method used to create the sample masks, as indicated in Section 2.2. In addition, the terrain data included during training and inference also helps the model to distinguish permanent water bodies from ephemeral floods due to their effect on the topography. This application demonstrates only one potential utilisation of the dataset. It is anticipated that the incorporation of multiple data layers, particularly the before-flood and during-flood segments, will facilitate the development of more robust models.

SAR data band differencing of images acquired at distinct timestamps has been used by the remote sensing community to map flood extent (Amitrano et al., 2024), but few have taken advantage of comparing images from different dates to specifically map flood extent with machine learning on a global scale. This dataset allows training and evaluation of deep learning algorithms with a global coverage that reduces the probability of

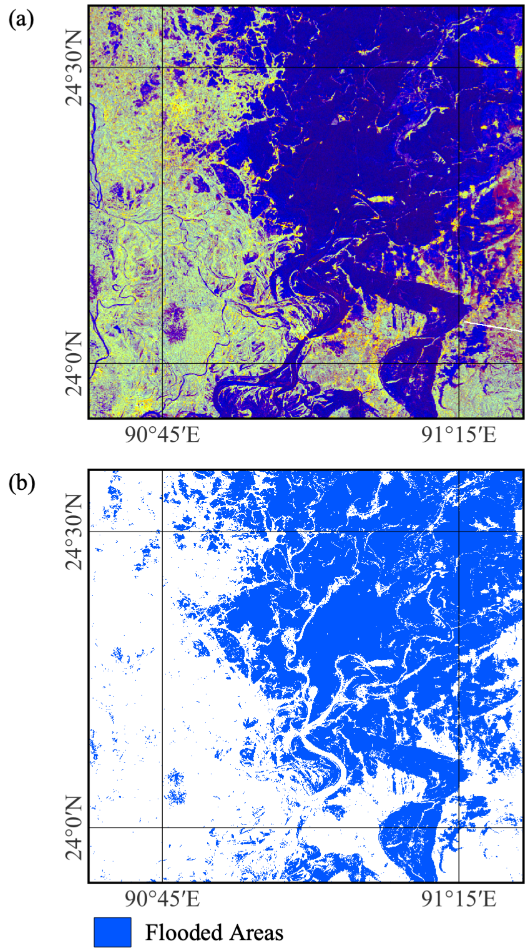

Figure 5. (a) Sentinel-1 data for a region in Bangladesh affected by the floods started at the end of May 2024. False-color composition where RGB bands are VV, VH, and VV/VH; (b) Flood mask created for the same region, using the trained U-Net model. It can be noticed that permanent water bodies such as rivers are not identified as flood, even though only one image during the event was used to generate the final map.

overfitting and improves transferability. Existing change detection methodologies based on Adversarial Networks or diffusion models can be adapted to take advantage of our dataset (Li et al., 2022, Li et al., 2023, Luo et al., 2024, Ayala et al., 2023), especially considering the flooded and non-flooded image patches of samples, radar and optical pairs, with topographic and land cover information from the same area. Furthermore, the highly accurate flood extent reference constructed from the CEMS data can be used for validation or fine-tuning.

4. Conclusion

In this work, a dataset was presented, tailored for rapid training and evaluation of robust deep learning models using images captures prior and during the flood. This dataset, which may evolve into a benchmark dataset by rigorous testing and assessment, provides SAR and optical imagery at $10\,m$ spatial resolution, topography data, LULC, and labels computed automatically or obtained directly from a highly accurate source. More than 370 flood events around the globe have been included, which we expect will help to produce models with high transferability. In total, 40,770 distinct 512×512 pixel samples were included, each with 7 different information layers (a total of 285,390 image files). In the future, the dataset will be used in a series of benchmarking flood detection experiments conducted by the authors.

5. Acknowledgements

This work was supported in part by the Christian Doppler Research Association, Doctors Without Borders-Section Austria, the Austrian Federal Ministry of Labour and Economy, and the National Foundation for Research, Technology and Development.

References

Adegun, A. A., Viriri, S., Tapamo, J.-R., 2023. Review of deep learning methods for remote sensing satellite images classification: experimental survey and comparative analysis. *Journal of Big Data*, 10(1). http://dx.doi.org/10.1186/s40537-023-00772-x.

Amitrano, D., Di Martino, G., Di Simone, A., Imperatore, P., 2024. Flood Detection with SAR: A Review of Techniques and Datasets. *Remote Sensing*, 16(4), 656. http://dx.doi.org/10.3390/rs16040656.

Ayala, C., Sesma, R., Aranda, C., Galar, M., 2023. Diffusion models for remote sensing imagery semantic segmentation. *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 5654–5657.

Bonafilia, D., Tellman, B., Anderson, T., Issenberg, E., 2020. Sen1floods11: a georeferenced dataset to train and test deep learning flood algorithms for sentinel-1. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, 835–845.

Brakenridge, G. R., 2024. Global Active Archive of Large Flood Events. *DFO - Flood Observatory, University of Colorado, USA*. http://floodobservatory.colorado.edu/ Archives/. Access in 1 October 2024.

Copernicus DEM, 2022. Copernicus DEM. *European Space Agency*. http://dx.doi.org/10.5270/ESA-c5d3d65.

Li, J., Wu, Z., Sheng, Q., Wang, B., Hu, Z., Zheng, S., Camps-Valls, G., Molinier, M., 2022. A hybrid generative adversarial network for weakly-supervised cloud detection in multispectral images. *Remote Sensing of Environment*, 280, 113197. http://dx.doi.org/10.1016/j.rse.2022.113197.

Li, Y., Shi, T., Zhang, Y., Ma, J., 2023. SPGAN-DA: Semantic-Preserved Generative Adversarial Network for Domain Adaptive Remote Sensing Image Semantic Segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1–17. http://dx.doi.org/10.1109/TGRS.2023.3313883.

Luo, Z., Pan, J., Hu, Y., Deng, L., Li, Y., Qi, C., Wang, X., 2024. RS-Dseg: semantic segmentation of high-resolution remote sensing images based on a diffusion model component with unsupervised pretraining. *Scientific Reports*, 14(1). http://dx.doi.org/10.1038/s41598-024-69022-1.

Pasquarella, V. J., Brown, C. F., Czerwinski, W., Rucklidge, W. J., 2023. Comprehensive quality assessment of optical satellite imagery using weakly supervised video learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2125–2135.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. Pytorch: An imperative style, high-performance deep learning library. H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, R. Garnett (eds), *Advances in Neural Information Processing Systems*, 32, Curran Associates, Inc.

Rambour, C., Audebert, N., Koeniguer, E., Le Saux, B., Crucianu, M., Datcu, M., 2020. FLOOD DETECTION IN TIME SERIES OF OPTICAL AND SAR IMAGES. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2020, 1343–1346. http://dx.doi.org/10.5194/isprs-archives-XLIII-B2-2020-1343-2020.

Ronneberger, O., Fischer, P., Brox, T., 2015. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Springer International Publishing, 234–241.

Salamon, P., Mctlormick, N., Reimer, C., Clarke, T., Bauer-Marschallinger, B., Wagner, W., Martinis, S., Chow, C., Bohnke, C., Matgen, P., Chini, M., Hostache, R., Molini, L., Fiori, E., Walli, A., 2021. The new, systematic global flood monitoring product of the copernicus emergency management service. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, IEEE.

Skakun, S., Wevers, J., Brockmann, C., Doxani, G., Aleksandrov, M., Batič, M., Frantz, D., Gascon, F., Gómez-Chova, L., Hagolle, O., López-Puigdollers, D., Louis, J., Lubej, M., Mateo-García, G., Osman, J., Peressutti, D., Pflug, B., Puc, J., Richter, R., Roger, J.-C., Scaramuzza, P., Vermote, E., Vesel, N., Zupanc, A., Žust, L., 2022. Cloud Mask Intercomparison eXercise (CMIX): An evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2. *Remote Sensing of Environment*, 274, 112990. http://dx.doi.org/10.1016/j.rse.2022.112990.

Wieland, M., Fichtner, F., Martinis, S., Groth, S., Krullikowski, C., Plank, S., Motagh, M., 2024. S1S2-Water: A Global Dataset for Semantic Segmentation of Water Bodies From Sentinel-1 and Sentinel-2 Satellite Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17, 1084–1099. http://dx.doi.org/10.1109/JSTARS.2023.3333969.

Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N.-E., Xu, P., Ramoino, F., Arino, O., 2022. Esa worldcover 10 m 2021 v200.