Near Real-Time Flood Mapping from Sentinel Data Using Machine Learning Techniques

Matija Gerčer [1], Tanja Grabrijan [2], Jernej Tekavec [3], Anka Lisec [4], Krištof Oštir [5]

[1] Faculty of Civil and Geodetic Engineering, University of Ljubljana, Slovenia – matija.gercer@fgg.uni-lj.si
[2] Faculty of Civil and Geodetic Engineering, University of Ljubljana, Slovenia – tanja.grabrijan@fgg.uni-lj.si
[3] Faculty of Civil and Geodetic Engineering, University of Ljubljana, Slovenia - jernej.tekavec@fgg.uni-lj.si
[4] Faculty of Civil and Geodetic Engineering, University of Ljubljana, Slovenia - anka.lisec@fgg.uni-lj.si
[5] Faculty of Civil and Geodetic Engineering, University of Ljubljana, Slovenia – kristof.ostir@fgg.uni-lj.si

Keywords: Earth observation, Flood detection, Copernicus EMS, Sentinel-1, Sentinel-2, DEM, Deep learning, CNN.

Abstract

This study presents a near-real-time flood-mapping approach that integrates satellite-based Earth observation (EO) data, digital elevation models (DEMs), and machine-learning (ML) techniques. Several publicly available flood datasets were evaluated; however, none fully met the requirements for spatial coverage, data quality, and thematic diversity needed for robust model development. To address these limitations, a dedicated training dataset was constructed using Copernicus Emergency Management Service (EMS) Rapid Mapping products, comprising 38 flood events from 2022 to 2025. A modular workflow was developed to generate ML-ready datasets from satellite imagery, including data acquisition, advanced preprocessing, flood mask generation, and image tiling. Additional steps, such as co-registration, rescaling, data fusion, and masking irrelevant regions, were implemented to ensure spatial and temporal consistency across heterogeneous inputs. The developed model demonstrates reliable performance in delineating flood extents, achieving an average IoU of 0.70 on the validation dataset. Although the system remains under active development, the results indicate strong potential for operational deployment in near-real-time flood monitoring.

1. Introduction

Floods are among the most destructive natural hazards, causing significant damage to infrastructure, agriculture, and human life. Rapid and accurate flood mapping is therefore essential for effective emergency response, enabling timely resource allocation and prioritisation of affected areas. While aerial and drone-based imagery can provide high spatial resolution, their applicability is often constrained by weather conditions, limited spatial coverage, and logistical challenges. In contrast, satellite-based Earth observation (EO) enables systematic, large-scale monitoring with regular revisit times and global accessibility, making it a key data source for near-real-time flood detection.

The Sentinel-1 and Sentinel-2 missions, part of the European Union’s Copernicus programme, provide complementary EO data for flood mapping. Sentinel-1 Synthetic Aperture Radar (SAR) enables all-weather, day-and-night imaging, ensuring reliable observations even under cloud cover and during extreme weather events. Sentinel-2 optical imagery provides detailed surface information under clear-sky conditions. When combined with digital elevation models (DEMs), these data sources can significantly improve the spatial and temporal accuracy of flood-extent mapping.

Despite these advances, publicly available flood datasets often remain insufficient for training robust machine learning (ML) models due to limitations in spatial coverage, data quality, and event diversity. To overcome these limitations, this study constructs a dedicated dataset using Copernicus Emergency Management Service Rapid Mapping (Copernicus EMS, 2025) products and develops a modular workflow to prepare ML-ready data from satellite imagery. The workflow includes data acquisition, preprocessing, flood mask generation, and image tiling, ensuring consistency across heterogeneous inputs.

In this study, flood detection is addressed by analysing changes in surface water extent using multi-temporal satellite observations acquired before and after a flood event. The problem is formulated as a binary water segmentation task, where models are trained to detect water bodies in individual images. The approach first detects water bodies in individual satellite images, after which flood extent is derived in a post-processing step by comparing pre-flood and post-flood observations. Sentinel-1 SAR data were used as the primary input, with DEM data incorporated as an additional explanatory variable, as Sentinel-2 imagery proved less suitable due to limited data availability and frequent cloud coverage. The constructed dataset was used to train and evaluate initial ML models, demonstrating its suitability for flood detection tasks. Convolutional neural networks (CNNs) were used as the primary modelling approach, with additional experiments involving architectures designed to capture temporal dynamics.

The developed approach is designed with operational applicability in mind, and initial prototypes have already been integrated into the ESA Charter Mapper platform (Arcorace, 2025), enabling scalable, near-real-time flood mapping for disaster response.

2. Related Work

2.1 Flood Mapping Datasets

A structured review of publicly available flood mapping datasets was conducted as an initial step, as the quality of reference labels and the representativeness of flood events strongly influence the generalisation capability of ML models. The reviewed datasets included ETCI-2021 (NASA Interagency Implementation and Advanced Concepts Team, 2021), FloodNet (Rahnemoonfar et al., 2021, WorldFloods MMFLOOD (Montello, 2022), OMBRIA (Drakonakis et al., 2022), S1GFloods (Saleh et al., 2023), SEN12-FLOOD (Rambour et al., 2020), Sen1Floods11 (Bonafilia et al., 2020), ML4Floods (ML4Floods, 2023), Floodbase (Euro Data Cube, 2020), and the CEMS C2S-MS Floods dataset (ECMWF, 2023). These resources differ substantially in sensor modality, event definition, spatial resolution, and suitability for supervised learning, particularly for near-real-time flood mapping. A summary of their main characteristics and suitability for this study is provided in Table 1.

Several datasets were found to be unsuitable for the target use case due to limitations in data quality, consistency, and applicability for ML. FloodNet relies on UAV imagery, which limits transferability to medium-resolution satellite observations. WorldFloods MMFLOOD contains a very small proportion of flooded pixels per scene, resulting in severe class imbalance that hampers effective model training. ML4Floods is valuable for optical data experiments but lacks radar observations, reducing its applicability under cloudy conditions. OMBRIA, SEN12-FLOOD, and CEMS C2S-MS Floods combine optical and SAR data, introducing heterogeneity that complicates the development of a consistent Sentinel-1-based operational workflow. S1GFloods provides Sentinel-1 data, but with insufficient image quality and consistency for reliable model training. In addition, some datasets lack clearly defined training, validation, and test splits or sufficient documentation, further limiting their reproducibility and practical usability. Sen1Floods11 represents the closest match; however, its spatial coverage and event diversity remain limited for robust model generalisation.

Dataset	Main characteristics	Suitability
ETCI-2021	Public challenge dataset; limited transparency on processing and split definition.	Limited use
FloodNet	High-resolution UAV imagery for post-flood scene understanding.	Not suitable
WorldFloods MMFLOOD	Sentinel-based flood dataset with very low flooded-pixel ratios.	Not suitable
OMBRIA	Optical and SAR data with a strong methodological design but limited event coverage.	Partly suitable
S1GFloods	Sentinel-1 flood imagery of lower quality for the present workflow.	Not suitable
SEN12-FLOOD	Combined optical and radar observations for flood detection.	Partly suitable
Sen1Floods11	The Sentinel-1 dataset focused on flood segmentation.	Partly suitable
ML4Floods	Well-designed framework centred on Sentinel-2 optical imagery.	Not suitable
Floodbase	Global flood-event catalogue rather than a model-ready supervised dataset.	Not suitable
CEMS C2S-MS Floods	Open Copernicus flood data with mixed sensor inputs.	Partly suitable

Table 1. Summary of reviewed flood datasets and their suitability for this study

Based on the reviewed literature and available resources, existing flood mapping datasets exhibit several limitations that restrict their applicability for ML-based flood detection. In general, these datasets often lack sufficient diversity in flood events, consistent data quality, or adequate geographic coverage, and may not align fully with workflows based on openly available EO data. In this context, the Copernicus Emergency Management Service (EMS) Rapid Mapping represents a particularly relevant data source, especially in scenarios where openly accessible EO data is required. EMS provides event-based flood mapping products with global coverage, regularly updated activations, and standardised geospatial outputs. (Copernicus EMS, 2025) The datasets include detailed metadata and delineation layers derived from expert interpretation, which are commonly used as reference data in flood mapping applications. A further advantage of the EMS Rapid Mapping archive is its continuous expansion through newly activated events, which are regularly added as disasters occur worldwide. This results in a growing collection of flood cases with associated delineation products, providing an increasingly comprehensive source of labelled data.

2.2 ML Approaches for Flood Mapping

While this study primarily focuses on developing a complete workflow for data acquisition and constructing an ML-ready dataset, the resulting dataset was also evaluated using ML models to assess its applicability for flood detection tasks. This section, therefore, provides only a brief overview of relevant ML approaches, while a more detailed analysis of model architectures and performance will be presented in future research.

Flood mapping approaches can generally be framed either as direct change detection between pre-event and post-event observations or as water body detection followed by temporal comparison (Zhang et al., 2022). While change detection methods explicitly model differences between multi-temporal inputs, water segmentation approaches focus on detecting surface water in individual images and subsequently deriving flood extent through comparison. In both cases, multi-temporal analysis remains a key component for distinguishing permanent water bodies from newly flooded areas.

In recent years, methods based on EO data have evolved from traditional threshold-based image analysis towards ML and deep learning (DL) approaches (Ghosh et al., 2024). Among these, U-Net and related encoder–decoder architectures are widely used due to their ability to balance localisation accuracy and computational efficiency in pixel-wise segmentation tasks (Ghosh et al., 2024). More advanced architectures, such as DeepLab variants, further improve performance by incorporating multi-scale contextual information through techniques such as atrous convolutions and spatial pyramid pooling, which are particularly beneficial for detecting fragmented flood patterns and narrow inundation features (Zhang et al., 2023).

The literature also highlights the importance of multi-source data integration. SAR data provide robust observations under cloudy and low-visibility conditions, while optical imagery can enhance class separability when clear-sky data are available (Slagter et al., 2020). In addition, ancillary data such as DEMs are highly relevant in flood mapping, as topographic characteristics strongly influence flood propagation and water accumulation.

These observations guided the design choices in this study. The problem was formulated as a binary water body segmentation task, where flood extent is derived from multi-temporal observations. This approach simplifies model training, enables consistent use of pre-event and post-event data, and aligns naturally with the requirements of an operational flood mapping workflow.

3. Data and Methods

3.1 Data Acquisition

The dataset was constructed using flood events from the EMS Rapid Mapping component, which provides on-demand mapping products for disaster response. Within the Copernicus ecosystem, participating countries can request rapid mapping services during major natural disasters. Each activation corresponds to a specific event (e.g. a flood) and typically includes multiple Areas of Interest (AOIs), representing spatially distributed impacts within a single event and requiring consistent handling during dataset construction.

The EMS Rapid Mapping portal offers structured access to event metadata and geospatial products. For each activation, detailed information is available, including event location, acquisition dates, geographic extent, disaster type, and mapping status. In addition, the service provides vector layers delineating flood extent, hydrography, infrastructure, and affected settlements. These layers are generated through semi-automatic processing and expert interpretation. The portal also includes references to the EO data used for mapping, such as acquisition timestamps and sensor types, although the original satellite imagery is not directly distributed through the EMS platform.

The training dataset was assembled from EMS Rapid Mapping activations between 1 January 2022 and 1 January 2025. In total, 75 flood-related activations were reviewed. After filtering for data quality, availability of corresponding satellite imagery, and compatibility with open Sentinel data sources, 38 flood events were retained. These events comprised 163 AOIs, of which 106 were in Europe and 57 outside Europe, providing both regional relevance and broader geographic diversity.

Data acquisition was carried out in two phases. The first phase involved retrieving EMS metadata and associated vector layers describing flood-affected areas. The second phase focused on acquiring the corresponding EO data and elevation information based on the previously extracted metadata.

The corresponding EO data were primarily retrieved automatically from the Copernicus Data Space Ecosystem (https://dataspace.copernicus.eu/) using API-based queries derived from the extracted metadata, targeting Sentinel products via their SAFE naming. However, as the exact EO products used in the original EMS delineations were not always explicitly provided, a partial manual harmonisation step was required in some cases. Descriptive references from the activation materials were matched to the corresponding acquisitions using acquisition date, time, and sensor type, ensuring temporal and spatial consistency between the reference flood maps and the underlying observations.

In addition to raw Sentinel data, Terradue services within the mCube (Terradue, 2025) environment provided DEM data. The resulting archive comprised 102 Sentinel-1 acquisitions and 29 Sentinel-2 acquisitions, together with the corresponding DEM layers. Excluding raw imagery, the curated working archive amounted to approximately 250 GB.

3.2 Data Preparation

The training dataset was prepared to closely reflect the conditions under which the model is expected to be applied. To achieve this, the preprocessing steps and data organisation were designed to mimic the intended prediction setting. Although both Sentinel-1 and Sentinel-2 data were initially considered, the number of suitable Sentinel-2 acquisitions was limited, and their quality was often affected by cloud coverage. Therefore, the dataset preparation relied exclusively on Sentinel-1 imagery, which provided more consistent coverage across the selected flood events.

Data alignment and multi-source fusion: Sentinel-1 imagery was pre-processed following an established processing chain (Arcorace, 2023). As the final objective of this work was to integrate the model into the ESA Charter Mapper platform, this preprocessing approach was adopted to ensure consistency with the operational environment. Alternative preprocessing strategies may introduce variations in the input data and potentially affect model performance. The processing included orbit refinement, border noise removal, radiometric calibration, terrain correction, and conversion to sigma nought (dB) (Arcorace, 2023). VV and VH polarisations were used as the primary input for flood mapping.

In addition to SAR imagery, DEM provided by Terradue within the mCube (Terradue, 2025) were also used. As these datasets were available in different coordinate reference systems and resolutions, they were reprojected to match the Sentinel-1 spatial grid using bilinear resampling. The input DEM had a spatial resolution of 30m and was resampled to match the 10m Sentinel-1 spatial resolution. This ensured full spatial alignment between the DEM and radar data, allowing their direct use in subsequent processing. All input data were then clipped to the AOI boundaries defined in the EMS activations. Sentinel-1 scenes typically cover areas far beyond the regions of interest, and clipping reduced the data volume while retaining only relevant regions. This step was applied consistently to Sentinel-1 imagery, DEMs, and flood extent layers, ensuring precise spatial alignment within each AOI.

Image tiling and flood mask generation - In this phase, vector layers from EMS activations were converted into a raster mask suitable for ML. Hydrography and flood extent polygons were merged and rasterised into a binary water body mask, where 0 represents non-water and 1 represents water. The rasterised masks were aligned with the Sentinel-1 grid to ensure pixel-level correspondence between the input imagery and ground truth.

Given the large spatial extent of Sentinel-1 scenes, the data were divided into smaller image tiles to improve ML computational efficiency and model performance. Working with full scenes would be computationally demanding and introduce significant variability. Therefore, each AOI was partitioned into fixed-size tiles (e.g., 512 × 512 pixels) with corresponding water-body masks. To maintain uniform tile dimensions, edge regions were handled by introducing overlaps, ensuring that no incomplete tiles were generated.

The tiling procedure was applied consistently across all data layers, including Sentinel-1 VV and VH polarisations, the aligned DEM, and the rasterised masks (Figure 1). This ensured that each tile contained a spatially aligned set of input features and corresponding labels, forming complete training samples for the ML pipeline.

(1) (2) (3)

Figure 1. Example of tiles consisting of Sentinel-1 VV (1), reprojected DEM (2), and the corresponding rasterised water body mask (3).

https://doi.org/10.5194/isprs-archives-XLIX-B3-2026-797-2026

Training dataset in numbers - From 38 EMS flood events covering 163 AOIs, a total of 5.942 aligned tiles were generated. Each tile comprised multi-source inputs, including Sentinel-1 VV and VH polarisations and DEM data, along with the corresponding binary water mask.

As a large proportion of tiles contained little or no flooding, a filtering step was applied to reduce class imbalance. In total, 2.523 tiles contained no flooded pixels, while an additional 1,953 tiles had less than 2% flooded area, contributing limited information for model training. Including such samples would introduce redundancy and bias the dataset towards the majority non-flood class.

Threshold (%)	Remaining tiles	Flood pixel ratio
1.0	1,911 (55.9%)	7.41%
2.0	1,466 (42.9%)	9.20%
5.0	834 (24.4%)	13.73%
10.0	423 (12.4%)	20.19%

Table 2. Effect of flooded-pixel thresholding on the final learning set.

To address this, a threshold of 2% flooded pixels per tile was introduced as a compromise between dataset size and class balance. After filtering, 1.466 tiles were retained for model development, representing 42.9% of the original dataset. This selection increased the proportion of flooded pixels to approximately 9.2%, compared to a significantly lower ratio in the unfiltered data. As shown in Table 2, higher thresholds further improve class balance, but at the cost of substantially reducing the number of training samples. The selected threshold, therefore, represents a trade-off between representativeness and dataset size.

3.3 ML Techniques

Flood detection was formulated as a change detection problem, where flood extent is derived by comparing EO data acquired before and after an event (Figure 2). Instead of predicting flood extent directly, the models perform per-pixel water segmentation for each scene, reducing the task to a water-versus-non-water classification. Flooded areas are then identified through temporal differencing in a post-processing step.

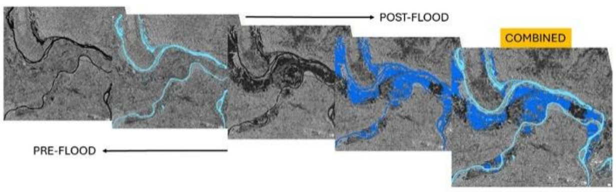

Figure2. Workflow overview - Water bodies are detected on pre-event and post-event images, and the differences between them are subtracted to identify newly flooded areas

The modelling approach is based on the dataset described in the previous section, consisting of Sentinel-1 SAR imagery (VV and VH polarisations) and corresponding binary water masks. During training, the models learn to map SAR inputs to water masks representing flooded areas. A DEM was incorporated in a later stage of post-processing to refine the results.

Architecture Selection - CNNs were adopted as the primary modelling approach, with U-Net selected as the baseline architecture due to its strong performance in EO segmentation tasks and computational efficiency. To better capture the heterogeneous spatial patterns of flood events, additional encoder–decoder architectures were evaluated, particularly DeepLabV3+ with modern backbones and atrous spatial pyramid pooling (ASPP). These models improve multi-scale context aggregation and are better suited for delineating complex or fragmented flood extents. However, this increased representational capacity comes at the cost of higher computational requirements and sensitivity to hyperparameter settings. Additional experiments with transformer-based and recurrent architectures were conducted; however, CNN-based models remained the primary approach due to their stability under the available data volume and operational constraints.

Training Strategy and Feature Integration - The modelling pipeline was primarily based on Sentinel-1 SAR inputs, while DEM data were incorporated later in the workflow to provide additional topographic context. DEM layers were aligned with the Sentinel-1 grid through reprojection and resampling, enabling their use in post-processing to improve the physical plausibility of the results and reduce false detections in topographically unlikely areas. Training was performed on tiles of size 512 × 512 pixels derived from the prepared dataset. To mitigate class imbalance, a 2% flooded-pixel threshold was applied, yielding a final dataset of 1,466 samples. The data were split into training (70%), validation (15%), and test (15%) subsets without overlap. The training objective combined Binary Cross-Entropy (BCE) and Dice loss to balance per-pixel accuracy and spatial overlap. Model optimisation was performed using AdamW for U-Net and SGD with momentum for DeepLab-based models, together with learning rate scheduling strategies. Early stopping based on validation IoU was applied to prevent overfitting, and the best-performing model was retained for inference.

3.4 Post-processing

Differences between predicted water body masks acquired at different time steps form the basis for flood detection. By comparing masks from pre-event and post-event observations, it is possible to distinguish between baseline water conditions and newly inundated areas. When multiple acquisitions are available, this approach also allows the identification of different flood stages. The choice of a suitable reference date that represents non-flood conditions is therefore critical.

By combining information across timestamps, two primary classes are derived: (i) permanent water bodies corresponding to the normal hydrological state, and (ii) newly detected water occurring outside these extents, indicating flooding. After inference, predicted water masks are converted into flood maps by differencing post-event and pre-event water detections.

Several post-processing operations are applied to suppress noise and produce cleaner cartographic outputs. Small, isolated components are removed, small holes are filled, median filtering reduces pixel-scale irregularities, and morphological opening and closing smooth object boundaries. A stricter filter for connected-component size is then applied before vector export.

An additional refinement step can be performed using terrain information through the Height Above Nearest Drainage (HAND) concept (Rennó, 2008). HAND represents the vertical distance between a pixel and its nearest drainage line and is therefore more hydrologically meaningful than absolute elevation alone. In FloodsDL, HAND is computed automatically from the Copernicus DEM and a derived drainage network. By default, only flood detections below a 5 m HAND threshold are retained when this option is enabled. This helps suppress implausibly high-elevation detections and improves floodplain delineation.

4. Results

4.1 Dataset

The final curated dataset is one of the study's main outcomes. Compared with commonly reused benchmark sets, it is event-oriented, based on a consistent EMS reference workflow, and readily extensible through future rapid-mapping activations. Its geographic distribution across Europe and non-European regions supports better transferability than a narrowly regional archive, while the metadata harmonisation enables reproducibility and traceability back to the original Copernicus products. In addition, the entire data processing workflow is designed to be reproducible and updatable, allowing the dataset to be continuously extended as new flood events occur. This ensures that the dataset can evolve over time, incorporating future rapid-mapping activations and progressively improving its representativeness and utility. Furthermore, several processing parameters (e.g., tile size) are defined dynamically, enabling the dataset to be adapted to different ML requirements and experimental setups.

From an ML perspective, the filtering strategy proved essential. Without removing near-empty tiles, the learning task would have been dominated by the non-water class. The 2% threshold retained 42.9% of candidate tiles while increasing the flood-pixel ratio from a highly skewed distribution to a still challenging, but more informative, 9.2%. This compromise preserved enough diversity for learning while reducing redundancy.

4.2 ML Results

The models presented in this study constitute an initial baseline for assessing the applicability of the constructed dataset and processing workflow. More advanced modelling approaches are currently under active development and will be addressed in future work. Model performance was evaluated using the Intersection over Union (IoU) metric, which directly quantifies the overlap between predicted flood masks and reference data. IoU is defined as the ratio of true positives to the union of true positives and false positives and false negatives, making it particularly suitable for segmentation tasks. Across the validation dataset, the models achieved an average IoU of approximately 0.70, indicating reliable segmentation performance given the heterogeneity of flood morphology and SAR backscatter conditions. Performance varied across scenes, with the best cases reaching IoU values close to 0.80, while more challenging cases dropped to approximately 0.15. Lower performance was typically observed in fragmented flood scenarios, urban environments with complex radar responses, and scenes affected by strong speckle noise or ambiguous water signatures.

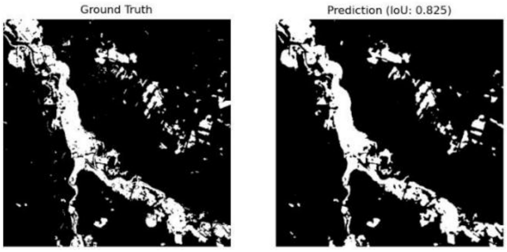

Figure 3. Best-performing prediction. The ground truth (left) and predicted mask (right) show high agreement (IoU = 0.825).

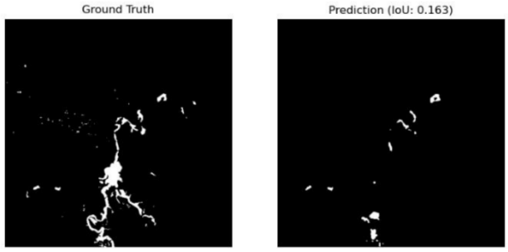

Figure 4. Worst-performing prediction. The ground truth (left) and predicted mask (right) show low agreement (IoU = 0.163).

These extremes (Figure 3 and Figure 4) highlight both the model's strong performance under well-represented conditions and its limitations in more complex or edge-case scenarios, as illustrated by the best- and worst-case examples.

Training dynamics were further analysed using the evolution of the IoU metric on the training set. As illustrated in Figure 5, the U-Net model demonstrated stable, consistent convergence, achieving a training IoU above 0.80. In contrast, DeepLabV3+ exhibited slower convergence behaviour and less stable optimisation in this experimental setup. Although DeepLabV3+ is generally considered a more expressive architecture, its performance in this study was likely constrained by the limited dataset size and sensitivity to hyperparameter selection. Consequently, U-Net provided more robust and reliable results under the given conditions.

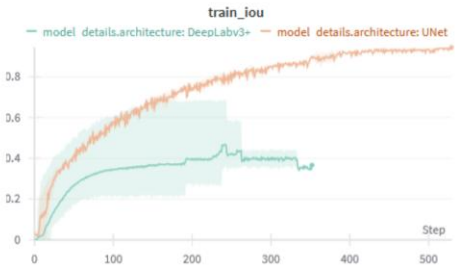

Figure 5. Training IoU curves for U-Net and DeepLabV3+

These findings are further supported by visual inspection of predicted flood masks. U-Net produced more coherent and spatially consistent segmentations, whereas DeepLabV3+ often resulted in under-segmented or fragmented masks, particularly in more complex scenes with heterogeneous backscatter or fragmented flood patterns. This behaviour is consistent with the quantitative evaluation and reflects the sensitivity of the more complex DeepLabV3+ architecture to limited training data and parameter settings. Based on these observations, U-Net was selected as the final model for further analysis. In this setup, it proved more robust to varying scene conditions, converged more reliably during training, and produced more consistent flood delineations across different AOIs.

Hyperparameter optimisation for the U-Net architecture was carried out using automated parameter sweeps to assess the impact of different training configurations on IoU performance. The analysis included parameters such as batch size, learning rate, weight decay, dropout rate, and BCE loss weighting. The results indicate that better performance is generally achieved with batch sizes around 32, low dropout rates (below 0.02), and learning rates on the order of 10^{-4}. Overall, these findings suggest that moderate batch sizes, limited regularisation, and appropriately tuned learning rates contribute to more stable and reliable training in this setup.

4.3 Post-processing Results

The post-processing stage focused primarily on reconciling tile-based predictions and deriving consistent flood extent differences across the full scene. Since the model operates on individual tiles, an important step was to aggregate these predictions and identify newly flooded areas by comparing them with pre-event water conditions.

The post-processing chain further improved the practical usability of raw model outputs by removing isolated artefacts and enforcing more hydrologically plausible shapes. The optional HAND filter was particularly valuable for rejecting spurious flood detections on local topographic highs. Although this refinement does not replace explicit hydraulic modelling, it provides a computationally efficient terrain constraint that is well aligned with near-real-time operations.

The final operational flood mask uses three classes: 0 for non-water, 1 for permanent water, and 2 for newly detected flooded areas after the event. This output design preserves permanent hydrography while making flood extent explicit, which is beneficial for downstream visualisation and integration with other emergency mapping products.

4.4 Model Integration

A major practical outcome of this study is that the developed workflow is not limited to offline experimentation but has been successfully integrated into an operational environment. The model, together with its preprocessing and post-processing chain, was encapsulated in a Docker-based microservice and exposed through a command-line interface (CLI), enabling automated execution and orchestration. The service was further described using the Common Workflow Language (CWL), enabling seamless integration with the ESA Charter Mapper and mCube platforms. The ESA Charter Mapper platform provides a cloud-based environment for rapid disaster response using EO data, where authorised users can access, process, and analyse satellite imagery within a standardised framework. In this context, the FloodsDL service was designed as a modular, interoperable component that can be invoked directly from the platform interface. To ensure reproducibility and portability, the service was deployed within a containerised environment using Docker. All dependencies, including ML and geospatial libraries, as well as the trained model, were packaged into a single execution unit. This approach eliminates dependency conflicts and allows the service to run consistently across different infrastructures. In the operational setting, the workflow is initiated by an authorised user who defines an AOI and specifies the flood event date within the platform interface. Based on these inputs, the system automatically selects a suitable post-event Sentinel-1 acquisition and retrieves corresponding pre-event imagery representing baseline hydrological conditions. The selected datasets, together with the associated DEM and AOI geometry, are then passed to the FloodsDL service. The processing chain is executed automatically, encompassing data ingestion, preprocessing, and model inference applied to both pre- and post-event observations. In cases where the AOI spans multiple Sentinel-1 scenes, processing is performed independently for each scene, after which the resulting tiles are merged and mosaicked into a single, consistent flood extent layer. The resulting water masks are subsequently combined to derive the initial flood extent. A dedicated post-processing stage is then applied to refine the results by removing isolated artefacts, improving spatial consistency, and, optionally, incorporating terrain-based constraints, such as HAND filtering. This step enhances the physical plausibility of the detected flood extent while remaining computationally efficient for near-real-time applications. The final output explicitly distinguishes between non-water areas, permanent water bodies, and newly inundated regions. The outputs are produced in standard geospatial formats, enabling direct visualisation within the ESA Charter Mapper environment and integration into downstream analysis and emergency response workflows. An overview of the processing sequence is presented in Figure 6.

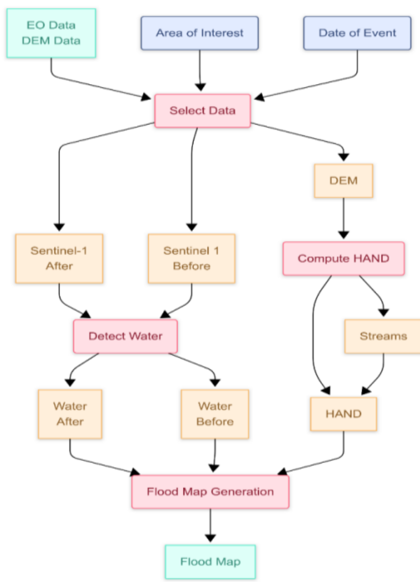

Figure 6. User-driven workflow of the FloodDL service within ESA Charter Mapper

The result of the workflow is a flood extent map (Figure 7) that clearly delineates areas of permanent water and newly flooded regions within the defined AOI. By explicitly separating pre-existing hydrography from event-driven inundation, the product provides an intuitive and operationally relevant representation of flood dynamics.

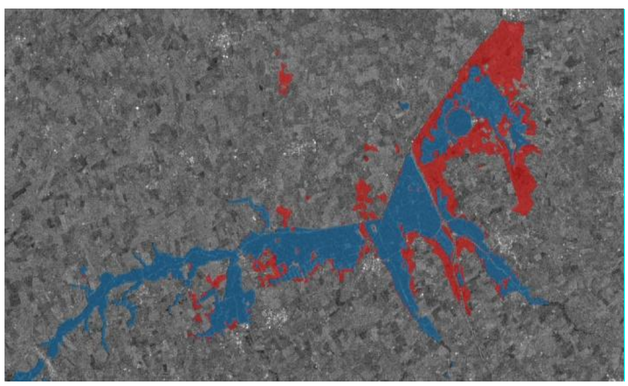

Figure 7. Example of a flood event map generated by the proposed workflow. Permanent water bodies are shown in blue, while newly flooded areas are highlighted in red.

5. Conclusion

This study presents a complete workflow for near-real-time flood mapping using Sentinel data, DEM data, and DL techniques. A key contribution is the construction of a curated, event-based dataset derived from Copernicus EMS Rapid Mapping products. The dataset is built on expert-validated delineation layers and ensures high-quality, consistent flood masks aligned with satellite observations. In addition, the dataset is inherently dynamic, and continuously extensible, as new EMS activations can be systematically incorporated through the automated data acquisition and preprocessing pipeline. This enables the archive to grow over time, improving its geographic coverage, event diversity, and overall representativeness for ML applications.

From a modelling perspective, the results show that the proposed approach enables reliable flood delineation across heterogeneous scenarios, achieving an average IoU of approximately 0.70. Among the evaluated architectures, U-Net achieved the most robust performance, demonstrating stable convergence behaviour and more consistent segmentation results than more complex models such as DeepLabV3+. This highlights the importance of selecting model complexity appropriate to the available data and operational constraints.

Beyond model performance, an important outcome of this work is the successful integration of the complete workflow into the ESA Charter Mapper and mCube platforms. By combining automated data preparation, tile-based inference, post-processing, and containerised deployment, the system supports scalable and reproducible flood mapping in a near-real-time operational setting.

At the same time, several limitations remain. The training dataset, although carefully curated, remains relatively limited in size and diversity for global generalisation, while the integration of optical Sentinel-2 data is constrained by cloud cover and data availability. In addition, uncertainty estimation has not yet been fully addressed. Future work will therefore primarily focus on further developing the dataset and improving the modelling approach. This includes expanding the event archive, enabling continuous and automated dataset updates, and preparing the dataset for potential public release. In parallel, efforts will be directed towards exploring more advanced model architectures, improving model robustness and generalisation, and enhancing uncertainty quantification. Multi-sensor data fusion will also be investigated to further improve performance under diverse environmental conditions.

Overall, the presented workflow demonstrates a practical and scalable approach to flood mapping that bridges the gap between research and real-world application, providing a solid foundation for further development of EO-based emergency response services.

Acknowledgements

The authors gratefully acknowledge financial support from the Slovenian Research Agency (core funding No. P2-0406 Earth observation and geoinformatics and GC-0006 Geospatial information technologies for resilient and sustainable society) and European Space Agency (project EO4MASRISK – Ground Deformation Detection and Risk Information Service).

References

Arcorace, M. (2023) *Multi Mission Mapper - mCube*. Terradue documentation.

Arcorace, M. (2025). *ESA Charter Mapper*. ESA Charter Mapper documentation.

Bonafilia, D., Tellman, B., Anderson, T. and Issenberg, E. (2020) ‘Sen1Floods11: a georeferenced dataset to train and test deep learning flood algorithms for Sentinel-1’, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 835–845. doi: 10.1109/CVPRW50498.2020.00113.

Copernicus Emergency Management Service (2025) *EMS Rapid Mapping.*

Drakonakis, G.I., Tsagkatakis, G., Fotiadou, K., Tsakalides, P. (2022). ‘OmbriaNet - supervised flood mapping via convolutional neural networks using multitemporal Sentinel-1 and Sentinel-2 data fusion’. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, pp. 2341–2356.

ECMWF (2023) *Open CEMS-Flood Data - Copernicus Emergency Management Service*. Available at: https://confluence.ecmwf.int/display/CEMS/Open+CEMS-Flood+Data (Accessed: 6 November 2023).

Euro Data Cube (2020) *Cloud to Street - Microsoft Flood Dataset - Sentinel-2* . Available at: https://collections.eurodatacube.com/microsoft-floods-s2/ (Accessed: 6 November 2023).

Ghosh, S., Mandal, D. and Kumar, V. (2024) ‘Deep learning approaches for SAR-based flood mapping: A comparative study’, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.

ML4Floods (2023) *The WorldFloods database* . Available at: https://spaceml-org.github.io/ml4floods/content/worldfloods_dataset.html (Accessed: 6 November 2023).

Montello, F., Arnaudo, E. and Rossi, C. (2022) ‘MMFlood: A multimodal dataset for flood delineation from satellite imagery’, *IEEE Access*, 10, pp. 96774–96787. doi: 10.1109/ACCESS.2022.3205419.

NASA Interagency Implementation and Advanced Concepts Team (2021) *ETCI 2021 Competition on Flood Detection*. Available at: https://nasa-impact.github.io/etci2021/ (Accessed: 6 November 2023).

Rahnemoonfar, M., Chowdhury, T., Sarkar, A., Varshney, D., Yari, M., Murphy, R.R., 2021. FloodNet: A High Resolution Aerial Imagery Dataset for Post Flood Scene Understanding. IEEE Access 9, 89644-89654.

Rambour, C., Audebert, N., Koeniguer, E., *et al*. (2020) ‘SEN12-FLOOD: a SAR and multispectral dataset for flood detection’. *IEEE DataPort*.

Rennó, C.D., Nobre, A.D., Cuartas, L.A., *et al.* (2008) ‘HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia’, *Remote Sensing of Environment*, 112(9), pp. 3469–3481.

Saleh, T., Weng, X., Holail, S., Hao, C. and Xia, G.-S. (2023) ‘DAM-Net: Global Flood Detection from SAR Imagery Using Differential Attention Metric-Based Vision Transformers’. Available at: https://arxiv.org/pdf/2306.00704.pdf.

Slagter, B., Schulte, S. and van der Sande, C. (2020) ‘Flood extent mapping using Sentinel-1 and Sentinel-2 data: A multi-sensor approach’, Remote Sensing, 12(15), 2461.

Terradue (2025). mCube documentation and processing environment. Available at: https://docs.mcube.terradue.com/

Zhang, X., Chen, J. and Liu, Y. (2023) ‘Multi-scale deep learning framework for flood mapping using remote sensing imagery’, Remote Sensing of Environment, 295, 113673.

Zhang, Y., Li, Y., Wang, C. and Zhang, H. (2022) ‘Deep learning-based flood mapping from multi-temporal remote sensing data: A review’, Remote Sensing, 14(5), 1235.