OmbriaNet—Supervised Flood Mapping via Convolutional Neural Networks Using Multitemporal Sentinel-1 and Sentinel-2 Data Fusion

Georgios I. Drakonakis, Grigorios Tsagkatakis, Konstantina Fotiadou,
and Panagiotis Tsakalides, *Member, IEEE*

***Abstract*—Regions around the world experience adverse climate-change-induced conditions that pose severe risks to the normal and sustainable operations of modern societies. Extreme weather events, such as floods, rising sea levels, and storms, stand as characteristic examples that impair the core services of the global ecosystem. Especially floods have a severe impact on human activities, hence, early and accurate delineation of the disaster is of top priority since it provides environmental, economic, and societal benefits and eases relief efforts. In this article, we introduce OmbriaNet, a deep neural network architecture, based on convolutional neural networks, that detects changes between permanent and flooded water areas by exploiting the temporal differences among flood events extracted by different sensors. To demonstrate the potential of the proposed approach, we generated OMBRIA, a bitemporal and multimodal satellite imagery dataset for image segmentation through supervised binary classification. It consists of a total number of 3.376 images, synthetic aperture radar imagery from Sentinel-1, and multispectral imagery from Sentinel-2, accompanied with ground-truth binary images produced from data derived by experts and provided from the Emergency Management Service of the European Space Agency Copernicus Program. The dataset covers 23 flood events around the globe, from 2017 to 2021. We collected, co-registrated and preprocessed the data in Google Earth Engine. To validate the performance of our method, we performed different benchmarking experiments on the OMBRIA dataset and we compared with several competitive state-of-the-art techniques. The experimental analysis demonstrated that the proposed formulation is able to produce high-quality flood maps, achieving a superior performance over the state-of-the-art. We provide OMBRIA dataset, as well as OmbriaNet code at: https://github.com/geodrak/OMBRIA.**

***Index Terms*—Convolutional neural network (CNN), deep learning, flood mapping, remote sensing, Sentinel-1, Sentinel-2.**

I. Introduction

FLOODS are natural disasters that have a great impact on human societies, affecting economic activity at both local and regional scales. Their main driver on such extreme events is meteorological phenomena with an increase in frequency and magnitude observed during the last decades due to climate change [1]. Communities are pushed to poverty as agriculture production output is reduced and infrastructure is damaged. Studies from insurance companies show that between 1980 and 2019, hydrological disasters caused overall losses of approximately 1 billion U.S. dollars [2].

Detecting flooded areas is a tedious task that requires human expertise and many work hours. Remote sensing data are used to delimit flood extents and integrated with GIS to produce maps [3]. Satellite systems constitute the most widely used platform for large area mapping and emergency management [4], with Coperincus Emergency Management Service being a prime example. Earth observation satellites are equipped with instruments operating in wavelengths extending from the visible to microwave range. Traditional approaches exploit the capacity of water to absorb light at certain wavelengths [5], [6]. Early and current works use indicators, such as the normalized difference water index (NDWI) [7] and its improved version [8], which are suitable for enhancing and extracting water information.

Recently, the combination of the satellite imagery infrastructure with artificial intelligence technologies has provided a new path toward successfully addressing the problem of flood detection and mapping. Traditional machine learning techniques have been employed in Earth observation (EO) data analysis, including support vector machines (SVMs) [9] and random forests [10]. convolutional neural networks (CNNs), as particular types of deep learning architectures, represent the most promising and prolific machine learning models, and have become a predominate tool due to their efficiency in learning data representations [11]. Advances in the fields of deep learning and computer vision have taken remote sensing to a new level [12] and are proven to be more accurate in tasks such as land cover classification [13]–[16] and object detection [17]–[19], outperforming the traditional methods [20]–[22]. Although most works focus on land applications, there is an increasing interest in water applications, such as water detection [23]–[25].

Manuscript received May 30, 2021; revised October 13, 2021 and January 11, 2022; accepted February 12, 2022. Date of publication March 1, 2022; date of current version March 23, 2022. This work was supported by the Marie Skłodowska-Curie project CALCHAS project under Grant 842560 within the H2020 Framework Program of the European Commission. *(Corresponding author: Georgios I. Drakonakis.)*

Georgios I. Drakonakis, Konstantina Fotiadou, and Panagiotis Tsakalides are with the Department of Computer Science, University of Crete, 700 13 Heraklion, Greece, and also with the Institute of Computer Science and Foundation of Research and Technology Hellas, University of Crete, 70013 Heraklion, Greece (e-mail: drakonakis@ics.forth.gr; kfot@ics.forth.gr; tsakalid@ics.forth.gr).

Grigorios Tsagkatakis is with the Institute of Computer Science and Foundation of Research and Technology Hellas, University of Crete, 70013 Heraklion, Greece (e-mail: greg@ics.forth.gr).

Digital Object Identifier 10.1109/JSTARS.2022.3155559

In this article, we address two challenges. The first is to compensate for the lack of ground-truth data and flooded area annotations and to provide to the scientific community a new dataset for supervised classification. Second, we propose a novel deep learning architecture for supervised segmentation that is able to detect changes in water presence using a bitemporal set of high-resolution 3-D imagery.

The main novelties of this article include the following.

1) OMBRIA: The generation of a new dataset for addressing the problem of flood mapping.
2) OmbriaNet: A novel multimodal and bitemporal CNN designed for supervised image segmentation and change detection.

The rest of this article is organized as follows. In Section II, we present the state-of-the-art on image segmentation with deep learning and flood detection in remote sensing. In Section III, the OMBRIA dataset creation is discussed. In Section IV, we present some theoretical background of deep learning and the proposed OmbriaNet network is presented. Furthermore, in Section V, the experimental results are shown. In Section VI, discussion over the analysis results and directions for future steps is made. Finally, Section VII concludes this article.

II. Related Work

A. Computer Vision and Image segmentation

There are many applications in remote sensing that require assigning a label to every pixel in an image. This classification task is addressed with semantic segmentation algorithms. Computer vision is contributing significantly in remote sensing tasks such as cloud detection [26], urban planning [27], and land cover classification [28]. During the last years, semantic segmentation algorithms with three channels or multispectral imagery have been developed employing machine learning and especially deep learning [29]. Pretrained networks like VGG-16 [30] and ResNet [31] have been used not for segmentation per se, but for scene understanding in a coarse scale classification [32]. The DeepLab network [33] improves segmentation performance and produces sharp boundaries by substituting convolution layers with atrous convolutions. Atrous filters have zeros between sample points resulting in increased filter sizes with a constant number of parameters.

Deep learning frameworks for classifying multispectral images have been explored thoroughly in the community. However the scarcity of annotated data has limited most work to unsupervised methods [34], [35]. Lack of supervised data has an impact on model generalization between different datasets. Stacked autoencoders are unsupervised neural networks that encode efficiently the training data and learn high feature representations [36], [37]. In [38], a hybrid autoencoder—multilayer perceptron is introduced to map floods in two study areas in Iran and India. The autoencoder was used to reduce the feature count, enhance the training process, and increase the performance compared to the traditional MLP.

Fusion of high-resolution imagery with digital surface models in a fully convolutional network has achieved state-of-the-art performance on a multimodal semantic segmentation scheme. Experiments on the International Society of Photogrammetry and Remote Sensing (ISPRS) Vahingen benchmark test set showed that the overall accuracy is among the top performers according to [39]. Transfer learning was used in [40] fusing a U-net based deep,[1] called TL-DenseUNet, with an encoder subnetwork transferring pretrained DenseNet to fuse multiscale information, performing multiobject semantic segmentation with an imbalanced class distribution. Experimental results showed that transfer learning is effective and achieves a better performance than other models.

B. Remote Sensing and Flood Detection

Remote sensing technologies have evolved rapidly during the recent years and their advantages in analyzing the Earth surface by its spectral properties have been utilized in environmental monitoring and emergency and disaster relief [41]. Automated methods for waterbodies segmentation with satellite imagery can be divided in two categories, namely rule-based systems and machine learning models. Thresholding a water index such as the modified normalized difference water index (MNDWI) or the multiband water index (MBWI) [8], [42] is a simple rule-based approach. Less commonly used by remote sensing community techniques that include expert systems, with visual analysis combining human cognitive abilities and evidential reasoning to deal with problems related to both uncertainties and quality issues in the dataset are used in [43]. While these methods produce accurate results, this comes under specific conditions and it lacks generalization ability and transferability. Machine learning methods can learn flood characteristics given a set of labeled samples [44]. CNNs are able to learn features from images and segment water from land, outperforming hand-crafted features, and report higher accuracy and generalization ability [45].

Synthetic aperture radar (SAR) is proven to provide more reliable information on flood extent [46]. Precipitation events have as a result long-lasting cloud coverage periods and SAR sensors' ability to penetrate clouds both day and night make them more suitable for the task. Lidar technologies can give precise digital elevation models that can be combined with SAR to estimate flood extend along with precise depth [47]. Sentinel-2 multispectral imagery and CNN's have been used in supervised classification producing good results but in limited spatial scale (national) and water segmentation for labeling was produced by visual interpretation with results relative to a human analyst and not in absolute ground truth [48].

C. Flood Mapping With Deep Learning

Machine learning models have not been utilized extensively in flood mapping problems because there is a lack of available datasets. Sen1Floods11 is the first dataset for flood detection. It was introduced to assist efforts to operationalize deep learning algorithms for flood mapping in global scale. It contains Sentinel-1 imagery and is trained with a fully CNN to perform multiple classifications and compare performance with common remote sensing algorithms like backscatter thresholding [49].

[1] Convolutional neural network (CNN).

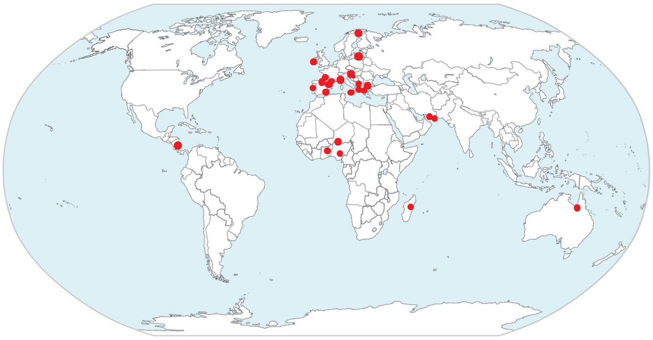


Fig. 1. Map of emergency management service activations for flood events.

TABLE I
Table With Flood Events Used in OMBRIA Dataset as Emergency Management Service Rapid Mapping Activations

EMS ID	**Country**	**Date 1**	**Date 2**	**UTM Zone**
271	Greece	01/05/2017	28/02/2018	34 N
273	Albania	01/05/2017	11/03/2018	34 N
275	Croatia	01/05/2017	22/03/2018	33 N
279	Spain	01/05/2017	15/04/2018	30 N
324	France	01/05/2018	16/10/2018	31 N
342	Australia	15/04/2018	13/02/2019	54 S
388	Spain	01/05/2019	14/09/2019	30 N
416	France	01/05/2019	15/12/2019	30 N
417	Portugal	01/05/2019	23/12/2019	29 N
419	Iran	01/05/2019	13/01/2020	41 N
422	Spain	01/05/2019	26/01/2020	31 N
424	Madagascar	01/05/2019	29/01/2020	39 S
429	Ireland	01/05/2019	23/02/2020	29 N
441	Finland	01/05/2019	04/06/2020	34 N
465	Greece	01/05/2020	20/09/2020	31 N
466	Niger	01/05/2020	27/09/2020	32 N
468	Italy	01/05/2020	10/10/2020	32 N
470	Togo	01/05/2020	17/10/2020	31 N
482	Honduras	01/05/2020	22/11/2020	17 N
492	France	01/05/2020	02/01/2021	30 N
501	Albania	01/05/2020	15/02/2021	35 N
507	Timor	01/05/2020	06/04/2021	51 S
514	Guyana	01/05/2020	06/06/2021	21 S

Deep learning architectures tailored for water segmentation are also numbered. H20-Net is the first to our knowledge to address the problem. The network learns SWIR signal synthesis in low resolution data as a domain adaptation mechanism for accurate flood segmentation. It uses as input red, green, blue, and near infrared channels in a self-supervised classification and achieves high accuracy [50]. Besides EO, unmanned aerial vehicles (UAVs) constitute another source of data that makes observation easy and more frequent. High-resolution UAV imagery has been used for postnatural disaster damage assessment using state-of-the-art deep learning algorithms [51].

III. OMBRIA Dataset

The Copernicus program [52] provides data with global coverage, high temporal frequency, and high spatial resolution. The European Space Agency (ESA) launched the Sentinel-2 mission in July 2015, putting into orbit two twin satellites with a better resolution and a more frequent acquisition, compared to previous missions, e.g., NASA's LANDSAT 8. The ESA satellites provide Copernicus Emergency Management Service (EMS) with data necessary for mapping products to support emergency activities immediately following a disaster [53]. In addition, LANDSAT 8 constellation, the longest record of global scale EO providing data since 1972 [54], is widely used in delineating flood extents with change detection approaches [55]. Active EO sensors can also be utilized for flood mapping. SAR instruments have been used in several studies to create inundation maps [56]–[58].

The Emergency Management Service of the Copernicus Programme provides data packages of delineation and grading products for emergency situations and natural disasters. These products have been produced by field experts in semiautomated procedures. For the creation of the dataset, 20 flood event activations around the globe were picked between 2017 and 2020. Vector files were converted into raster images and used as ground truth. In Fig. 1, we present a world map with all flood events used in this article. In Table I, the flood events are categorized in a chronological order from the oldest to the most recent. The table includes the EMS Activation ID, the name of the country, the corresponding geographic zone, and the two dates used for data collection. More details regarding the data collection and preproccessing is given in Section V-A.

Imagery from Sentinel-1 was acquired at Level-1 Ground Range Detected (GRD), with VV polarization (single copolarization, vertical transmit/vertical receive). Level-1, GRD products consist of focused SAR data, that has been detected, multilooked and projected to ground range using the Earth ellipsoid model WGS '84. The ellipsoid projection of the GRD products is corrected using the terrain height from the SRTM Digital Elevation Model [59]. Ground range coordinates correspond to the slant range coordinates projected onto the ellipsoid of the Earth. Additionally, pixel values represent the detected amplitude, while phase information is lost. To reduce the speckle effect, morphological filtering was applied with a 30 × 30 meters median value kernel. As a result, the resulting patches have dimensions of $256 \times 256 \times 1$.

The Sentinel-2 MSI instrument bands that were exploited are: Band 3—Green (0.560 μm), Band 8—Near Infrared (0.842 μm), and Band 11—SWIR (1.610 μm). Bands 3 and 8 have a spatial resolution of 10 m, while band 11 has a spatial resolution of 20 m. Additionally, the bottom-of-atmosphere reflectance in cartographic geometry product (i.e., Level 2 A) was selected, since it is atmospherically corrected and orthorectified with the SRTM. The resulting patches have dimensions of $256 \times 256 \times 3$.

Imagery from every flood event were divided into nonoverlapping tiles of size 256×256. This division led to 844 tiles for each timestamp (preevent and postevent) and each modality (Sentinel-1 and Sentinel-2) creating 1.688 optical and 1688 SAR images adding up to of 3.376 input image patches. Given that the ground sample distance is 10 m the total cover area is 553 km^2.

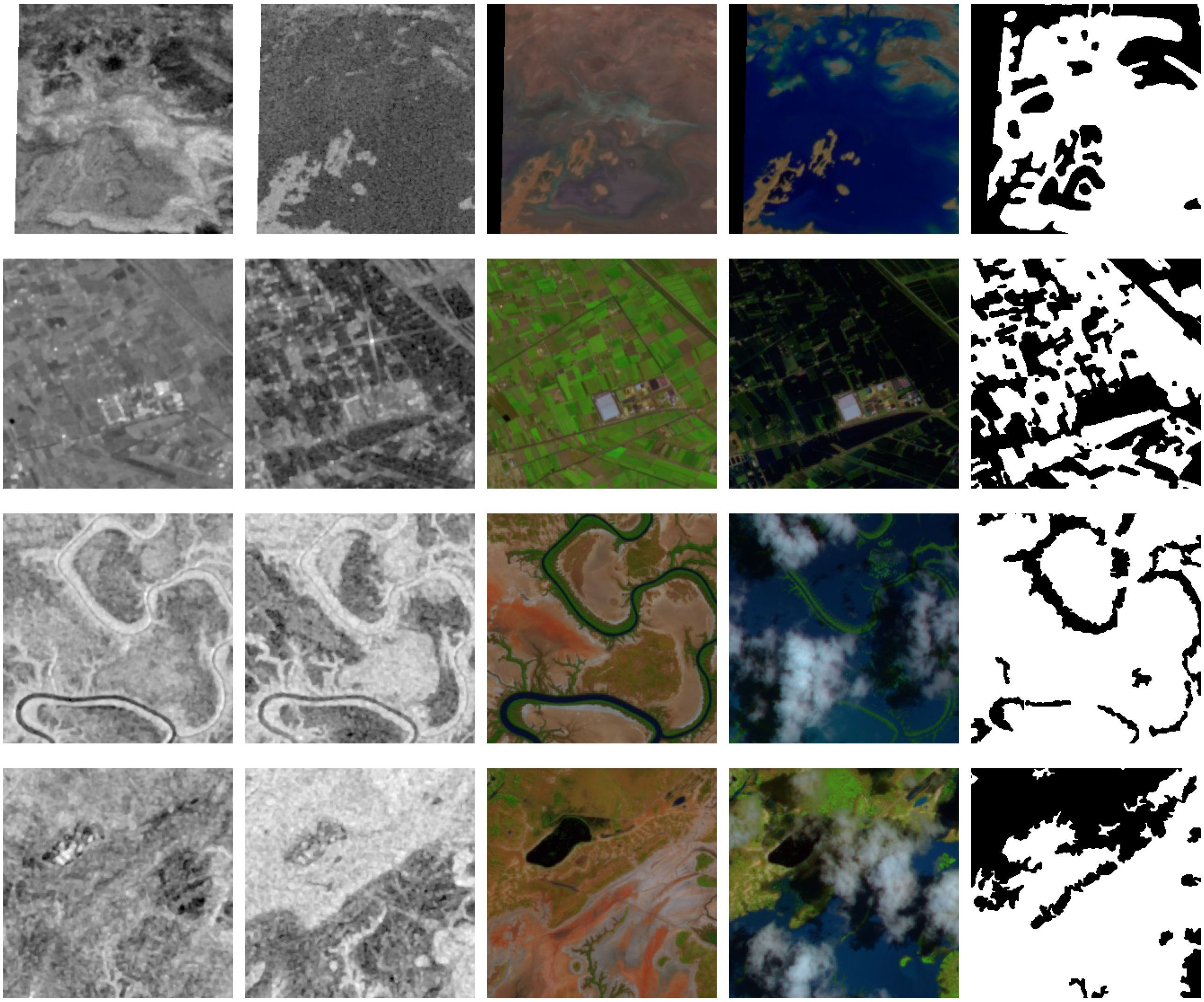


Fig. 2. OMBRIA Dataset samples. From left to right: Sentinel-1 preevent, Sentinel-1 postevent, Sentinel-2 preevent, Sentinel-2 postevent, and ground truth (white pixels is flood).

In Fig. 2, selected samples are presented. During or after a flood event, there is a high probability of cloud presence. We took that into consideration during data collection and included cloudy samples (Sentinel-2 only as SAR has the ability to penetrate clouds) to simulate realistic data capture scenarios.

The coordinate reference system of the imagery is the World Geodetic System 1984 [60], the so-called WGS ‘84, which stands as a gold standard in geodesy, satellite navigation, and cartography. It consists of an ellipsoid that is an oblate spheroid, a geodetic datum (horizontal and vertical), and the geographical coordinates, which are the angles measured in terms of latitude (ϕ), north or south of the equatorial plane, and longitude (λ), east or west of the prime meridian. The coordinate reference system utilized by the EMS is the Universal Transverse Mercator (UTM) that divides the Earth into 60 longitudinal and 20 latitudinal segments. For the coregistration task of the flooded areas, it is necessary to reproject the Sentinel-2 imagery. A map projection is defined as a systematic transformation of ellipsoidal coordinates to a plane coordinate system $(x, y) = (f_1(\phi, \lambda), f_2(\phi, \lambda))$.

The equations for reprojecting latitude and longitude to the cartesian coordinates as presented in [61] are formulated as

$$x = k_0 N \Bigg[A + \frac{(1 - T + C)A^3}{6} + \frac{(5 - 18T + T^2 + 72C - 58e'^2)A^5}{120} \Bigg]$$
$$y = k_0 \Bigg[M - M_0 + N \tan\phi \Bigg(\frac{A^2}{2} + \frac{(5 - T + 9C + 4C^2)A^4}{24} + \frac{(61 - 148T + 16T^2)A^6}{720} \Bigg] \quad (1)$$

where

$$e'^2 = \frac{e^2}{1 - e^2},\ N = \frac{\alpha}{(1 - e^2 \sin^2 \phi)^{1}/2},\ T = \tan^2 \phi$$
$$C = e'^2 \cos^2 \phi,\ A = (\lambda - \lambda_0) \cos\phi$$

$$M = \alpha\left[\left(1 - \frac{e^2}{4} - \frac{3e^4}{64} - \frac{5e^6}{256}\right)\phi - \left(\frac{3e^2}{8} + \frac{3e^4}{32} + \frac{45e^6}{1024}\right)\sin^2\phi + \left(\frac{15e^4}{256} + \frac{44e^6}{1024}\right)\sin^4\phi\right]. \tag{2}$$

Latitude and longitude are expressed in radians, *M* is the true distance from the Equator to ϕ along the central meridian, and M_0 is computed with respect to the latitude ϕ_0 that crosses the central meridian at the origin of the (x, y) coordinates. For these equations, k_0 is specified for a 6° longitude window at 0.9996 [61].

IV. Proposed Method: OmbriaNet

A. Problem Formulation and Theoretical Background

The goal in semantic segmentation has several variations from scene prediction to dense, fine-grained predictions and instance separation [62]. In this article, we focused on dense prediction, i.e., per-pixel class segmentation. To that end, we formulate the pixel labeling problem as assigning a class from a label space $\mathcal{L} = \{\ell_1, \ell_2, \ldots, \ell_k\}$ to each pixel of a set of 2-D or high-dimensional images $\mathcal{X} = \{x_1, x_2, \ldots, x_N\}$. For the problem addressed in this article, there is one class for "water" and another for "not water." Therefore,

$$\mathcal{L} = \begin{cases} 0, & \text{if not flooded} \\ 1, & \text{if flooded.} \end{cases}$$

Classic machine learning approach to this problem is by means of SVMs. SVMs build a discriminating function that simulates the optimal discriminating surface between classes, using training data [9]. When linear separation is impossible, kernel techniques are used so that the hyperplane defining the SVMs corresponds to a nonlinear decision boundary in the input space that is mapped to a linearized higher dimensional space [63].

Over the last years, deep learning (DL) has revolutionized several remote sensing analysis tasks, including the challenging problem of change detection, among others. Due to its fully data-driven structure, DL-based approaches learn automatically higher level features providing more faithful and representative approximations of input feature space. DL architectures with several intermediate hidden layers efficiently encode the internal representations of the raw data, and thus, they exhibit a superior performance compared to the traditional, shallow machine learning (ML)-based techniques. Additionally, DL techniques are quite robust and effective in remote sensing image segmentation tasks.

An efficient DL approach is based on CNNs. CNNs are composed of convolutional layers that alternate with nonlinear activation and possibly subsampling (pooling) layers, resulting in a hierarchy of increasingly abstract features. CNNs involve the mathematical operation of convolution. For the case of 2-D imagery, the convolution is applied on input features $I(i, j, k)$, where (i, j) represent the input image height and width and k is the input depth or channel, and it is expressed as

$$\begin{aligned} Y^{(l)} &= I^{(k)} * K^{(l,k)} \\ &= \sum_m \sum_n \sum_k I(m, n, k)K(i - m, j - n, l, k) \end{aligned} \tag{3}$$

where $K^{(l,k)} \in \mathbb{R}^{m,n}$ is the convolution kernel of size $m \times n$, associated with input channel k and output channel l.

Features generated from convolving the inputs are passed through a nonlinear function, the activation function. We use the leaky rectified linear unit (ReLU) [64], expressed as

$$y_i = \begin{cases} x, & \text{if} x \geq 0 \\ \frac{x}{\alpha}, & \text{if} x < 0 \end{cases} \tag{4}$$

where α is a hyperparameter to be tuned. Leaky ReLU can assist to overcome the vanishing gradient problem and avoid network saturation.

To determine the optimal values for the weights during training, it is necessary to select the proper loss function. The loss function is defined as $\mathcal{L} = \mathbb{E}(\hat{\mathbf{y}} - \mathbf{y})$, where $\hat{\mathbf{y}}$ is the predicted value and $\mathbf{y}$ the ground truth. One common loss function that is widely used is the binary cross-entropy loss function

$$\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}}) = -(y\log(\hat{y}) + (1 - y)\log(1 - \hat{y})). \tag{5}$$

B. Deep Learning Architectures

1) Baseline Architecture: U-Net: The majority of visual classification tasks target to assign a label to a provided image. Nevertheless, in remote sensing applications, a key objective also concerns the localization of the labels, i.e., the corresponding class that is assigned to each pixel. In this article, we use U-Net, a state-of-the-art image segmentation neural network, as the baseline and we further develop *novel* multimodal and multitemporal DL architectures designed specifically for the problem of flood mapping change detection.

U-Net [65] is a fully connected CNN architecture particularly applied on biomedical segmentation problems. It consists of a contracting path to imprint context and an expanding path, which is symmetric and enables precise localization. The U-Net architecture has also been exploited in various remote sensing tasks.

In our analysis, we perform adjustments on the basic U-Net architecture to perform flood mapping, and we further use it in our experimental setup as a baseline evaluation. The new proposed contracting path consists of a repeated pattern of two 3×3 convolutions activated via a Leaky ReLU function [66], instead of the ReLU activation that is used in the original implementation of U-Net. Additionally, a 2×2 max pooling operation with stride 2 is used for down sampling. In the expansive path, the feature maps are up-sampled, followed by a 2×2 convolution that halves the number of feature maps. Then, a concatenation with the corresponding cropped feature maps from the contracting path is implemented, followed by two 3×3 convolutions activated with the Leaky ReLU function. Moreover, a final 1×1 convolution layer is exploited that maps each feature vector to the desired class. Regarding the loss function, the binary cross-entropy was selected, since we

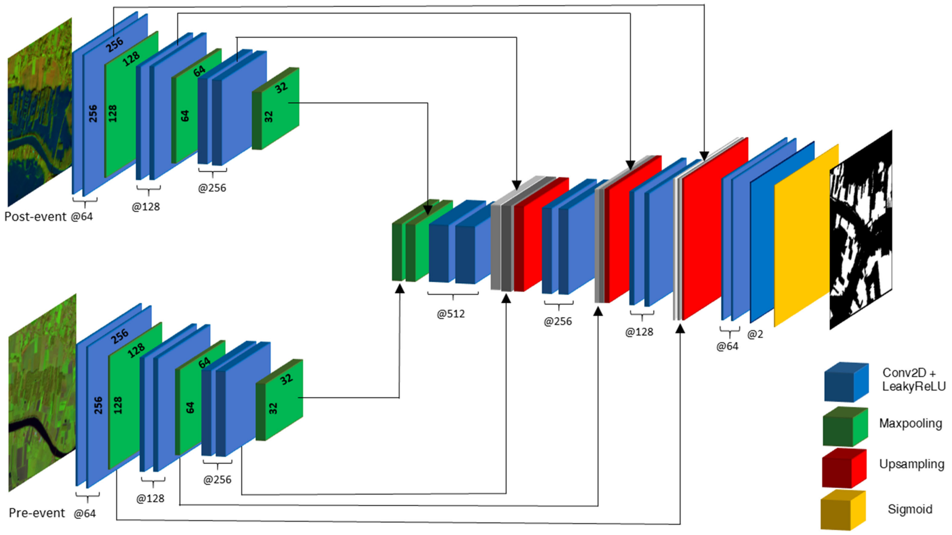


Fig. 3. Bitemporal OmbriaNet architecture. Two input images of the same region in different timestamps are imported into our proposed two-branch architecture attempting to take advantage of the bitemporality in detecting the change that is present.

address a binary classification task. Concerning the optimization algorithm, the Adam optimizer [67] was selected. Finally, the total number of parameters is 31 032 837.

Despite U-Net's superior performance in several remote sensing tasks and its promise for the proposed flood detection task, a major drawback still remains that the traditional U-Net considers single input imagery. This limitation makes the network incapable of distinguishing permanent from flooded water territories. Consequently, the basic U-Net formulation results in misclassification of water bodies (i.e., lakes, rivers, oceans, etc.) as flooded water areas resulting in the deterioration of the overall system detection accuracy.

To compensate the inability of U-Net in detecting accurately flooded areas, we introduce the novel *OmbriaNet* architecture. The intuition behind our approach is that there are three sources of water: main water bodies such as sea oceans and rivers, temporal streams only in winter seasons, and flood water. Temporal streams and flood water present a periodicity. This event periodicity may be captured if the proposed DL architecture is fed with the same area of interest in two different chronological moments that represent characteristic timestamps before and after the event.

Intense precipitation is related to dense cloud coverage. These clouds cause severe obstructions and heavy shadowing effects to satellite optical sensors resulting in significant data information loss. We leverage this discrepancy by introducing multimodal input in the proposed DL architecture. For this purpose, we exploit SAR data, since it is not limited by illumination or cloud coverage conditions. Additionally, SAR technology is proved to be a valuable source of information due to its higher probability of capturing imagery right after the flood event. In our proposed OmbriaNet architecture, two different approaches were designed: the Bitemporal OmbriaNet and the Multimodal OmbriaNet. OmbriaNet's intuition is based the U-Net framework but designed to receive multiple input. Different input is categorized as same sensor and different timestamp or both different sensor and different timestamp. In the first instance, the network is referred as Multitemporal OmbriaNet and in the second is referred as Multimodal OmbriaNet. The following paragraphs provide their complete description.

2) Bitemporal OmbriaNet: Creating meaningful feature maps from multitemporal images improves the change detection accuracy because the network detects modifications based on the temporal information from the feature maps generated from temporal images. Bitemporal OmbriaNet takes as input two images of a region in two different timestamps. The first is before the event and the second is right after the event. Three blocks of double 3 × 3 Convolutions, with a 2 × 2 max pooling factor and Leaky ReLU activations, encode the input data into the deepest point of the network. The resulting feature maps are concatenated, and then, a dropout layer with a 0.3 probability value is applied for regularization. Additionally, three blocks of double 3 × 3 convolutions and a 2 × 2 up-sampling factor are combined with skip connections and are activated via a sigmoid activation layer for obtaining the final predictions. Regarding the loss function, we utilize the binary cross entropy, as it was also used in the original U-Net implementation. Additionally, the exploited optimizer is the Adam. Finally, the Bitemporal OmbriaNet architecture has a total of 10 796 485 parameters. In Fig. 3, we depict the detailed Bitemporal OmbriaNet scheme.

3) Multimodal OmbriaNet: To fully exploit multimodal data fusion, we constructed the so-called Multimodal OmbriaNet architecture, as an improved version of the Bitemporal OmbriaNet that considers *four input images*, corresponding to one pre- and one postevent image acquired by Sentinel-1, and one pair of pre- and postevent images acquired by Sentinel-2, depicting the same spatial territory. Similar to the Bitemporal OmbriaNet architecture, three convolutional 3 × 3 blocks followed by 2 × 2 max-pooling layers were exploited and activated via the Leaky ReLU function. The resulting feature maps are concatenated, while the dropout regularization term is also set to 0.3. Regarding the inverse process, three 3 × 3 convolutional blocks

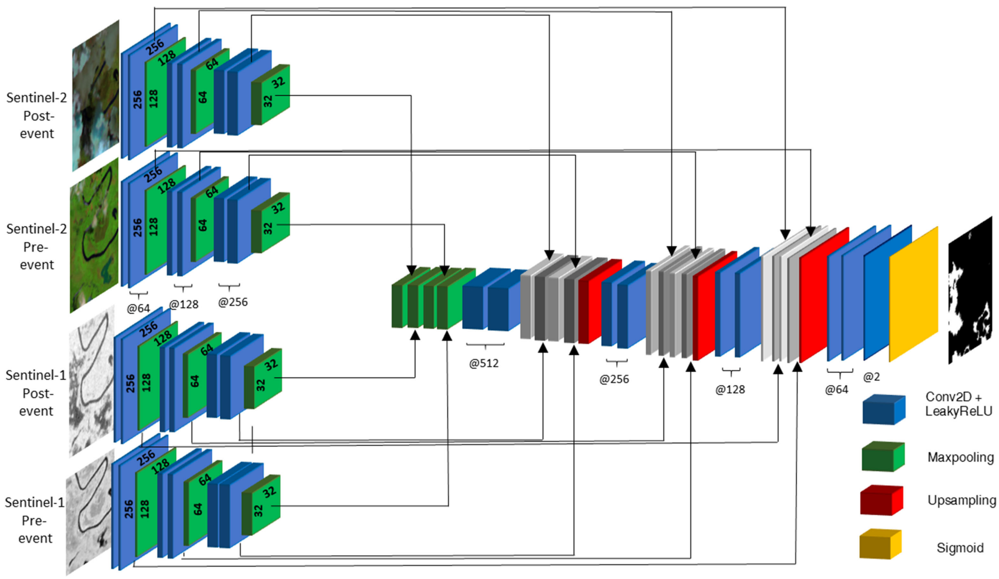


Fig. 4. Multimodal OmbriaNet architecture. In our final proposed method, four input images of different modalities and different timestamps are exploited in a multibranch architecture and utilized to segment the flood.

followed by 2 × 2 up-sampling layers are combined with two more skip connections, for completeness purposes. The final layer is a fully connected activated using the sigmoid function and is responsible for the segmentation task. The Multimodal OmbriaNet architecture resolves a total of 18 108 101 trainable parameters. In Fig. 4, we present the proposed Multimodal OmbriaNet architecture.

V. Experimental Results

A. Analysis Ready Dataset

Data preprocessing was conducted on the widely utilized Google Earth Engine [68] platform for geo-spatial science data and analysis. Data collections from Sentinel-1 and Sentinel-2 were accessed via this platform. In order to implement our proposed bitemporal approach, we selected two time stamps as follows. The Date 1 ranges from the 1st to the 31st of May of the year before the flood event. In this window, available pixels with cloud coverage below a threshold between 10% and 30% were averaged to yield the final input intensities. The Date 2 ranges from the date that the event has occurred until 15 days after it. The same cloud coverage threshold was applied and instead of pixel averaging, the first available pixels were selected. All pixels were reprojected to the corresponding UTM zone, depending on the region that is illustrated in Table I. We consider a significant preprocessing step by adopting a generalized data augmentation approach. We apply in each patch of the input EO imagery the following set of transformations:left-right flip, horizontal and vertical shift, shearing, and random rotations. Via this approach, our dataset size increased by a factor of 2, covering the dominant possible scenarios regarding data transformation. The dataset was then shuffled and further divided into training (80%), validation (10%), and testing (10%). All preproccessing steps are visualized as a flow chart in Fig. 5.

B. Evaluation Metrics

Evaluation using standard and well-known metrics is of critical importance, since it enables fair comparisons with the state-of-the-art. Additionally, the context of the application determines the importance of the metrics. We also provide execution time as an evaluation parameter, bearing in mind that computation times depend on the available hardware resources. In semantic segmentation, the most popular performance metrics are *pixel accuracy* and *intersection over union (IoU)*. In our case, we have a total of two classes (ℓ_0 and ℓ_1). Let p_{ij} be the number of pixels of the class i inferred to belong to the class j. Then, p_{ii} is the number of correctly classified or true positives, while p_{ij} and p_{ji} represent the false positives and true negatives, respectively.

1) *Pixel accuracy (PA)*: PA is defined as the ratio of the amount of correctly classified pixels to their total number:

$$\mathrm{PA} = \frac{\sum_{i=0}^{1} p_{ii}}{\sum_{i=0}^{1}\sum_{j=0}^{1} p_{ij}}. \tag{6}$$

2) *Intersection over union (IoU)*: IoU is the most frequently used metric for image segmentation. It stands as the ratio between the intersection and the union of two sets. In our formulation, it represents the prediction and the ground truth. It is formulated as the number of true positives over the sum of true positives, false negatives, and false positives. It is computed in a per-class basis and averaged. For the binary problem, it is written as

$$\mathrm{IoU} = \frac{1}{2}\sum_{i=0}^{1} \frac{p_{ii}}{\sum_{j=0}^{1} p_{ij} + \sum_{j=0}^{1} p_{ji} - p_{ii}}. \tag{7}$$

3) *Frequency weighted intersection over union (FWIoU)*: FWIoU is an improved version of the IoU metric that takes into consideration class appearance frequencies and

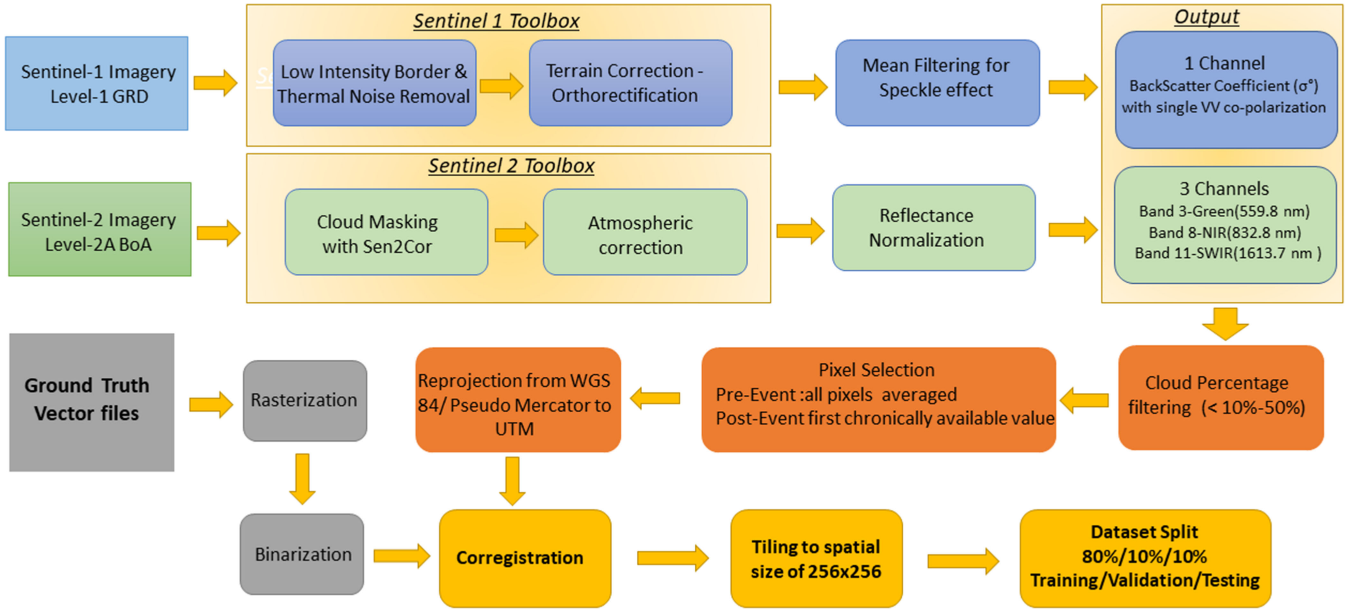

Sentinel-1 Imagery Level-1 GRD
Sentinel 1 Toolbox
Low Intensity Border & Thermal Noise Removal
Terrain Correction - Orthorectification
Mean Filtering for Speckle effect
Output
1 Channel BackScatter Coefficient (σ°) with single VV co-polarization
Sentinel-2 Imagery Level-2A BoA
Sentinel 2 Toolbox
Cloud Masking with Sen2Cor
Atmospheric correction
Reflectance Normalization
3 Channels Band 3-Green(559.8 nm) Band 8-NIR(832.8 nm) Band 11-SWIR(1613.7 nm)
Ground Truth Vector files
Rasterization
Reprojection from WGS 84/ Pseudo Mercator to UTM
Pixel Selection Pre-Event :all pixels averaged Post-Event first chronically available value
Cloud Percentage filtering (< 10%-50%)
Binarization
Corregistration
Tiling to spatial size of 256x256
Dataset Split 80%/10%/10% Training/Validation/Testing

Fig. 5. Data preproccessing flowchart. All preproccessing was performed in Google Earth Engine API.

TABLE II
U-NET BASELINE SCHEME: QUANTITATIVE PERFORMANCE OF TRAINING AND VALIDATION PHASES

Methods	U-Net (Sentinel-2)			U-Net (Sentinel-1)		
Batch size	**Training Accuracy**	**Validation Accuracy**	**Training Time**	**Training Accuracy**	**Validation Accuracy**	**Training Time**
2	0.7876	0.7712	23	0.7268	0.8487	22
4	0.7976	0.7796	20	0.7325	0.8523	20
6	**0.8205**	0.8442	18	0.7364	**0.8558**	18
8	0.8163	**0.8493**	18	**0.7631**	0.8388	18
12	0.8132	0.8480	17	0.7592	0.8510	17

weighs their importance

$$\text{FWIoU} = \frac{1}{\sum_{i=0}^{1}\sum_{j=0}^{1} p_{ij}} \times \sum_{i=0}^{1} \frac{\sum_{j=0}^{1} p_{ii} p_{ij}}{\sum_{j=0}^{1} p_{ij} + \sum_{j=0}^{1} p_{ji} - p_{ii}}. \quad (8)$$

C. Evaluation

In this Section, we provide the complete experimental analysis that has been benchmarked on the developed OMBRIA dataset. Specifically, we provide a detailed experimental setup corresponding to various training parameters, including different batch sizes and number of epochs. Additionally, we evaluate the proposed DL methods in terms of the training, validation, and testing accuracy. Moreover, we provide quantitative and qualitative results of our proposed Bitemporal and Multimodal architectures and we compare with competitive state-of-the-art ML and DL algorithmic formulations.

1) Baseline: U-Net: The initial experiments were conducted using the U-Net architecture on both Sentinel-1 and Sentinel-2 data, separately. Since U-Net allows single input data, only post-flood event imagery was utilized. For the specific architecture, the best case scenario is achieved using 50 epochs, a batch size of 8 for Sentinel-2, and a batch size of 12 for Sentinel-1.

Table II presents the quantitative performance regarding the training and validation phases of the U-Net baseline architecture. We observe, that using Sentinel-2 input data, the highest training accuracy is achieved with a batch size of 6, and it is $\mathbf{82.05}\%$, while the highest validation accuracy is $\mathbf{84.93}\%$, using a batch size of 8. Additionally, using Sentinel-1 input data, the highest training accuracy is $\mathbf{76.31}\%$ using a batch size of 8, while the best validation accuracy of $\mathbf{85.58}\%$ is achieved using a batch size 6. We also observe that the validation score is higher than the training score, especially in Sentinel-1, probably due to the sampling batch during the training process.

In Table III, we report the scores of the trained models on the testing set. We observe that the highest overall score is achieved with Sentinel-2 for a batch size of 8, and it is $\mathbf{84.67}\%$ in PA, $\mathbf{43.46}\%$ in IoU, and $\mathbf{75.51}\%$ in FW IoU. The results indicate that Sentinel-2 slightly outperforms Sentinel-1 when examined in terms of accuracy.

In Fig. 6, a sample from the test set is illustrated that highlights the advantage of using SAR over optical imagery. Due to a large concentration of clouds, Sentinel-2 was not able to capture any flooded water areas, which is demonstrated by the relatively low value for the classification accuracy.

TABLE III
QUANTITATIVE PERFORMANCE COMPARISON OF U-NET WITH SENTINEL-1 AND SENTINEL-2 IN TERMS OF PIXEL ACCURACY, IOU, FW IOU, AND TRAINING TIME FOR 50 EPOCHS

Methods	**U-Net(Sentinel-2)**			**U-Net(Sentinel-1)**		
Batch size	**Pixel Accuracy**	**IoU**	**FW IoU**	**Pixel Accuracy**	**IoU**	**FW IoU**
2	0.7596	0.3378	0.6339	0.7520	0.3912	0.6567
4	0.7700	0.3686	0.6652	0.7716	0.4526	0.6823
6	0.8386	0.4262	0.745	0.7372	0.3180	0.6343
8	**0.8467**	**0.4346**	**0.7551**	0.7781	0.4014	0.6828
12	0.8412	0.4143	0.7461	**0.7926**	**0.4619**	**0.7030**

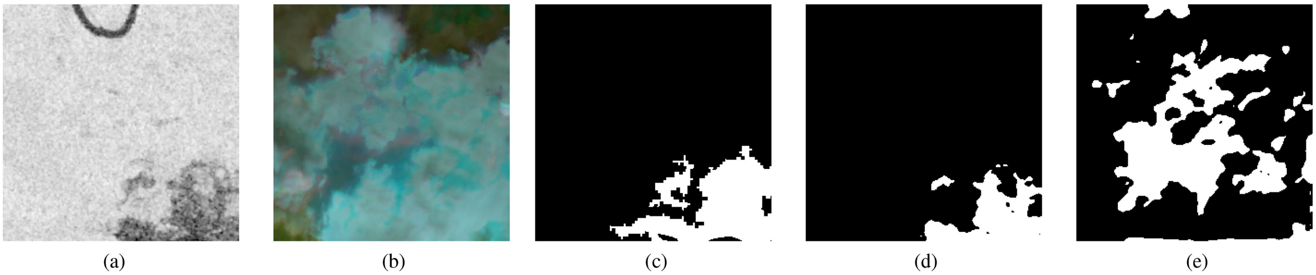


Fig. 6. U-Net baseline scheme: Qualitative comparison of selected samples, in terms of both visual perception and the corresponding IoU metric. (a) Sentinel-1 Input. (b) Sentinel-2 Input. (c) Ground Truth (White pixelsis flood). (d) Sentinel-1 Output (58.19%). (e) Sentinel-2 Output (9.24%).

TABLE IV
QUANTITATIVE PERFORMANCE OF TRAINING AND VALIDATION FOR BITEMPORAL OMBRIANET

Methods	**Bitemporal OmbriaNet(50 epochs)**			**Bitemporal OmbriaNet(100 epochs)**		
Batch size	**Training Accuracy**	**Validation Accuracy**	**Training Time (min)**	**Training Accuracy**	**Validation Accuracy**	**Training Time (min)**
2	0.8312	0.8269	26	0.8425	0.8269	53
4	**0.8264**	**0.8492**	25	0.8448	**0.8566**	51
6	0.8298	0.8266	25	0.8407	0.7026	49
8	0.8327	0.8241	25	0.8445	0.8231	50
12	0.8302	0.8381	23	**0.8455**	0.8397	47

TABLE V
QUANTITATIVE PERFORMANCE OF TRAINING, VALIDATION ACCURACY, AND COMPUTATIONAL TIME FOR MULTIMODAL OMBRIANET

Methods	**Multimodal OmbriaNet(50 epochs)**			**Multimodal OmbriaNet(100 epochs)**		
Batch size	**Training Accuracy**	**Validation Accuracy**	**Training Time (min)**	**Training Accuracy**	**Validation Accuracy**	**Training Time (min)**
2	0.8540	0.7987	37	0.8719	0.7864	82
4	0.8415	0.8715	35	0.8755	0.8578	79
6	0.8521	0.8496	35	0.8713	0.8066	70
8	0.8533	0.8735	35	0.8445	0.8231	70
12	**0.8554**	**0.8659**	38	**0.8765**	**0.8596**	76

2) OmbriaNet—Bitemporal: In this subsection, we report on experiments meant to evaluate the improvement in our overall system accuracy when introducing bitemporal imagery. The goal is to demonstrate that the proposed OmbriaNet architecture described in Section IV, which takes as input preevent and postevent images, is capable of "learning" the change between the absence of water in normal conditions and the presence of water when the region is flooded. In our experiments, we experimented with different batch sizes, as with the traditional U-Net case.

In Table IV, the quantitative performance comparison of training and validation metrics for the Bitemporal OmbriaNet architecture is demonstrated. In this scenario, for a fixed batch value of 4, the best validation accuracy that is achieved is **84.92%** for 50 epochs and **85.66%** for 100 epochs. Multimodal OmbriaNet quantitive performance comparison is shown in Table V. Fora batch size of 12 the validation accuracy is 86.59% and 85.96% for 50 and 100 epochs respectively.

In Table VI, we illustrate the quantitative performance comparison on the testing set with the OmbriaNet architecture. We

TABLE VI
QUANTITATIVE PERFORMANCE COMPARISON ON TESTING SET FOR BITEMPORAL AND MULTIMODAL OMBRIANET IN TERMS OF PIXEL ACCURACY, IOU, FW, AND IOU (100 EPOCHS)

Methods	**Bitemporal OmbriaNet**			**Multimodal OmbriaNet**		
Batch size	PA	IoU	FW IoU	PA	IoU	FW IoU
2	0.8614	0.6330	0.7779	0.8755	0.6492	0.8029
4	0.8613	**0.6415**	0.7793	0.8776	0.6805	0.8043
6	0.8531	0.6306	0.7706	0.8887	**0.7093**	0.8218
8	**0.8709**	0.6208	**0.7865**	0.8898	0.6918	0.8177
12	0.8572	0.6014	0.7634	**0.8994**	0.7049	**0.8302**

TABLE VII
OMBRIANET PERFORMANCE COMPARISON ON OMBRIA DATASET ON 90%/10% TRAINING/TESTING SPLIT

Methods	**PA**	**IoU**	**FW IoU**
Otsu's Thresholding (Sentinel-2)	0.7930	0.333	0.6889
Multimodal SVM	0.8504	0.6245	0.7631
U-Net (Sentinel-1)	0.7925	0.5734	0.6971
U-Net (Sentinel-2)	0.8251	0.5418	0.7221
Bitemporal OmbriaNet (Sentinel-1)	0.7203	0.5181	0.6229
Bitemporal OmbriaNet (Sentinel-2)	0.8733	0.6457	0.7919
Multimodal OmbriaNet (Sentinel-1 & Sentinel-2)	**0.9010**	**0.7236**	**0.8330**

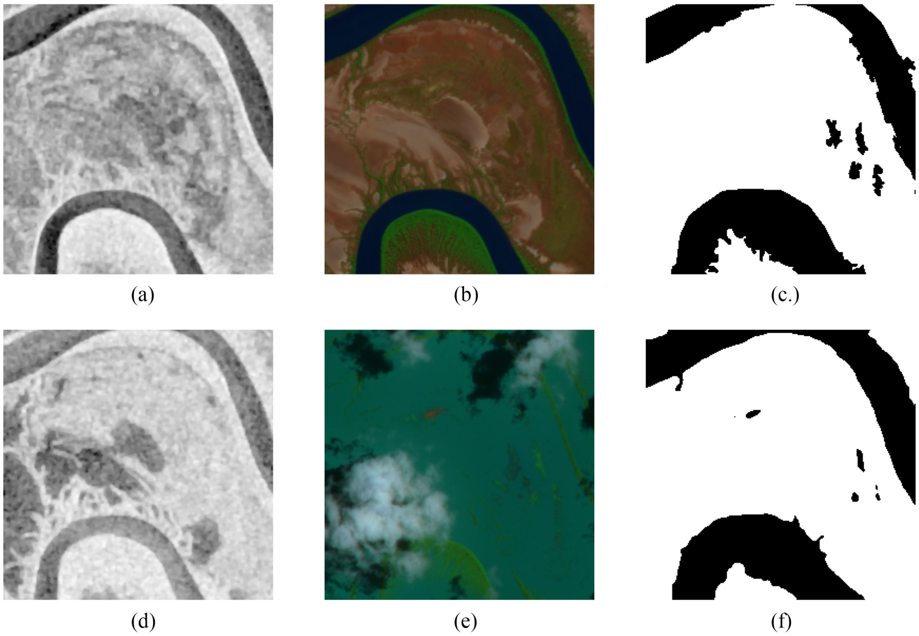


Fig. 7. Qualitative comparison of selected samples with Bitemporal OmbriaNet and their corresponding IoU metric. (a) Sentinel-1 (Pre-event). (b) Sentinel-2 (Pre-event). (c) Ground Truth (White pixels is flood). (d) Sentinel-1 (Post-event). (e) Sentinel-2 (Post-event). (f) Bitemporal Output (94.64%).

observe a significant improvement in the IoU metric (about 5%). If examined individually, the range of IoU values for the different batch sizes is far larger for the U-Net (i.e., approximately 10%) as compared to the OmbriaNet (i.e., approximately 4%). This observation indicates a higher model robustness. Additionally, pixel accuracy exhibits a small increase, approximately 2–3% when comparing the different architectures, and demonstrates similar variations when examined for different batch sizes (i.e., about 10% for U-Net and only 2% for OmbriaNet). Experiments showed that the model behavior presents minimum change toward its performance, when increasing the number of epochs. Specifically increasing the training epochs yields into a limited gain (i.e., approximately 1%), while the training time is almost doubled when doubling the training iterations. In this scenario, the best evaluation metrics values are achieved using a batch size of 8, regarding pixel accuracy and FW IoU.The resulting predictions in Table VII were implemented from the corresponding models that were trained with a batch size of 8.

In Fig. 7, a sample of inundated areas with rivers is illustrated to demonstrate the model ability to segregate permanent water

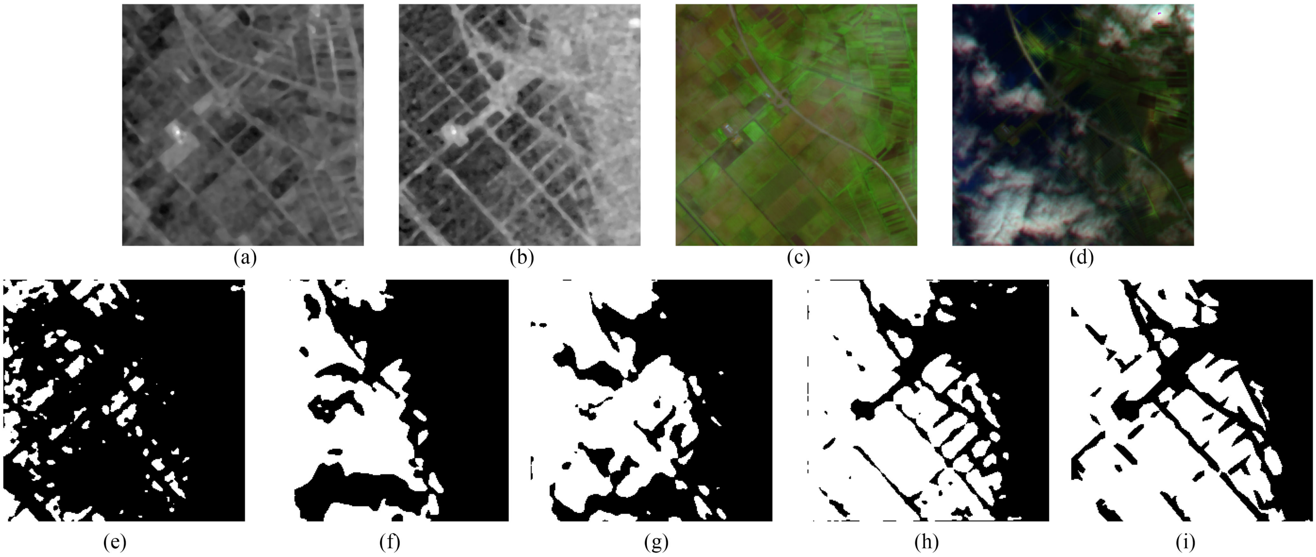


Fig. 8. Qualitative comparison of selected sample with Multimodal OmbriaNet and their corresponding IoU metric. (a) Sentinel-1 (Pre-event). (b) Sentinel-1 (Post-event). (c) Sentinel-2 (Pre-event). (d) Sentinel-2 (Post-event). (e) Sentinel-1 U-Net (22.19%). (f) Sentinel-2 U-Net (59:13%). (g) Bitemporal OmbriaNet (60:96%). (h) Multimodal OmbriaNet (81:90%). (i) Ground Truth (White pixels is flood).

from flooded territories. As we may observe, the Bitemporal OmbriaNet depicts a high classification IoU score of $\mathbf{96.64}\%$, validating its better robustness compared to the traditional U-Net scheme.

3) OmbriaNet—Multimodal: Bitemporal OmbriaNet showed a significant improvement over the baseline of U-Net motivating the fusion of Sentinel-1 and Sentinel-2 preevent and postevent imagery via our proposed multimodal OmbriaNet, which indeed gave the best results. Specifically, running the experiments with the same hyperparameters as in Section V-C2, Multimodal OmbriaNet outperforms the baseline of U-Net by over 10% and Bitemporal OmbriaNet by 7% in IoU score achieving a value of 70.93%. Increasing the number of epochs did not show a significant increase in accuracy.

In Fig. 8, a sample comparison between different experiments is presented. The Multimodal OmbriaNet improves the prediction performance about 21% compared to Multitemporal OmbriaNet achieving an IoU score 81.90% over 60.96%. A remarkable note is that although in the Sentinel-2 postevent instance, there is extensive cloud coverage over the flooded areas, our multimodal network manages a very good delineation of the event. Table VI gives the quantitative results over the test set for Multimodal OmbriaNet. A significant improvement in overall accuracy is observed, especially in the IoU score, in which Multimodal OmbriaNet outperforms the Bitemporal OmbriaNet by about 10%. The highest score is achieved for batch size of 12, and that is the model used in Table VII.

4) Overall Comparison: In this paragraph, we illustrate the final comparisons among the proposed architectures and the related state-of-the-art approaches. In order to obtain a fair comparison, we retrained all models (i.e., the baseline U-Net, proposed Multimodal, and proposed Bitemporal) with 90% of the OMBRIA dataset, and we tested on the resulting 10% test set. Regarding the state-of-the-art approaches, we compared against Otsu's method for completeness in benchmarking evaluation. Since Otsu's algorithm input is based on a single value pixel intensity, we derive the modified normalized difference water index (MNDWI) [69] from the Sentinel-2 imagery, provided by

$$\text{MNDWI} = \frac{B_{\text{GREEN}} - B_{\text{SWIR}}}{B_{\text{GREEN}} + B_{\text{SWIR}}}$$

where B_{GREEN} and B_{SWIR} are reflectance values of the corresponding channels. Additionally, for the ML-based SVM algorithm, we choose a linear kernel over the RBF as it results in a smaller time complexity and converges much faster on large datasets (i.e., about 45 million pixels). The Hinge loss function was selected, L2 norm for penalization, while the parameter C is set to 10. Feature selection was performed on raw pixel values and no complex features were constructed, but different sets of features were tested depending on sensor and timestamp. The multimodal combination was selected for the SVM.

Regarding all DL architectures, the batch sizes and number of epochs were chosen for best performance, resulting in a batch size of 12 for the U-Net with Sentinel-1, a batch size of 8 for the U-Net with Sentinel-2, and 50 epochs for both schemes. Concerning the Bitemporal OmbriaNet, both with Sentinel-1 and Sentinel-2, the batch size number was set to 8, while the number of epochs is fixed to 100. Finally, for the Multimodal OmbriaNet, the batch size was fixed to 12 and the number of epochs to 100. The corresponding results are presented in Table IX.

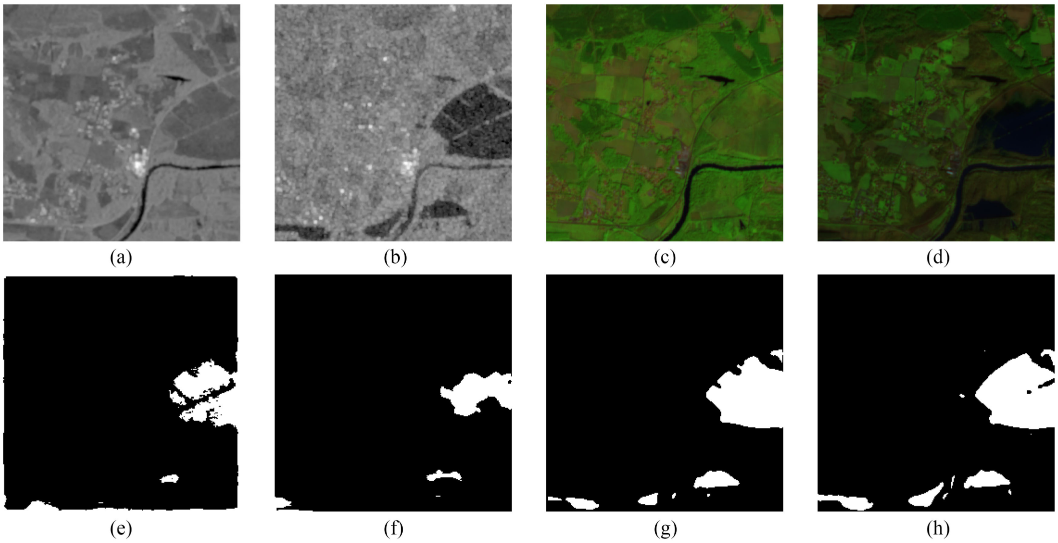
Performance scores show that networks trained with Sentinel-2 imagery perform better than Sentinel-1 in both U-Net and Bitemporal OmbriaNet. Sentinel-2 data utilize three reflectance channels (Band 3, Band 8, and Band 11). These channels were selected due to their high response on water, according to literature. IR (Band8) and SWIR (Band11) are infrared channels and water absorbs infrared radiation. Sentinel-1 one channel (VV polarization) has a similar response as the emitting signal

TABLE VIII
QUANTITATIVE PERFORMANCE OF TRAINING/VALIDATION AND TESTING OF OMBRIANET WITH MISSING INPUT

Methods	**Missing Sentinel-1 Post-event**					**Missing Sentinel-2 Post-event**				
Batch size	**Training Accuracy**	**Validation Accuracy**	**PA**	**IoU**	**FW IoU**	**Training Accuracy**	**Validation Accuracy**	**PA**	**IoU**	**FW IoU**
8	0.8324	0.8129	0.8607	0.6325	0.7774	0.7181	0.5788	0.7241	0.5321	0.6332
12	0.8378	0.8076	0.8633	0.6229	0.7767	0.7436	0.6673	0.7569	0.5315	0.6569

TABLE IX
OMBRIANET PERFORMANCE COMPARISON ON NEW FLOOD EVENTS

Flood Event	**Methods**	**PA**	**IoU**	**FW IoU**
France (ID 492)	U-Net (Sentinel-1)	0.9258	0.4629	0.8681
	U-Net (Sentinel-2)	0.9114	0.5007	0.8644
	Multimodal OmbriaNet	**0.9817**	**0.673**	**0.9641**
Albania (ID 501)	U-Net (Sentinel-1)	0.7780	0.4906	0.6754
	U-Net (Sentinel-2)	0.7733	0.5526	0.6851
	Multimodal OmbriaNet	**0.9050**	**0.7142**	**0.8390**
Timor (ID 507)	U-Net (Sentinel-1)	0.7292	0.4483	0.6554
	U-Net (Sentinel-2)	0.7864	0.5422	0.7211
	Multimodal OmbriaNet	**0.8993**	**0.6871**	**0.8461**
Guyana (ID 514)	U-Net (Sentinel-1)	0.7292	0.4483	0.6554
	U-Net (Sentinel-2)	0.8100	0.4864	0.7427
	Multimodal OmbriaNet	**0.951**	**0.7238**	**0.9137**


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

(e) (f) (g) (h)

Fig. 9. Comparison of selected sample from ID492 flood in France (IoU metric score). (a) Sentinel-1 (Pre-event). (b) Sentinel-1 (Post-event). (c) Sentinel-2 (Pre-event). (d) Sentinel-2 (Post-event). (e) Sentinel-1 U-Net (66:98%). (f) Sentinel-2 U-Net (63:19%). (g) Multimodal OmbriaNet (90:03%). (h) Ground Truth (White pixels is flood).

reflects away on the water's smooth surface. The higher channel number in Sentinel-2 leads to information surplus, thus increasing performance.

Experimental results for Bitemporal OmbriaNet are lower than the U-Net when using Sentinel-1 data. This can be explained by the speckle effect that is present in a number of preevent imagery. This effect creates problems in predictions in Bitemporal OmbriaNet. Although this is not a setback for the Multimodal OmbriaNet.

5) Case of Missing Input: As it is discussed in previous sections, OMBRIA dataset was constructed taking into account realistic conditions as the purpose is for the network to be used in real data. Clouds and drained parts of land are present in the dataset. For completeness reasons, we performed several experiments replacing successively each modality postevent in Multimodal OmbriaNet with black imagery to simulate missing input. We conducted experiments for two different batch sizes, 8 and 12, and we trained the networks for 50 epochs. The results

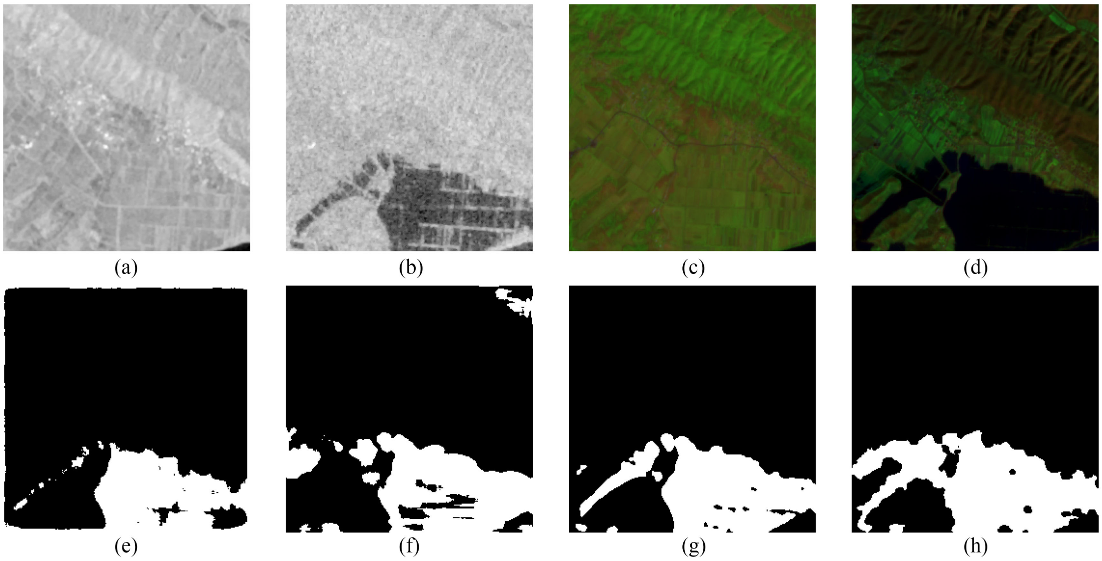


Fig. 10. Comparison of selected sample from ID501 flood in Albania (IoU metric score). (a) Sentinel-1 (Pre-event). (b) Sentinel-1 (Post-event). (c) Sentinel-2 (Pre-event). (d) Sentinel-2 (Post-event). (e) Sentinel-1 U-Net (63.82%). (f) Sentinel-2 U-Net (78:30%). (g) Multimodal OmbriaNet (89:14%). (h) Ground Truth (White pixels is flood).

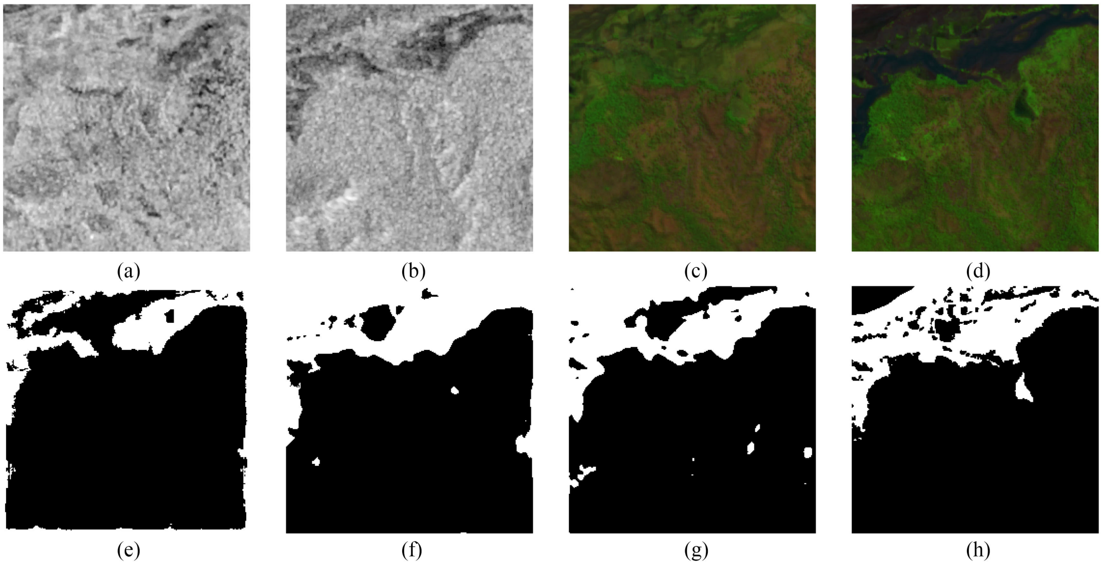


Fig. 11. Comparison of selected sample from ID507 flood in Timor (IoU metric score). (a) Sentinel-1 (Pre-event). (b) Sentinel-1 (Post-event). (c) Sentinel-2 (Pre-event). (d) Sentinel-2 (Post-event). (e) Sentinel-1 U-Net (69:50%). (f) Sentinel-2 U-Net (79:08%). (g) Multimodal OmbriaNet (79:44%). (h) Ground Truth (White pixels is flood).

are shown on Table VIII. For the case of missing Sentinel-1 postevent data, the network scores 86.07% in PA, 63.25% in IoU, and 77.74% in FW IoU. For the case of missing Sentinel-2 postevent data, the scores are 75.69% in PA, 53.15% in IoU, and 65.69% in FW IoU. Both cases achieve almost equal scores with Bitemporal OmbriaNet.

6) Testing OmbriaNet Performance in New Floods: Samples in OMBRIA dataset have spatial-temporal relationships as they were produced from same regions. This relationship can result two neighboring regions to share similar semantic segmentation information and leading to overfitting. We performed experiments on new flood events that are not included in the training process. These events are the EMS ID 492 (France) 501 (Albania), 507 (Timor), and 514 (Guyana). Event ID 492 took place in France in 2021 and includes regions that have flooded in 2019 and mapped in event ID 416. The rest are completely new events. For each flood, we used the pretrained models of the U-Net with Sentinel-1, U-Net with Sentinel-2, and Multimodal OmbriaNet that were used in Section V-C4. The results are presented in Table IX. Multimodal OmbriaNet again surpassed U-Net both with Sentinel-1, by about 20%–30% in IoU score, and with Sentinel-2, by about 15%–25%. This

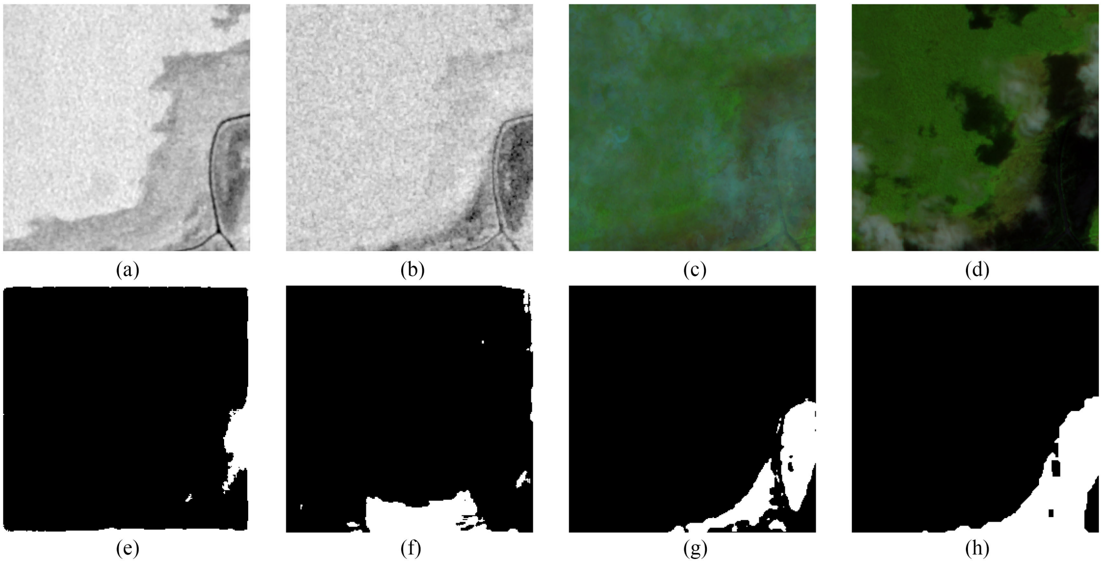


Fig. 12. Comparison of selected sample from ID514 flood in Guyana (IoU metric score). (a) Sentinel-1 (Pre-event). (b) Sentinel-1 (Post-event). (c) Sentinel-2 (Pre-event). (d) Sentinel-2 (Post-event). (e) Sentinel-1 U-Net (58.19%). (f) Sentinel-2 U-Net (53:72%). (g) Multimodal OmbriaNet (79:97%). (h) Ground Truth (White pixels is flood).

proves the network's superiority on change detection for flood events. U-Net with Sentinel-2 outperformed the U-Net with Sentinel-1 in all cases. The scores indicate that our proposed model is robust and effective on new and "unseen" flood events as the performance is at least equal or better than Table VII. Figs. 9–12 show the qualitative results on selected samples with their respective IoU score for each flood event.

VI. Discussion

The experimental results reveal that machine learning and deep learning methods outperform the traditional thresholding algorithms by at least 25% in IoU. SVMs perform decently if both optical and SAR data are given. One remarkable note is that bitemporality in our SAR data does not necessarily improve the performance. Examining Sentinel-1 VV band, we observed that the results are better with postevent imagery rather than combined with preevent. This is caused by the speckle effect that is inevitably present. The best model is Multimodal OmbriaNet that outperforms the SVM by 10% in IoU, Bitemporal OmbriaNet with Sentinel-2 by 8%, Bitemporal OmbriaNet with Sentinel-1 by 20%, and U-Net by more than 20%. Results also suggest that OmbriaNet is robust and effective in flood mapping under realistic conditions where input data are not ideal, cloudy, or even missing. OMBRIA dataset includes reference imagery that has already masked out permanent water bodies resulting our proposed architectures to learn how to automatically distinguish flooded from permanent water. This ability eliminates the necessity for manual permanent water body annotation and drastically reduces the workload. OMBRIA is an analysis ready dataset, which is introduced in this article. It is the first fused dataset comprising both optical and Radar imagery with different timestamps, and conducts a complete comparison with other state-of-the art methods not subjective. The closest work that includes fused data from Sentinel-1 and Sentinel-2 is [49], which we outperform by more than 30% in IoU. We also outperform [50] by 3% in PA and by 6% in FW IoU, which are, to our knowledge, the highest scores in flood delineation with remote sensing data.

VII. Conclusion

In this article, we presented a novel approach to address the problem of flood delineation, allowing detection of floods with satellite imagery. We introduced the OmbriaNet network that employs multimodal and multitemporal satellite imagery for semantic segmentation using supervised learning under realistic conditions. Our hope is that this article will contribute to the efforts of flood disaster management. Our approach showed that new platforms such as Google Earth Engine can be employed to construct a supervised dataset for remote sensing applications.

Ground-truth annotations provided by ESA can be of crucial importance for tackling the flood mapping problem, as hand labeling such types of data requires expertise in remote sensing photointerpretation and is time consuming. Computer vision contributes significantly to remote sensing problems such as land cover classification and cloud detection. There are few datasets available for training flood detection algorithms using publicly available satellite imagery. In this article, we focused on flooded water detection in an effort to operationalize monitoring for crisis situations. We formed and provided the OMBRIA dataset to the research community in order to train deep learning algorithms for flood detection without the overhead of generating training and validation datasets.

Future plans include the expansion of the dataset with more samples so that it generalizes better. Furthermore, we will investigate the expansion of the architecture to include more spectral bands in order to exploit the full potential of the Sentinel satellite constellation. Finally, it is worth exploring the transfer learning possibility on data from unmanned aerial vehicles, which can provide high-resolution imagery and will greatly improve the mapping accuracy.

REFERENCES

[1] Y. Hirabayashi *et al.*, "Global flood risk under climate change," *Nature Climate Change*, vol. 3, no. 9, pp. 816–821, 2013.

[2] MunuchRE, "Risks from floods, storm surges and flash floods," 2019. [Online]. Available: https://www.munichre.com

[3] P. Brivio, R. Colombo, M. Maggi, and R. Tomasoni, "Integration of remote sensing data and GIS for accurate mapping of flooded areas," *Int. J. Remote Sens.*, vol. 23, no. 3, pp. 429–441, 2002.

[4] A. Goffi, D. Stroppiana, P. A. Brivio, G. Bordogna, and M. Boschetti, "Towards an automated approach to map flooded areas from Sentinel-2 MSI data and soft integration of water spectral features," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 84, 2020, Art. no. 101951.

[5] Y. Du, Y. Zhang, F. Ling, Q. Wang, W. Li, and X. Li, "Water bodies' mapping from Sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the SWIR band," *Remote Sens.*, vol. 8, no. 4, p. 354, 2016, Art. no. 354.

[6] R. Brakenridge and E. Anderson, "Modis-based flood detection, mapping and measurement: The potential for operational hydrological applications," in *Transboundary Floods: Reducing Risks Through Flood Management*. Berlin, Germany: Springer, 2006, pp. 1–12.

[7] B.-C. Gao, "NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space," *Remote Sens. Environ.*, vol. 58, no. 3, pp. 257–266, 1996.

[8] H. Xu, "Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery," *Int. J. Remote Sens.*, vol. 27, no. 14, pp. 3025–3033, 2006.

[9] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th Annu. Workshop Comput. Learn. Theory*, 1992, pp. 144–152.

[10] T. K. Ho, "Random decision forests," in *Proc. 3rd Int. Conf. Document Anal. Recognit.*, vol. 1, 1995, pp. 278–282.

[11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[12] X. X. Zhu *et al.*, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.

[13] C. Karakizi, K. Karantzalos, M. Vakalopoulou, and G. Antoniou, "Detailed land cover mapping from multitemporal landsat-8 data of different cloud cover," *Remote Sens.*, vol. 10, no. 8, 2018, Art. no. 1214.

[14] M. Mahdianpari, B. Salehi, M. Rezaee, F. Mohammadimanesh, and Y. Zhang, "Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery," *Remote Sens.*, vol. 10, no. 7, 2018, Art. no. 1119.

[15] R. Stivaktakis, G. Tsagkatakis, and P. Tsakalides, "Deep learning for multilabel land cover scene categorization using data augmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 7, pp. 1031–1035, Jul. 2019.

[16] G. Tsagkatakis, A. Aidini, K. Fotiadou, M. Giannopoulos, A. Pentari, and P. Tsakalides, "Survey of deep-learning approaches for remote sensing observation enhancement," *Sensors*, vol. 19, no. 18, 2019, Art. no. 3929.

[17] M. Vakalopoulou, K. Karantzalos, N. Komodakis, and N. Paragios, "Building detection in very high resolution multispectral data with deep learning features," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2015, pp. 1873–1876.

[18] Z. Deng, H. Sun, S. Zhou, J. Zhao, L. Lei, and H. Zou, "Multi-scale object detection in remote sensing imagery with convolutional neural networks," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 3–22, 2018.

[19] Y. Long, Y. Gong, Z. Xiao, and Q. Liu, "Accurate object localization in remote sensing images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2486–2498, May 2017.

[20] S. Ahmad, A. Kalra, and H. Stephen, "Estimating soil moisture using remote sensing data: A machine learning approach," *Adv. Water Resour.*, vol. 33, no. 1, pp. 69–80, 2010.

[21] M. Pal and P. Mather, "Support vector machines for classification in remote sensing," *Int. J. Remote Sens.*, vol. 26, no. 5, pp. 1007–1011, 2005.

[22] C. Huang, L. Davis, and J. Townshend, "An assessment of support vector machines for land cover classification," *Int. J. Remote Sens.*, vol. 23, no. 4, pp. 725–749, 2002.

[23] F. Isikdogan, A. C. Bovik, and P. Passalacqua, "Surface water mapping by deep learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 11, pp. 4909–4918, Nov. 2017.

[24] Y. Wang, J. Colby, and K. Mulcahy, "An efficient method for mapping flood extent in a coastal floodplain using landsat TM and DEM data," *Int. J. Remote Sens.*, vol. 23, no. 18, pp. 3681–3696, 2002.

[25] A. Rango and V. V. Salomonson, "Regional flood mapping from space," *Water Resour. Res.*, vol. 10, no. 3, pp. 473–484, 1974.

[26] F. Xie, M. Shi, Z. Shi, J. Yin, and D. Zhao, "Multilevel cloud detection in remote sensing images based on deep learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3631–3640, Aug. 2017.

[27] I. Masser, "Managing our urban future: The role of remote sensing and geographic information systems," *Habitat Int.*, vol. 25, no. 4, pp. 503–512, 2001.

[28] K. Karantzalos, D. Bliziotis, and A. Karmas, "A scalable geospatial web service for near real-time, high-resolution land cover mapping," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 10, pp. 4665–4674, Oct. 2015.

[29] K. Fotiadou, G. Tsagkatakis, and P. Tsakalides, "Deep convolutional neural networks for the classification of snapshot mosaic hyperspectral imagery," *Electron. Imag.*, vol. 2017, no. 17, pp. 185–190, 2017.

[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[31] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.

[32] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[33] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[34] A. Romero, C. Gatta, and G. Camps-Valls, "Unsupervised deep feature extraction for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1349–1362, Mar. 2016.

[35] R. Kemker and C. Kanan, "Self-taught feature learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2693–2705, May 2017.

[36] A. Mughees and L. Tao, "Hyperspectral image classification based on deep auto-encoder and hidden Markov random field," in *Proc. 13th Int. Conf. Natural Comput., Fuzzy Syst. Knowl. Discov.*, 2017, pp. 59–65.

[37] M. Ahmad, A. M. Khan, M. Mazzara, and S. Distefano, "Multi-layer extreme learning machine-based autoencoder for hyperspectral image classification," in *Proc. Int. Joint Conf. Comput. Vis., Imag. Comput. Graph.*, 2019, pp. 75–82.

[38] M. Ahmadlou *et al.*, "Flood susceptibility mapping and assessment using a novel deep learning model combining multilayer perceptron and autoencoder neural networks," *J. Flood Risk Manage.*, vol. 14, no. 1, 2020, Art. no. e12683.

[39] W. Sun and R. Wang, "Fully convolutional networks for semantic segmentation of very high resolution remotely sensed images combined with DSM," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 474–478, Mar. 2018.

[40] B. Cui, X. Chen, and Y. Lu, "Semantic segmentation of remote sensing images using transfer learning and deep convolutional neural network with dense connection," *IEEE Access*, vol. 8, pp. 116744–116755, 2020.

[41] S. Van Ackere, J. Verbeurgt, L. De Sloover, S. Gautama, A. De Wulf, and P. De Maeyer, "A review of the internet of floods: Near real-time detection of a flood event and its impact," *Water*, vol. 11, no. 11, 2019, Art. no. 2275.

[42] X. Wang *et al.*, "A robust multi-band water index (MBWI) for automated extraction of surface water from landsat 8 OLI imagery," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 68, pp. 73–91, 2018.

[43] J.-F. Pekel, A. Cottam, N. Gorelick, and A. S. Belward, "High-resolution mapping of global surface water and its long-term changes," *Nature*, vol. 540, no. 7633, pp. 418–422, 2016.

[44] A. Hollstein, K. Segl, L. Guanter, M. Brell, and M. Enesco, "Ready-to-use methods for the detection of clouds, cirrus, snow, shadow, water and clear sky pixels in Sentinel-2 MSI images," *Remote Sens.*, vol. 8, no. 8, 2016, Art. no. 666.

[45] Y. Li, S. Martinis, and M. Wieland, "Urban flood mapping with an active self-learning convolutional neural network based on TerraSAR-X intensity and interferometric coherence," *ISPRS J. Photogramm. Remote Sens.*, vol. 152, pp. 178–191, 2019.

[46] A. Refice *et al.*, "SAR and inSAR for flood monitoring: Examples with cosmo-skymed data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 7, pp. 2711–2722, Jul. 2014.

[47] F. Cian, M. Marconcini, P. Ceccato, and C. Giupponi, "Flood depth estimation by means of high-resolution SAR images and lidar data," *Natural Hazards Earth Syst. Sci.*, vol. 18, no. 11, pp. 3063–3084, 2018.

[48] M. Wieland and S. Martinis, "Large-scale surface water change observed by Sentinel-2 during the 2018 drought in Germany," *Int. J. Remote Sens.*, vol. 41, no. 12, pp. 4742–4756, 2020.

[49] D. Bonafilia, B. Tellman, T. Anderson, and E. Issenberg, "Sen1floods11: A georeferenced dataset to train and test deep learning flood algorithms for Sentinel-1," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2020, pp. 835–845.

[50] P. Akiva, M. Purri, K. Dana, B. Tellman, and T. Anderson, "H2O-Net: Self-supervised flood segmentation via adversarial domain adaptation and label refinement," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2021, pp. 111–122.

[51] M. Rahnemoonfar, T. Chowdhury, A. Sarkar, D. Varshney, M. Yari, and R. Murphy, "Floodnet: A high resolution aerial imagery dataset for post flood scene understanding," *IEEE Access*, vol. 9, pp. 89644–89654, 2020.

[52] M. Berger, J. Moreno, J. A. Johannessen, P. F. Levelt, and R. F. Hanssen, "ESA's Sentinel missions in support of Earth system science," *Remote Sens. Environ.*, vol. 120, pp. 84–90, 2012.

[53] I. Caballero, J. Ruiz, and G. Navarro, "Sentinel-2 satellites provide near-real time evaluation of catastrophic floods in the west Mediterranean," *Water*, vol. 11, no. 12, 2019, Art. no. 2499.

[54] M. C. Hansen and T. R. Loveland, "A review of large area monitoring of land cover change using landsat data," *Remote Sens. Environ.*, vol. 122, pp. 66–74, 2012.

[55] S. Schlaffer, P. Matgen, M. Hollaus, and W. Wagner, "Flood detection from multi-temporal SAR data using harmonic analysis and change detection," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 38, pp. 15–24, 2015.

[56] S. Long, T. E. Fatoyinbo, and F. Policelli, "Flood extent mapping for Namibia using change detection and thresholding with SAR," *Environ. Res. Lett.*, vol. 9, no. 3, 2014, Art. no. 035002.

[57] L. Pulvirenti, N. Pierdicca, M. Chini, and L. Guerriero, "An algorithm for operational flood mapping from synthetic aperture radar (SAR) data using fuzzy logic," *Natural Hazards Earth Syst. Sci.*, vol. 11, no. 2, pp. 529–540, 2011.

[58] D. Amitrano, G. Di Martino, A. Iodice, D. Riccio, and G. Ruello, "Unsupervised rapid flood mapping using Sentinel-1 GRD SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3290–3299, Jun. 2018.

[59] F. Gascon *et al.*, "Copernicus Sentinel-2A calibration and products validation status," *Remote Sens.*, vol. 9, no. 6, 2017, Art. no. 584.

[60] M. Kumar, "World geodetic system 1984: A modern and accurate global reference frame," *Mar. Geodesy*, vol. 12, no. 2, pp. 117–126, 1988.

[61] J. P. Snyder, *Map Projections-A Working Manual*, vol. 1395. Washington, DC, USA: U.S. Government Printing Office, 1987.

[62] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," 2017, *arXiv:1704.06857*.

[63] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 988–999, Sep. 1999.

[64] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. Int. Conf. Mach. Learn.*, vol. 30, no. 1, 2013, p. 3.

[65] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2015, pp. 234–241.

[66] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," 2015, *arXiv:1505.00853*.

[67] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[68] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google Earth Engine: Planetary-scale geospatial analysis for everyone," *Remote Sens. Environ.*, vol. 202, pp. 18–27, 2017. [Online]. Available: https://doi.org/10.1016/j.rse.2017.06.031

[69] X. Han-Qiu, "A study on information extraction of water body with the modified normalized difference water index (MNDWI)," *J. Remote Sens.*, vol. 5, pp. 589–595, 2005.

Georgios I. Drakonakis received the Diploma (M. Eng.) in rural, surveying and geoinformatics engineering from the National Technical University of Athens, Athens, Greece, in 2016, and the M.Sc. degree in computer science and engineering from Computer Science Department, University of Crete, Heraklion, Greece, in 2021.

Since June 2020, he has been with the Institute of Computer Science of the Foundation for Research and Technology Hellas (FORTH-ICS), Heraklion, as a Graduate Research Assistant with the Signal Processing Laboratory. His main research interests include machine/deep learning applications in the fields of remote sensing, computer vision, and photogrammetry

Grigorios Tsagkatakis received the B.Sc. and M.Sc. degrees in electronics and computer engineering from the Technical University of Crete, Chania, Greece, in 2005 and 2007, respectively, and the Ph.D. degree in imaging science from the Center of Imaging Science, Rochester Institute of Technology, Rochester, NY, USA, in 2011.

He is an Associate Researcher with the Signal Processing Laboratory, Institute of Computer Science, Foundation for Research and Technology Hellas (FORTH-ICS), Heraklion, Greece. He is currently a Marie Skłodowska-Curie Fellow with the University of Southern California and FORTH on the CALCHAS project. His research interests include the domains of signal/image processing and machine learning for remote sensing and astrophysics.

Konstantina Fotiadou received the B.Sc. degree in applied mathematics from the Department of Applied Mathematics, University of Crete (UoC), Crete, Greece, in 2011, the M.Sc. and Ph.D. degrees in computer science from the Computer Science Department, UoC in 2014 and 2019, respectively.

Since 2012, she has been a Research Assistant with the Signal Processing Laboratory, Institute of Computer Science, Foundation for Research and Technology Hellas, Heraklion, Greece. From the past four years, she has been working in the industry domain in the fields of machine learning and big data analytics. Her main research interests include machine learning, image processing, and computational photography applications with an emphasis on satellite imaging systems.

Panagiotis Tsakalides (Member, IEEE) received the Diploma in electrical engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 1990, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 1995.

He is currently a Professor in computer science with the University of Crete, Heraklion, Greece, and the Head with the Signal Processing Laboratory, Institute of Computer Science, Foundation for Research and Technology Hellas (FORTH-ICS), Heraklion. He has an extended experience of transferring research and interacting with the industry. During the last ten years, he has been the Project Coordinator in seven European Commission and 12 national research and innovation projects totaling more than 5.5 Meuros in actual funding for FORTH-ICS and the University of Crete. His research interests include the fields of statistical signal processing and machine learning with emphasis in non-Gaussian estimation and detection theory, sparse representations, and applications in sensor networks, audio, imaging, and multimedia systems. He has co-authored more than 200 technical publications in these areas.