



## Article

# Synergistic Use of Geospatial Data for Water Body Extraction from Sentinel-1 Images for Operational Flood Monitoring across Southeast Asia Using Deep Neural Networks

Junwoo Kim , Hwison Kim, Hyungyun Jeon, Seung-Hwan Jeong, Juyoung Song, Suresh Krishnan Palanisamy Vadivel and Duk-jin Kim \*

School of Earth and Environmental Sciences, Seoul National University, Seoul 08826, Korea; darkcroa@snu.ac.kr (J.K.); rlaglnthd1@snu.ac.kr (H.K.); hgjeon1993@snu.ac.kr (H.J.); jheon2743@snu.ac.kr (S.-H.J.); %6daniel@snu.ac.kr (J.S.); krishnan92@snu.ac.kr (S.K.P.V.)

\* Correspondence: djkim@snu.ac.kr

**Abstract:** Deep learning is a promising method for image classification, including satellite images acquired by various sensors. However, the synergistic use of geospatial data for water body extraction from Sentinel-1 data using deep learning and the applicability of existing deep learning models have not been thoroughly tested for operational flood monitoring. Here, we present a novel water body extraction model based on a deep neural network that exploits Sentinel-1 data and flood-related geospatial datasets. For the model, the U-Net was customised and optimised to utilise Sentinel-1 data and other flood-related geospatial data, including digital elevation model (DEM), Slope, Aspect, Profile Curvature (PC), Topographic Wetness Index (TWI), Terrain Ruggedness Index (TRI), and Buffer for the Southeast Asia region. Testing and validation of the water body extraction model was applied to three Sentinel-1 images for Vietnam, Myanmar, and Bangladesh. By segmenting 384 Sentinel-1 images, model performance and segmentation accuracy for all of the 128 cases that the combination of stacked layers had determined were evaluated following the types of combined input layers. Of the 128 cases, 31 cases showed improvement in Overall Accuracy (OA), and 19 cases showed improvement in both averaged intersection over union (IOU) and F1 score for the three Sentinel-1 images segmented for water body extraction. The averaged OA, IOU, and F1 scores of the ‘Sentinel-1 VV’ band are 95.77, 80.35, and 88.85, respectively, whereas those of ‘band combination VV, Slope, PC, and TRI’ are 96.73, 85.42, and 92.08, showing improvement by exploiting geospatial data. Such improvement was further verified with water body extraction results for the Chindwin river basin, and quantitative analysis of ‘band combination VV, Slope, PC, and TRI’ showed an improvement of the F1 score by 7.68 percent compared to the segmentation output of the ‘Sentinel-1 VV’ band. Through this research, it was demonstrated that the accuracy of deep learning-based water body extraction from Sentinel-1 images can be improved up to 7.68 percent by employing geospatial data. To the best of our knowledge, this is the first work of research that demonstrates the synergistic use of geospatial data in deep learning-based water body extraction over wide areas. It is anticipated that the results of this research could be a valuable reference when deep neural networks are applied for satellite image segmentation for operational flood monitoring and when geospatial layers are employed to improve the accuracy of deep learning-based image segmentation.

**Keywords:** deep learning; U-Net; semantic segmentation; water body extraction; Sentinel-1; geospatial data



**Citation:** Kim, J.; Kim, H.; Jeon, H.; Jeong, S.-H.; Song, J.; Vadivel, S.K.P.; Kim, D.-j. Synergistic Use of Geospatial Data for Water Body Extraction from Sentinel-1 Images for Operational Flood Monitoring across Southeast Asia Using Deep Neural Networks. *Remote Sens.* **2021**, *13*, 4759. <https://doi.org/10.3390/rs13234759>

Academic Editor: Bruno Adriano

Received: 5 October 2021

Accepted: 20 November 2021

Published: 24 November 2021

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Floods, which make up 52.1% of natural disasters in frequency, occur unexpectedly and cause devastating damage over broad areas [1–3]. It was reported that hydrological disasters, including floods, were responsible for 19.3% of the total damage caused by natural disasters and 20.4% of the number of total victims [4]. Thus, flood monitoring,

including flooded area extraction and estimation, is critical to respond to, and recover from, such damage. Satellite remote sensing techniques have been used to estimate flooded areas, as they can provide visual information over wide areas [5–7] yet timely monitoring, and estimating inundated areas in flood situations have been limited by satellite data acquisition and analysing such data that includes the accuracy of classification for extracting flooded areas from available satellite data. Poor classification accuracy could cause more flood damages, as such damages depend heavily on the quality of flood forecasting, flood area estimation, and settlement patterns [8].

As acquiring optical satellite data is mainly limited by natural constraints, e.g., weather condition and cloud cover, during the rainy season [9], spaceborne Synthetic Aperture Radar (SAR) data have been considered to be suitable for flood monitoring [10–12]. Such data are almost independent of cloud cover, sunlight, and other weather conditions. For SAR image classification, backscatter intensity, polarimetric parameters, and interferometric coherence information have been mainly exploited [13]. For water body extraction using a single image, SAR data has been analysed with supervised or unsupervised classification methods that include thresholding, distance-based classification, decision tree/rule-based classification, image clustering approaches, and machine learning techniques [5,12–14]. Yet, threshold values and classification rules that are determined for a certain region were found to be difficult to be applied to other SAR images or regions [15], and accurately extracting flooded areas from SAR images is constrained by objects in the images that have similar reflectance values, such as roads, airports, mountainous areas, and radar shadow [16–18].

It was reported that combining many layers that can provide more information on targeted areas may allow for the discrimination of objects with similar backscattering values [19], and the accuracy of extracting water body from satellite data can thus be improved by the combined use of remote sensing data and other ancillary data, such as digital elevation model (DEM) products and digital topographic maps [20–22]. Flooding potential is determined by various conditions of river basins, including the characteristics of the climatic system and drainage basin conditions [23,24]. For predicting flood-prone areas by analysing spatial data, remotely sensed satellite data have been used in combination with geospatial data, such as DEM, Slope, and Aspect [2]. Yet, when analysing satellite data for water body extraction for flood monitoring, it seems that such factors have not been fully reflected in the processes as a form of ancillary data. This means that the effects of using geospatial layers for water body extraction remain uncertain, and further research is thus needed.

In image classification and segmentation, previous research showed that deep learning models outperform aforementioned traditional classification methods [25–28]. Deep learning methods such as convolutional neural networks (CNNs) have been widely applied for land cover classification, road extraction, ship detection, and other domains. Yet, even advanced deep learning methods have difficulties in discriminating water bodies in SAR images during image classification processes, due mainly to the misclassification of objects that have similar backscattering values. It could be assumed that the backscattering values of SAR data, which may contain insufficient information for clearly discriminating water bodies from images, could be supported by other data. Yet, only a very limited number of studies on such purpose have been conducted using SAR data [17]. For deep learning-based flood monitoring models, although it was reported that geospatial datasets could be used for spatial prediction of floods using machine learning approaches [29], the actual influences of such datasets on model performance are still poorly understood. To take account of such information in the geospatial datasets, existing deep learning models need to be optimised for water body extraction from satellite data. Yet, existing research focused mainly on producing more training data or on advancing network architecture to improve classification accuracy. In addition, even existing deep learning models have not been thoroughly tested and optimised for operational flood monitoring, as most of the results are confined to analysing specific bands of available satellite images or research sites [30]. Considering the existing literature, the synergistic use of geospatial data in deep

learning-based water body extraction over wide areas has yet to be demonstrated. To the best of our knowledge, this is the first research that demonstrates the synergistic use of geospatial data in deep learning-based water body extraction over wide areas. The aim of this research is to present a new deep learning-based flood monitoring model that has better predictive ability by testing the effectiveness of combining geospatial layers. For that, we conducted intensive and comprehensive experiments for examining synergistic use of geospatial data for water body extraction from Sentinel-1 data using deep learning and demonstrated that the accuracy of water body extraction from Sentinel-1 data is improved by utilising such data. In the process, we also constructed a geospatial database that contains structured and unified Shuttle Radar Topography Mission (SRTM) DEM, Slope, Aspect, Profile Curvature (PC), Topographic Wetness Index (TWI), Terrain Ruggedness Index (TRI), and Buffer layers for the Southeast Asia region and presented a novel flood area monitoring model based on deep learning, which is automated and optimised to the region for operational purposes.

This paper consists of six sections. Detailed explanations on producing input data and developed methods for this research are presented in the Sections 2 and 3, and then the experimental results of evaluating the effectiveness of using geospatial layers for water body extraction are presented in Section 4. The results are discussed in the discussion section, in addition to the relationship with other research, wider implications of the research, and limitations of the research, before presenting concluding remarks in Section 6.

## 2. Producing Input Data and Geospatial Database

### 2.1. Pre-Processing and Modification of Input Data

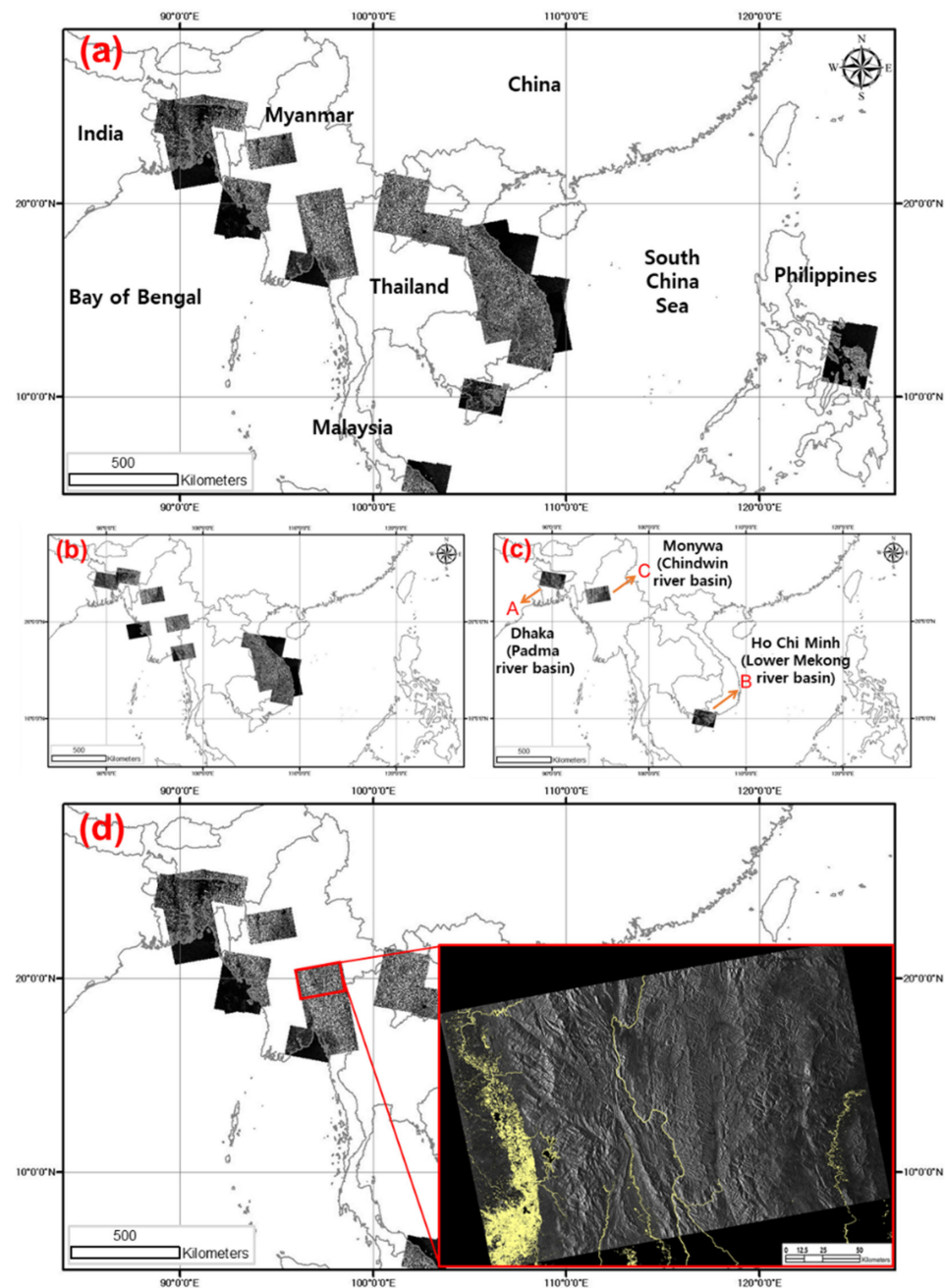
#### 2.1.1. Sentinel-1 and Ground Truth Data for the Southeast Asia Region

As almost all the countries in Southeast Asia suffer from floods during the rainy season, the Southeast area region was selected as the research area for this research. Due to the limited financial resources, infrastructure, and technological means to respond to floods, the impact of floods on countries in the region tends to be more severe than for other countries [2].

Sentinel-1 data and United Nations Satellite Centre (UNOSAT) flood datasets for the region were used as the main input data (Figure 1). UNOSAT provides analysed flood boundaries and flood extent in the shapefile spatial data format during and after flood events [31,32]. The UNOSAT flood datasets were produced with a thresholding method and validated through manual visual inspection and modification, which are freely available through the UNOSAT flood portal for various purposes (<http://floods.unosat.org/geoportal/catalog/search/search.page>, accessed on 18 September 2020), including flood model calibration and supporting post-disaster field assessments. Based on the locations and dates of the flood data, corresponding Sentinel-1 images were obtained from the Copernicus Open Access Hub (<https://scihub.copernicus.eu/dhus/#/home>, accessed on 21 September 2020).

Since empirical experiments showed that classification accuracy was not significantly improved by using other SAR information, such as VH polarisation and incident angle, as additional bands, and that VV polarization showed higher accuracy than VH polarization in water body classification, VV band was selected as the main input satellite data [33]. To use the data for extracting water bodies from multiple Level-1 Ground Range Detected (GRD) Sentinel-1 images, which were acquired with interferometric wide (IW) mode at a  $20\text{ m} \times 5\text{ m}$  spatial resolution, digital numbers of the images were converted into Sigma0 by performing radiometric calibration [17]. Applied pre-processing procedures of the Sentinel-1 images include 'Remove GRD border noise', 'Radiometric Calibration' (VV), 'Speckle Filtering', and 'Terrain Correction'. Therefore, the output SAR data that were pre-processed for further analyses have amplitude values in linear scale, with pixel spacing of  $10\text{ m} \times 10\text{ m}$ . A total of 50 scenes of Sentinel-1 images acquired between 2015 and 2018 were downloaded and pre-processed for visual inspection, which correspond to the dates and locations of the vector data of the UNOSAT flood datasets (Figure 1d). After

completing the process, 30 scenes were finally used as input data to train and validate deep learning models.



**Figure 1.** Sentinel-1 images used in this research and an example of UNOSAT flood data ((a) 50 scenes of Sentinel-1 images acquired between 2015 and 2018 for visual inspection, (b) 30 scenes of Sentinel-1 SAR images for training deep learning networks and evaluating model performance, (c) 3 scenes (A–C) for inference and validation, (d) flood data in shapefile format (yellow) overlaid on a Sentinel-1 image).

In accordance with geographical features and the locations of ground truth data, three areas were selected for accuracy assessment, i.e., Padma river basin (A) in Bangladesh, Lower Mekong river basin (B) in Vietnam, and Chindwin river basin (C) in Myanmar, which are reported to often experience flash floods during rainy seasons and have different topographical features. For reliable evaluation of model performance and segmentation results, of the 30 Sentinel-1 scenes, three scenes for the three regions were used for inference and validation (No.1–3 in Table 1).



**Table 1.** Examples of Sentinel-1 images for training and inference of deep learning models.

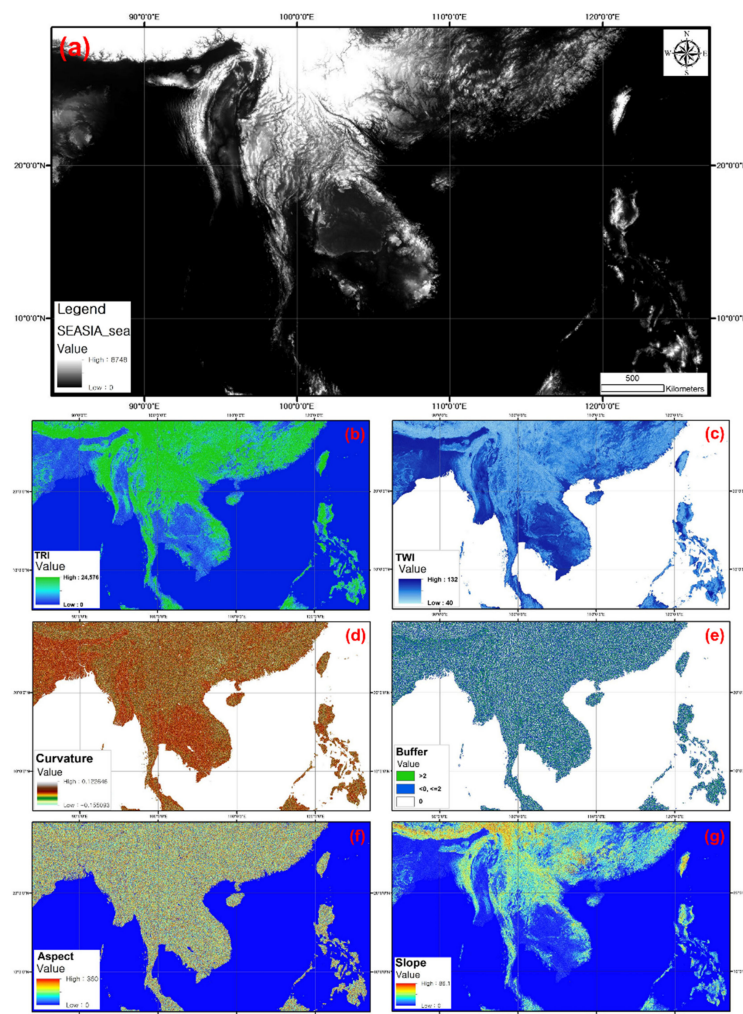
No.	Satellite	Type/Mode	Acquisition Time (UTC)	Product ID	Usage
1	Sentinel-1A	GRDH/IW	30 June 2016 23:55:28–23:55:53	0126A4_AB04	Inference
2	Sentinel-1A	GRDH/IW	7 November 2017 22:45:31–22:45:56	0206FC_1842	Inference
3	Sentinel-1A	GRDH/IW	18 July 2015 11:47:20–11:47:45	00942A_517D	Inference
4	Sentinel-1A	GRDH/IW	30 July 2017 11:04:20–11:04:45	01DA46_8ADC	Training
5	Sentinel-1A	GRDH/IW	15 June 2018 23:47:18–23:47:43	026C2D_EC7F	Training
6	Sentinel-1A	GRDH/IW	25 July 2018 11:04:26–11:04:51	027D94_F52C	Training
7	Sentinel-1A	GRDH/IW	11 July 2015 11:54:34–11:54:59	009133_64FC	Training
8	Sentinel-1A	GRDH/IW	6 August 2015 11:37:30–11:37:55	009BED_DE92	Training
9	Sentinel-1A	GRDH/IW	11 August 2015 11:47:21–11:47:46	009DE4_C4E2	Training
10	Sentinel-1A	GRDH/IW	6 August 2015 11:37:55–11:38:20	009BED_FB1C	Training
11	Sentinel-1A	GRDH/IW	24 July 2016 23:55:29–23:55:54	013213_0790	Training
12	Sentinel-1A	GRDH/IW	12 October 2016 22:51:27–22:51:52	015847_FB42	Training
13	Sentinel-1A	GRDH/IW	29 July 2018 22:44:19–22:44:44	027F8D_B944	Training
14	Sentinel-1A	GRDH/IW	13 July 2018 11:04:25–11:04:50	02780D_7D6F	Training
15	Sentinel-1A	GRDH/IW	13 December 2016 22:36:07–22:36:32	01747F_68A3	Training
16	Sentinel-1A	GRDH/IW	1 December 2016 22:36:08–22:36:33	016EE9_2752	Training

### 2.1.2. Data Modification and Producing Label Data

Although the original UNOSAT flood extent dataset has been verified by intensive data cleaning and manual visual inspection, some mismatching between the satellite data and the flood extent vector data was found during the labelling processes. Among the initial 30 input data pairs, only 12 input data pairs were selected by intensive visual interpretation. In the following cases, we excluded data pairs from input data: (1) the extent of an SAR image and label data was not matched; (2) the quality of the label data was low, as, for instance, water and non-water at flat beaches were mislabelled. We modified and included some data pairs into input data for the following cases: (1) the label was accurate but the ocean was not labelled as water. For this case, we modified the ocean part into water, based on the ideas of sea masking. (2) Label shapefiles only exist for some parts of SAR images. In such cases, the extent of SAR images was wider than that of label data, so stacked SAR images were cropped on the basis of the extent of label shapefiles. Lastly, rasterizing label shapefiles with the same spatial resolution of VV images was needed. We designed a code that created an empty field, which had the same area and a spatial resolution with Sentinel-1 VV band images, and executed in a way where pixels overlapped on shapefiles had meaningful values. Therefore, corresponding flood extent boundaries in the shapefile format were converted into binary raster data of 0 (non-water) and 1 (water), which are used as label data for training and ground truth data for validation.

### 2.1.3. Building a Geospatial Database

Geospatial data have been used to predict floods, manage flood emergencies, and produce flood-related maps, including floods risk and susceptibility maps [29,34]. Relying on the wealth of literature in hydrology and remote sensing sciences, geospatial layers that have a possibility of providing topo-hydrographic information were selected and produced for deep learning-based water body extraction [35–38], which had been evaluated to be related to the feature of flooded areas [39] and to be geographical variables having influence on river flood occurrence [19,23]. These layers include Digital Elevation Model (DEM), Slope, Aspect, Terrain Ruggedness Index (TRI), Profile Curvature (PC), Topographic Wetness Index (TWI), and ‘distance from water bodies’ (i.e., Buffer), which were expected to improve discriminating land surface and terrain effects that are caused by various physical characteristics (Figure 2).



**Figure 2.** Geospatial layers and digital map products produced for this research: (a) Digital Elevation Model (DEM) for Southeast Asia, (b) Terrain Ruggedness Index (TRI), (c) Topographic Wetness Index (TWI), (d) Profile Curvature (PC), (e) Buffer, (f) Aspect, and (g) Slope). All values shown in the legends are before normalisation.

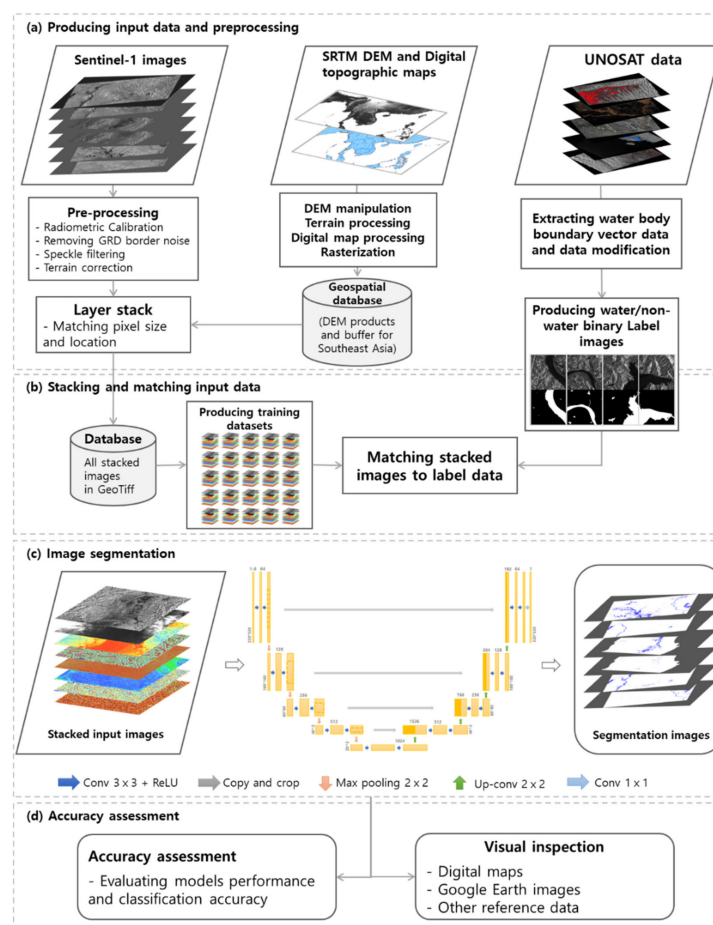
To produce geospatial layers as additional input layers, 1-ArcSecond Global Shuttle Radar Topography Mission (SRTM) DEM tiles, which were freely available through the USGS webpage (<https://earthexplorer.usgs.gov/>, accessed on 22 September 2020), were downloaded, mosaicked, gap-filled, and exported in the EPSG:4326 (WGS 84 latitude/longitude) coordinate at 1-arcsecond (around 30 m) spatial resolution. Using the mosaicked DEM layer, Slope, Aspect, and PC layers were produced at the same spatial resolution and coordinates.

In addition, TRI, which indicates terrain heterogeneity, and TWI layers were also generated based on [40–43], respectively. To produce a Buffer layer, which shows the proximity to rivers, digital topographic maps for the research area were merged, and then the water layers were extracted based on the attributes of the vector data. Before stacking up, the Buffer layer required an additional processing step, as the produced Buffer rings around river mouths protruded into coast lines. We masked out sea to prevent errors around the mouths of rivers. The size of each layer, which was produced for the extent of the whole Southeast Asia region, is around 50 Gb, and the size of all the layers saved in a geospatial database is thus around 400 Gb, excluding satellite images and label data.

### 3. Development of a Deep Learning-Based Water Body Extraction Model

#### 3.1. Deep Learning-Based Water Body Extraction Model for Operational Flood Monitoring

For this research, a deep learning-based water body extraction model for operational flood monitoring across the Southeast Asia region was presented, as shown in Figure 3. The model consists of four steps, including (a) producing input data and pre-processing, (b) stacking and matching input data, (c) semantic image segmentation, and (d) accuracy assessment (Figure 3). The first step is explained above, and detailed methods for steps two, three, and four of the model are explained in the following sections. Using the model, all 128 cases that were the possible combinations of satellite data and geospatial layers were examined to evaluate the effectiveness of adding ancillary layers and to evaluate model performance and segmentation accuracy.



**Figure 3.** Deep learning-based water body extraction model. The model consists of (a) producing input data and pre-processing, (b) stacking and matching input data, (c) image segmentation with a deep neural network, and (d) accuracy assessment. The architecture of the deep learning model in the figure is built on U-Net [44].

### 3.1.1. Customisation and Optimisation of the Deep Neural Network

For operational flood monitoring by extracting flood damage information from satellite data, Convolutional Neural Network (CNN)-based deep learning methods may not be effective if spatial resolutions of input satellite data are lost in the process of down-sampling with pooling layers. For this reason, instance segmentation is required to perform fine-grained inference, and by adopting fully connected layers, semantic segmentation based on fully convolutional networks (FCNs) was able to achieve pixel-wise labelling [45]. Yet, the FCNs have limitations in achieving highly accurate pixel-wise class labelling, due to difficulties in reconstructing the non-linear structures of object boundaries [28,46].

As the main purpose of this research is to present a reliable deep learning-based flood monitoring model that has better predictive ability by testing the effectiveness of geospatial layers, for this research, the U-Net architecture, which was developed for semantic segmentation with a relatively small amount of training data [44,47] and has thus been widely used to classify urban features with shorter training times and to minimise the loss of spatial resolution [48], was customised and optimised to utilise Sentinel-1 data and seven different types of geospatial layers. Unlike the original U-Net, our model can take geo-located and multi-layered SAR images and other ancillary data in the GeoTiff format as input data, and the optimised deep network presented for this research does not lose any spatial resolution or location information of the multi-modal input data. In addition, the size of the input data for inference is significantly larger than the U-Net to reduce processing times and to achieve the seamless merging of segmented image patches for inference.

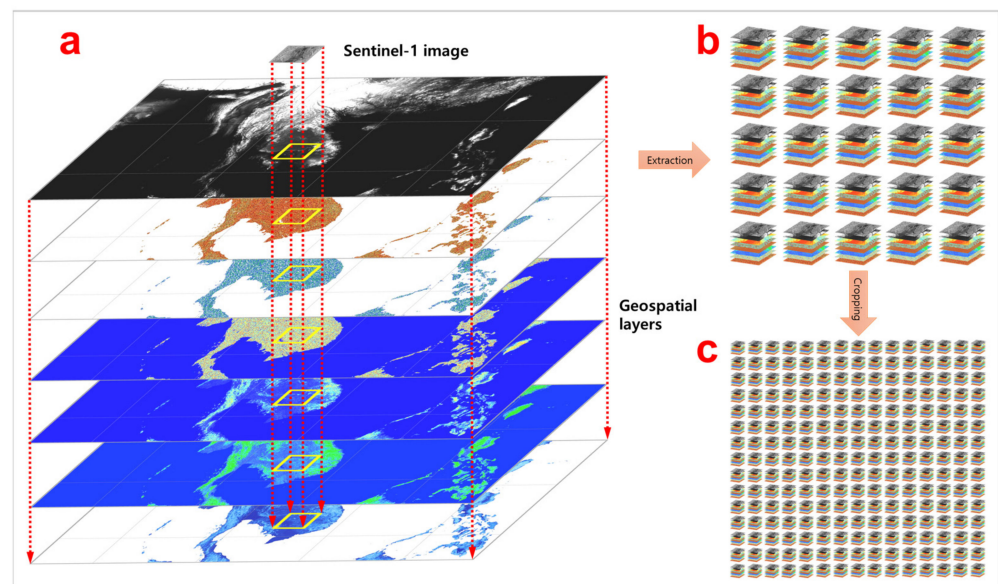
The architecture of the model for semantic image segmentation consists of 18,  $3 \times 3$  2D convolution layers, which have the Rectified Linear Unit (ReLU) as an activation function, and one  $1 \times 1$  2D convolution layer, which has Sigmoid as an activation function. The convolution layers in the contracting and expanding paths are followed by four  $2 \times 2$  2D max\_pooling layers and four  $2 \times 2$  up-sampling layers, which perform nearest interpolation. The contracting and expanding processes are combined by concatenation, and padding is added to the convolution layers to preserve the spatial resolution of input data. The number of total trainable parameters for the model is 31,379,521. The architecture of the model is presented in Figure 3c, and the hyper-parameters for training models are explained in detail in Section 3.1.3.

### 3.1.2. Stacking Input Data for Matching Layers and Normalisation

The last procedure to generate georeferenced input data for model training was stacking eight separate pieces of data into one single file with eight layers and normalising layer values. Since the geospatial layers were georeferenced images at various pixel sizes, geographic extents, datatypes and data formats, and deep learning models are trained based on the information in the pixels, the main points of stacking input data were matching such factors. The geospatial layers were georeferenced to the WGS 84 latitude/longitude coordinate system but had different pixel spacing and extents. Therefore, it was necessary to combine the other layers into a multi-band single image, and the first layer determined the output size and extent of the layer stack as a reference layer. To achieve this, we clipped input layers to input SAR data extent and stacked using Geospatial Data Abstraction Library (GDAL) libraries. In order to do that, we firstly defined the common grid from the input SAR data to which the auxiliary layers will be resampled and reprojected. Secondly, each auxiliary layer is clipped with those common extents. Following that, we resampled the geospatial layers that have different spatial resolutions to the target resolution and different extents to the target extent (here, input SAR data extent and resolution). To accurately clip without the influence of raster properties, a new python code was devised using a shapefile made with the VV raster pixel coordinate. As SAR images were pre-processed at 10 metre pixel spacing, the other bands were also interpolated at the same pixel size. Finally, the resampled datasets were stacked into a single dataset with separate eight bands using the gdal\_merge algorithm. Through the procedure, the final input data



that consist of eight raster layers with the same pixel size and coordinate system were produced for training (Figure 4a).



**Figure 4.** Producing and matching input data for model training and inference ((a) stacking geospatial layers for matching, (b) extracting the stacked geospatial layers for Sentinel-1 images (around  $30,000 \times 20,000$ ), and (c) producing  $320 \times 320 \times 8$  images for training and inference).

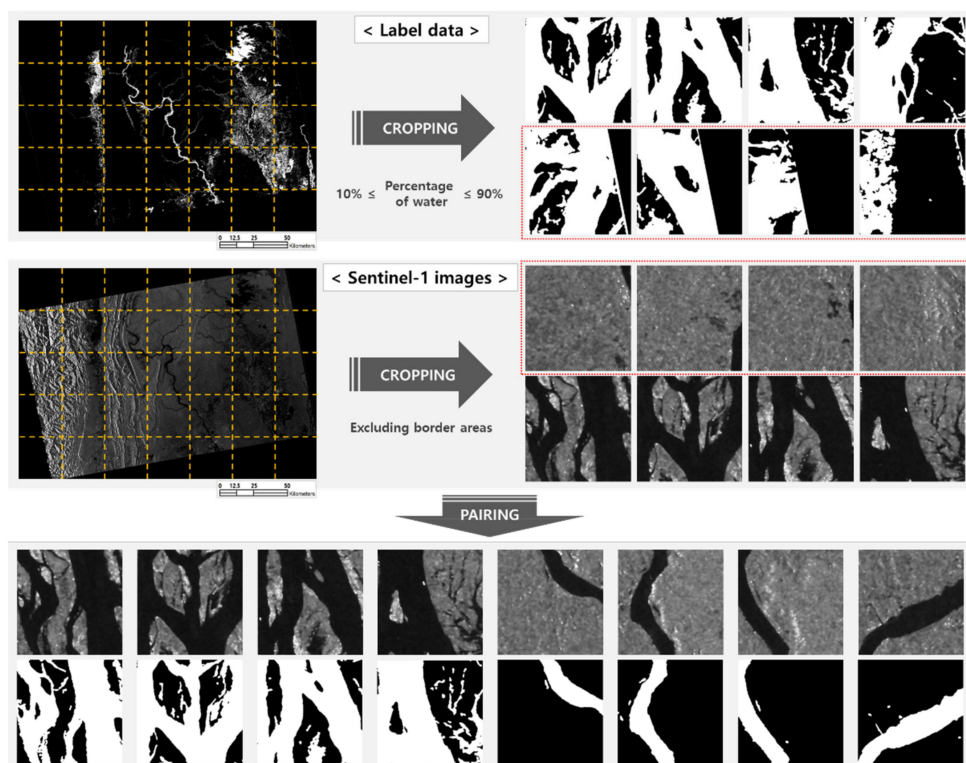
To train deep neural networks for semantic segmentation, all of the satellite images and geospatial data were normalised and standardised. To obtain more accurate models and reduce processing time, we reclassified all of the input layers into values between 0 and 1 by considering standard deviation and the histograms of the layer values. The normalisation was performed based on more than 300 times of experiments conducted to evaluate the effects of normalisation. To remove speckle values, the values of the VV layer that get out of 0 to 1 were adjusted into 0 and 1. The values of Slope and Aspect had a range by definition, so we divided the values by theoretical maximum value to make a range of 0 to 1. Therefore, VV, Slope, and Aspect layers contain continuous values between 0 and 1. Based on the real value ranges of the whole Southeast Asia region, the other bands were reclassified into discrete values of 0 to 1 (Table 2) after evaluating the effects of employing discrete values. The stacked dataset that was matched and normalised is exported to a database in geotiff format to be analysed with the deep learning algorithm as described in the following section.

### 3.1.3. Model Training

Deep learning models were trained from scratch using the stacked datasets produced for this research. Before starting model training, band extract was performed. Combinations of geographic information bands were automatically selected when the number of elements was set as the input for systematic evaluation. When a combination to test is chosen, stacked input data were copied only with a VV band and selected geospatial bands (Figure 4b). Copied and stacked images were cropped into  $320 \times 320$  pixels (Figure 4c) and saved only if the border of the VV image were not included. In addition, rasterized label data were cropped into  $320 \times 320$  pixels and saved only if the proportion of water pixels was between 10% and 90% (Figure 5). We matched cropped stacked images to corresponding cropped label images and filtered out based on the two conditions. The final number of pairs of SAR images and corresponding label images was 4326.

**Table 2.** Matching spatial resolution and normalisation of Sentinel-1 VV and geospatial layers.

Layer Order	Layer Name	Pixel Size (m)	Resampled Pixel Size (m)	Value Range	Normalised Value Range
1	Sentinel-1 data (VV)	10	10	0–1	0–1
2	SRTM Digital Elevation Model (DEM)	30	10	0–8220	0–1 (0, 0.2, 0.4, 0.6, 0.8, 1)
3	Slope	30	10	0–86.1	0–1
4	Aspect	30	10	0–360	0–1
5	Profile Curvature (PC)	30	10	−0.155093–0.122646	0–1 (0, 0.5, 1)
6	Terrain Wetness Index (TWI)	500	10	40–132	0–1 (0, 0.2, 0.4, 0.6, 0.8, 1)
7	Distance from water (Buffer)	30	10	0–3	0–1 (0, 0.5, 1)
8	Terrain Ruggedness Index (TRI)	30	10	0–24,576	0–1 (0, 0.2, 0.4, 0.6, 0.8, 1)



**Figure 5.** Examples of cropping, selecting, and pairing training data for testing the flood monitoring model developed for this research (452,000 pairs of label data (top) and Sentinel-1 image (bottom) patches were initially produced, and part of them were excluded based on the principle of water rate and the existence of borders in image patches).

The customised and optimised U-Net for water body extraction was trained for all of the possible combinations of the Sentinel-1 images and geospatial layers. For deep learning, hyper-parameters need to be tuned, which could often be set through heuristic ways, and it thus requires repetitive empirical tests. Repetitive systematic experiments were performed to decide optimal hyper-parameters for the model with minimum loss. The selected hyperparameters are as follows (Table 3): The activation function used for

each layer is RELU, and the activation function for the output layer is sigmoid. The kernel size of the convolution layer is  $3 \times 3$ ,  $2 \times 2$ , which is used for upsampling and maxpooling layers, and  $1 \times 1$  is used for the output convolution layer. To maintain the size of the input layer, the stride is fixed to  $1 \times 1$  and the same padding is used. Adadelata is used as an optimiser with an initial learning rate of 1 and a decay rate of 0.95.

**Table 3.** Hyper-parameters for training the deep learning models and inference.

Hyper-Parameters for the Deep Neural Network	
Kernel size (upsampling/output)	$3 \times 3/2 \times 2$
stride/padding	$1 \times 1$ /same
Maxpooling	$2 \times 2$
Activation function	RELU/sigmoid (output layer)
Learning rate/decay rate	Adadelata optimizer 1/0.95
Validation frequency	Every 20 iterations
Epoch/iteration	1000/170 per epoch
Early stopping	Validation criterion (No improvement of loss for five epochs)
Batch size	16
Patch size/channels	$320 \times 320/1-8$
Pair numbers/Water body rate	4326/0.1–0.9

Among 4326 input pairs, we randomly split them into three sets, training, validation, and test datasets [49]. The selected ratio of dataset split is 60%, 20%, and 20%. A minibatch size for training is 16 patches and iterated over the whole training dataset 170 times. To prevent overfitting, and to minimise training time, the early stopping function was adopted in the training process. When validation loss does not improve for five continuous epochs, the weights of the model with the minimum validation loss value are automatically saved for the segmentation of new Sentinel-1 images.

### 3.1.4. Inference

For prediction, the same preprocessing and band extraction procedures were applied to inference data. As mentioned in the model training section, combinations of geospatial bands were automatically selected and copied into a separate folder. Those copied images were cropped into patches and reclassified based on the same criteria for training input data. The patches were predicted into binary outputs using trained models. Values of outputs are 1 and 2, which are defined as non-water and water, respectively.

Meanwhile, VV images without other geospatial bands were copied into a result folder on which the predicted output was overlaid. The purpose of this procedure is to make the output georeferenced and to remove the borders of VV images. As SAR images are usually inclined rectangular shape, there are margins that have no information. Cropped patches had no geographic coordinates but predicted in the same way whether they were the part of the border or not. The cropped output combined in an order and overlaid on copied VV images only if the pixels in the patches were not part of margin. As a result, the outputs had coordinates information that was predicted with meaningful data.

A new code was developed to improve the inference procedure, which includes modifying: (1) the size of cropping for inference data and (2) the padding size for combining cropped patches. First, the size of cropping for inference data is different from training data. The size for trained patches is  $320 \times 320$  pixels, but that for inference patches is  $3040 \times 3040$  pixels. The reason for the different patch sizes is that increasing the patch size leads to a reduction in inferencing time. More importantly, there was an essential precondition to increase patch size that the quality of inference should be maintained. We

tested various cropping sizes from smaller to bigger than that for training and verified that increasing size does not interrupt the quality of outputs through visual interpretation and evaluating numerical indicators, such as accuracy, precision, IOU, recall, and F1 score. Second, for mosaicking predicted patches, the concept of overlapping of patches was not used. Without padding, inference time became shorter and duplication errors on borders were completely removed.

All experiments for testing the effects of geospatial data on image segmentation and the training and validating of the deep neural network were conducted with a GPU server that has four Nvidia GeForce RTX 3090 GPUs, which have 24 Gb memories and a 260 Gb RAM. The server also has 72 Intel(R) Xeon(R) Gold 6240 CPU @ 2.60 GHz CPUs, one SSD, and one 11TB HDD. The versions for NVIDIA driver, CUDA, and Python are 470.57.02, 11.4, and 3.9.4, respectively.

### 3.2. Accuracy Assessment

The performance of deep learning architectures can be evaluated with criteria such as Overall Accuracy (OA) of pixel-wise classification, time, and memory usage [28]. For evaluating image classification accuracy, OA has been commonly used, which is the proportion of correct predictions among the total number of predictions. Although, for supervised learning, using a confusion matrix for evaluating classification accuracy is common as a statistical indicator, accuracy metrics could mislead if the class representation for evaluation is unbalanced [28]. Therefore, in addition to OA, precision, recall, mean intersection over union (IOU), and F1 score were selected for a more precise model and inferred output evaluation [2,48]. Precision is the proportion of correct water pixels among the predicted water pixels, while Recall is the proportion of correct water pixels among the correct predictions. Mean IOU indicates the degree to which predicted bounding boxes are overlapped on ground truth bounding boxes. F1 score is derived with the ‘harmonic mean of precision and recall’ (see, [50]). The confusion table used in this research and the mathematical formulas for the five criteria are shown in Table 4.

**Table 4.** Criteria and equations for pixel-wise evaluation and accuracy assessment for output images.

Confusion Matrix for Pixel-Wise Evaluation			
Label Class	Predicted Class	Water	Non-Water
	Water		True Positive (TP)
Non-water		False Positive (FP)	True Negative (TN)
Formulas for Accuracy Assessment of Output Images			
Overall accuracy (OA)		$Accuracy (A) = \frac{TP + TN}{TP + TN + FP + FN}$	
Precision		$Precision (P) = \frac{TP}{TP + FP}$	
Recall		$Recall (R) = \frac{TP}{TP + FN}$	
Intersection over union (IOU)		$IOU = \frac{TP}{TP + FP + FN}$	
F1 Score		$F1\ score = \frac{2 * R * P}{R + P}$	

Trained models and segmentation accuracy were evaluated with those five-confusion matrices and the binary cross-entropy confusion function. As aforementioned, input data for training were divided into train: validation: test by a ratio of 6:2:2. Testing data were used to calculate the values of loss and the confusion matrix to show how the model is well-trained. The mathematical formulation of binary cross entropy is:

$$C = -\frac{1}{n} \sum_{i=1}^n [y_i \ln a_i + (1 - y_i) \ln(1 - a_i)] \quad (1)$$



where,  $C$  is the cross entropy,  $n$  is the total number of test data,  $y$  is the desired value, and  $a$  is the predicted value. The reason we used cross entropy instead of mean square error (MSE) is that MSE is a more time-consuming confusion function than cross entropy [51].

For inference evaluation, we relied on the same principle, i.e., the confusion matrix, but by random sampling. As the three images for inference are composed of about  $20,000 \times 30,000$  pixels, if the confusion matrix calculation is conducted on a pixel-by-pixel basis between ground truth data and prediction results, the process is a time-consuming task. As a shorter run time was required to efficiently compare between various band combinations and operational flood monitoring, we adopted a random sampling method for the calculation. Sample size was calculated on each inference datum, satisfying the conditions of a 99% confidence level, the observed percentages of 0.5 and 0.01 margins of error. From the population size, we randomly selected pixels with python code developed for this task and only evaluated selected pixels. To avoid the effect of margin areas, the number of total calculated pixels is far bigger than sample size. We checked the values of the entire calculated confusion matrix, and the randomly sampled calculated confusion matrix was statistically the same. In addition to the quantitative accuracy assessment, intensive visual inspection was performed to evaluate the quality of output images at various scales [28].

## 4. Results

### 4.1. Segmentation Results and Improved Cases

All of the 128 cases that had been determined by the combination of stacked layers were evaluated through the training of models and the inference of the three Sentinel-1 images for image segmentation; therefore, the number of total Sentinel-1 images segmented for this research is 384. Of the 128 cases being tested, 31 cases showed improvement in Overall Accuracy (OA), and 19 cases showed improvement in both averaged IOU and F1 score for the three images segmented for water body extraction, as shown in Figure 6. Most of the cases that consist of six, seven, and eight layers showed lower OA, IOU, and F1 scores, compared to those of the Sentinel-1 VV band image. In Figure 6, the numbers under  $x$ -axes indicate band combinations consisting of 1-VV, 2-DEM, 3-Slope, 4-Aspect, 5-PC, 6-TWI, 7-Buffer, and 8-TRI, and detailed information on stacked geospatial layers are in Table 1 above.

The training accuracy for model performance and the averaged inference results of the three Sentinel-1 images for the 19 cases are shown in Table 5. For training, the Overall Accuracy (OA), IOU, and F1 score of the Sentinel-1 VV band image are 94.91, 87.83, and 93.52, respectively, and those for inference are 95.77, 80.35, and 88.85, which were evaluated based on comparing sampled pixels of the output to corresponding ground truth data. Of the 19 cases, 'band combination 1358' (VV, Slope, PC, and TRI), and 'band combination 1357' (VV, Slope, PC, and Buffer) showed the best inference accuracy. The OA, IOU, and F1 score of 'band combination 1358' are 96.73, 85.42, and 92.08, and those of 'band combination 1357' are 96.89, 85.85, and 92.31, respectively. Compared to the Sentinel-1 VV band, 'band combination 1358' (VV, Slope, PC, and TRI) showed improvement in segmentation accuracy by 0.96, 5.07, and 3.23 in the three criteria.

### 4.2. Improvement in Inference Accuracy of the Three Cases

The F1 score of the 19 band combinations of the scenes A (Padma river basin in Bangladesh), B (Lower Mekong river basin in Vietnam) and C (Chindwin river basin in Myanmar), and the differences in F1 score between the 19 band combinations of those scenes and the Sentinel-1 VV band images are presented in Table 6, in addition to averaged OA, precision, recall, IOU, and F1 score for the 19 cases. The results show the improvement of stacked images combining geospatial layers in segmentation accuracy compared to that of the VV band images. For Scene A, minor improvements (up to 4.25) were observed, whereas a minor decrease ( $-2.46$ ) in F1 score was observed in the difference between B-VV. 'Band combination 1358' (VV, Slope, PC, and TRI) for Scene C showed improvement by 7.68

compared to the segmentation output of the Sentinel-1 VV band image for the same area. Of the 19 cases, 4 cases (band combination 134 (VV, Slope, and Aspect), 1357 (VV, Slope, PC, and Buffer), 1358 (VV, Slope, PC, and TRI), and 1578 (VV, PC, Buffer, and TRI)) showed improvements in F1 score of all of the three scenes compared to that of the Sentinel-1 VV band images. The segmentation results of 'band combination 1358 and 1357' for the three areas, which have different topographical features, are presented in Figure 7 for comparison. Scene C (Chindwin river basin in Myanmar), which contains mountainous areas, showed lower segmentation accuracy compared to that of Scenes A and B, but they also showed the highest improvement in segmentation accuracy.



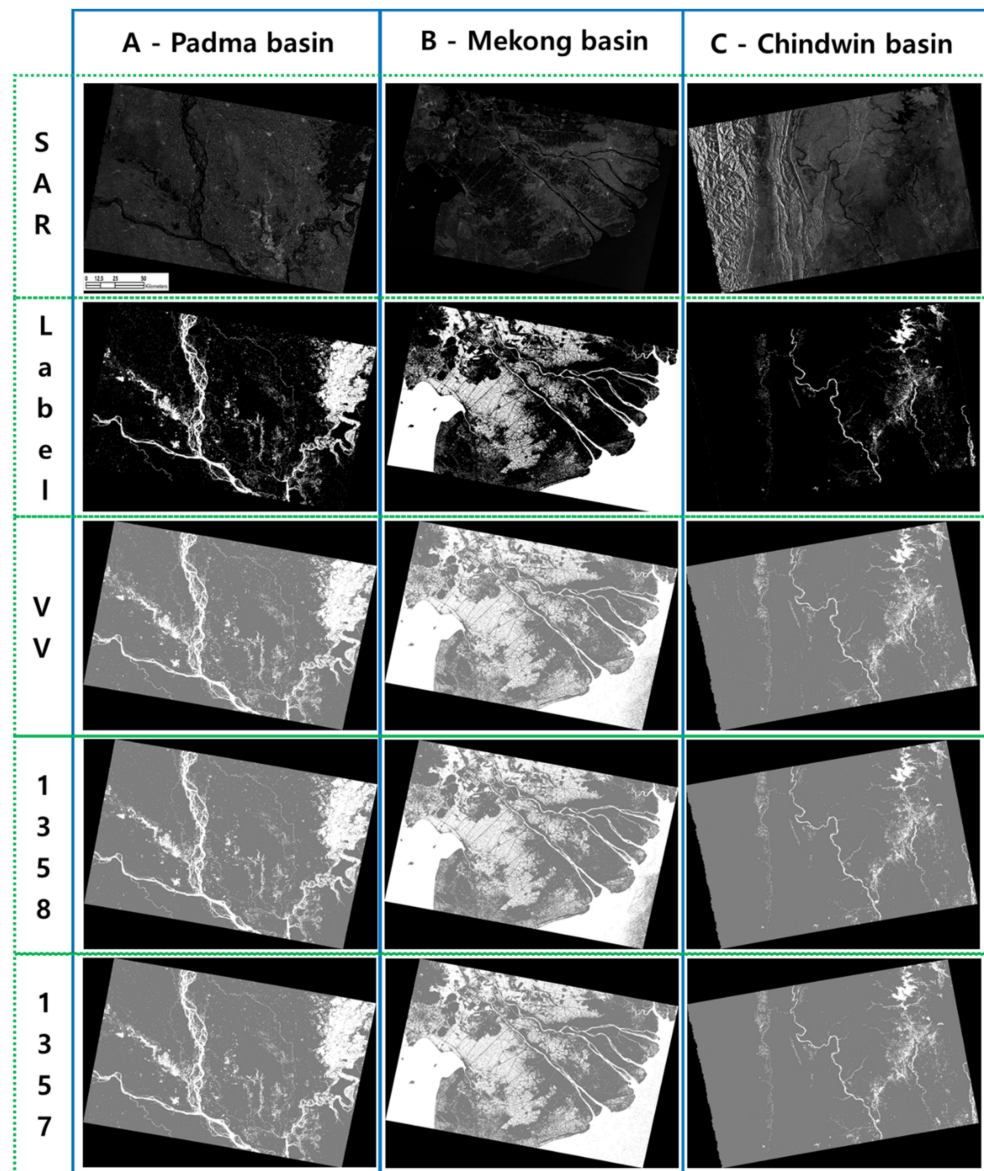
Figure 6. Averaged Overall Accuracy (OA), IOU, and F1 score of all of the 128 cases being tested.

**Table 5.** Selected training and inference results of water body extraction models. (Numbers in the band combination column indicate 1-VV/2-DEM/3-Slope/4-Aspect/5-PC/6-TWI/7-Buffer/8-TRI).

Band Combination	Training				Inference (Averaged)		
	Loss	Accuracy	IOU	F1 Score	Accuracy	IOU	F1 Score
1 (VV)	0.1398	94.91	87.83	93.52	95.77	80.35	88.85
134	0.1727	92.90	82.40	90.35	96.84	83.65	90.95
135	0.1280	95.06	88.02	93.63	95.81	80.41	88.89
148	0.1553	93.81	84.99	91.88	96.25	82.17	90.06
178	0.1414	94.55	87.21	93.17	96.08	80.89	89.19
1257	0.1653	93.32	83.29	90.89	96.87	81.58	89.49
1278	0.1659	92.87	82.02	90.12	96.75	82.09	89.95
1348	0.2095	91.50	79.09	88.32	96.35	81.71	89.81
<b>1357</b>	<b>0.1458</b>	<b>94.49</b>	<b>86.76</b>	<b>92.91</b>	<b>96.89</b>	<b>85.85</b>	<b>92.31</b>
<b>1358</b>	<b>0.1682</b>	<b>93.18</b>	<b>82.82</b>	<b>90.60</b>	<b>96.73</b>	<b>85.42</b>	<b>92.08</b>
1458	0.1596	93.78	84.46	91.58	96.35	80.96	89.19
1567	0.2331	90.94	77.79	87.51	96.83	81.68	89.74
1578	0.1216	95.06	87.79	93.50	96.23	82.58	90.28
12358	0.1446	94.14	85.69	92.30	96.64	82.02	89.87
12378	0.1489	94.26	86.17	92.57	97.12	83.65	90.86
12678	0.2186	91.33	78.17	87.75	96.32	81.29	89.57
13457	0.2086	91.12	77.64	87.41	96.88	82.21	90.04
13458	0.1463	94.06	85.86	92.39	96.68	82.69	90.32
14568	0.1701	93.08	82.96	90.69	96.33	80.49	88.96

**Table 6.** Selected training and inference results of water body extraction models by scene. (Numbers in the band combination column indicate 1-VV/2-DEM/3-Slope/4-Aspect/5-PC/6-TWI/7-Buffer/8-TRI).

Band Combination	Inference (Averaged)					Scenes			Differences		
	Accuracy	Precision	Recall	IOU	F1 Score	A	B	C	A–VV	B–VV	C–VV
1 (VV)	95.77	81.79	98.07	80.35	88.85	90.69	94.40	81.44	0.00	0.00	0.00
<b>134</b>	<b>96.84</b>	<b>89.05</b>	<b>93.05</b>	<b>83.65</b>	<b>90.95</b>	<b>92.89</b>	<b>94.70</b>	<b>85.25</b>	<b>2.20</b>	<b>0.30</b>	<b>3.80</b>
135	95.81	82.25	97.52	80.41	88.89	90.66	94.37	81.64	−0.03	−0.03	0.19
148	96.25	85.13	95.93	82.17	90.06	91.69	94.13	84.36	1.00	−0.27	2.91
178	96.08	82.79	97.39	80.89	89.19	91.61	94.15	81.80	0.92	−0.25	0.36
1257	96.87	88.09	91.11	81.58	89.49	93.50	94.67	80.29	2.81	0.27	−1.15
1278	96.75	92.35	87.72	82.09	89.95	92.51	94.33	82.99	1.82	−0.07	1.55
1348	96.35	88.77	90.98	81.71	89.81	92.90	91.94	84.60	2.20	−2.46	3.16
<b>1357</b>	<b>96.89</b>	<b>88.43</b>	<b>96.80</b>	<b>85.85</b>	<b>92.31</b>	<b>92.79</b>	<b>95.53</b>	<b>88.61</b>	<b>2.10</b>	<b>1.13</b>	<b>7.17</b>
<b>1358</b>	<b>96.73</b>	<b>90.14</b>	<b>94.27</b>	<b>85.42</b>	<b>92.08</b>	<b>91.97</b>	<b>95.15</b>	<b>89.12</b>	<b>1.27</b>	<b>0.75</b>	<b>7.68</b>
1458	96.35	84.72	94.58	80.96	89.19	94.03	92.48	81.05	3.33	−1.92	−0.39
1567	96.83	91.85	87.84	81.68	89.74	92.26	93.59	83.35	1.57	−0.81	1.91
<b>1578</b>	<b>96.23</b>	<b>84.79</b>	<b>97.11</b>	<b>82.58</b>	<b>90.28</b>	<b>91.32</b>	<b>95.11</b>	<b>84.40</b>	<b>0.63</b>	<b>0.72</b>	<b>2.96</b>
12358	96.64	87.36	93.29	82.02	89.87	94.64	92.55	82.41	3.95	−1.85	0.97
12378	97.12	87.71	94.66	83.65	90.86	94.95	93.99	83.65	4.25	−0.41	2.21
12678	96.32	88.31	90.89	81.29	89.57	92.06	92.00	84.66	1.36	−2.39	3.22
13457	96.88	90.67	89.52	82.21	90.04	93.30	93.45	83.36	2.60	−0.94	1.92
13458	96.68	86.11	95.22	82.69	90.32	94.39	93.09	83.47	3.70	−1.31	2.02
14568	96.33	84.88	93.67	80.49	88.96	92.62	92.62	81.63	1.92	−1.77	0.19



**Figure 7.** Segmented images of the three cases described in Figure 1c (A-Padma river basin in Bangladesh (first column), B-Lower Mekong river basin in Vietnam (second column), and C-Chindwin river basin in Myanmar (third column); the first-second row show the Sentinel-1 image and label data for the sites, and the third-last row show the classification results of VV, ‘band combination 1358’ (VV, Slope, PC, and TRI), and ‘band combination 1357’ (VV, Slope, PC, and Buffer) images of the corresponding site).

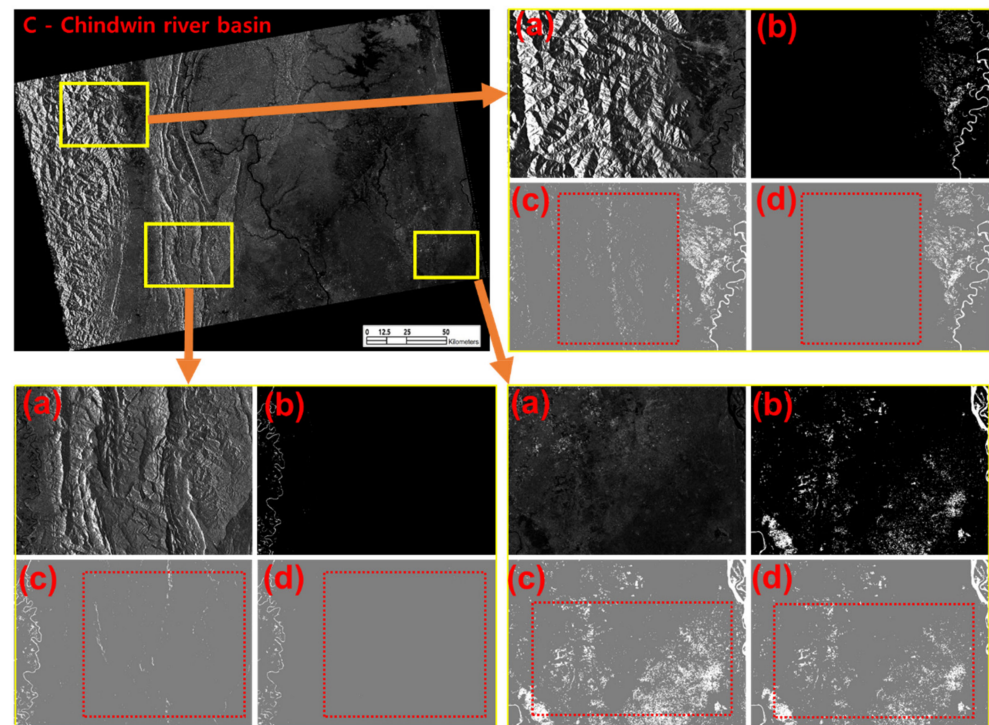
## 5. Discussion

### 5.1. Visual Interpretation

To evaluate semantic image segmentation accuracy at more detailed levels, visual interpretation was conducted for the three segmented Sentinel-1 images. Some examples of the evaluation of water body extraction results for the C-Chindwin river basin (‘band combination 1358’—VV, Slope, PC, and TRI) are presented in Figure 8. Enlarged images (a)–(d) in Figure 8 show: (a) Sentinel-1 images, (b) Label data, (c) Segmentation result of VV band, and (d) Segmentation result of ‘band combination 1358’. As shown in the dotted boxes in red in the output images, and compared to the segmentation result of VV bands, the segmentation result of ‘band combination 1358’ showed a significant reduction in mountain shadows, which are one of main sources of misclassification. As segmentation



accuracy was improved and the terrain effects in the Sentinel-1 images and output images were reduced, it can be said that the results of the qualitative accuracy assessment through visual inspection is consistent with that of the quantitative accuracy assessment. Such improvement was observed in other output images that were produced based on different band combinations and showed improvement in segmentation accuracy that was evaluated through quantitative assessment.



**Figure 8.** Segmentation results of C-Chindwin river basin (see the dotted boxes in red in the output images; (a) Enlarged Sentinel-1 images, (b) Label data, (c) Segmentation result of the Sentinel-1 VV band, (d) Segmentation result of 'band combination 1358' (VV, Slope, PC, and TRI)).

### 5.2. Training and Inference Time for Water Body Extraction

For operational flood monitoring through deep learning-based water body extraction, the training and inference times for the selected 19 cases are presented in Table 7. The training and inference times of the VV band were 1404.95 and 302.20 s, whereas those of the 'band combination 1358' (VV, Slope, PC, and TRI) were 590.25 and 847.41 s, respectively. The averaged training time by the number of bands was: 3 bands—1008.44, 4 bands—1020.91, and 5 bands—1150.58 s. Whereas the averaged inference time by the number of bands was: 3 bands—627.38, 4 bands—856.19, and 5 bands—1009.91 s. Training time was rather decreased as the number of bands was increasing, while inference time was gradually increased as the number of bands was increasing. It was assumed that the decreasing training time is because of the 'early stopping function' that was adopted for training and that the increasing inference time is because of the size of input data for inference that is in proportion to the number of bands. It was shown that although the inference time for water body extraction was increased by adding geospatial layers, it is still acceptable for operational flood monitoring, even when geospatial layers are added to the Sentinel-1 VV band as ancillary data.

**Table 7.** Selected training and inference time of water body extraction models. (Numbers in the band combination column indicate 1-VV/2-DEM/3-Slope/4-Aspect/5-PC/6-TWI/7-Buffer/8-TRI).

No. of Band(s)	Band Combination	Train Time (s)	Inference Time (s)
1	<b>1 (VV)</b>	<b>1404.95</b>	<b>302.20</b>
3	134	1425.92	659.20
	135	767.90	751.94
	148	1323.38	602.16
	178	516.57	496.22
	<b>average</b>	<b>1008.44</b>	<b>627.38</b>
4	1257	833.97	1029.08
	1278	1020.84	802.68
	1348	1146.71	701.48
	1357	621.58	738.80
	1358	590.25	847.41
	1458	1317.86	726.23
	1567	1064.65	872.16
	1578	1571.38	1131.64
	<b>average</b>	<b>1020.91</b>	<b>856.19</b>
5	12358	874.71	914.28
	12378	1182.14	995.53
	12678	655.49	866.57
	13457	1503.43	865.67
	13458	1552.98	1060.59
	14568	1134.72	1356.82
	<b>average</b>	<b>1150.58</b>	<b>1009.91</b>

### 5.3. Summary and General Discussion

Considering the existing previous research, it is clear that the disciplines most directly concerned with flood monitoring using satellite data, disaster management, or remote sensing science have not fully examined how flooded areas can be extracted with the state-of-the-art technique for image classification, i.e., deep learning. To see if our assumption that deep learning-based water body extraction could be improved by using geospatial layers as additional input layers is valid, we advanced an existing deep learning model by customising and optimising its network and processing procedures for more accurate and faster image segmentation.

Through the experiment, a novel water body extraction model based on a deep neural network that exploits Sentinel-1 data and flood-related geospatial datasets was presented for flood monitoring across the Southeast Asia region. For the model, the U-Net was customised and optimised to utilise Sentinel-1 data and other flood-related geospatial data, including digital elevation model (DEM), Slope, Aspect, Profile Curvature (PC), Topographic Wetness Index (TWI), Terrain Ruggedness Index (TRI), and Buffer, in GeoTiff format for the Southeast Asia region. The main features of our deep neural network for water body extraction from Sentinel-1 images are: (1) our model can take geo-located and multi-layered SAR images and other ancillary data in GeoTiff format as input data, (2) the optimised deep network presented for this research does not lose any spatial resolution and location information of the multi-modal input data, and (3) the size of input data for inference is significantly larger than the U-Net to reduce processing time and to achieve the seamless merging of segmented image patches for inference.

To test and validate the water body extraction model, it was applied to three areas in Vietnam, Myanmar, and Bangladesh, and model performance and segmentation accuracy for all of the 128 cases that had been determined by the combination of stacked layers were evaluated in accordance with the types of combined input layers. Therefore, the number of total Sentinel-1 images segmented for this research is 384. Of the 128 cases tested in this research, 31 cases showed improvement in Overall Accuracy (OA), and 19 cases showed improvement in both averaged IOU and F1 score for the three images classified for water body extraction. Most cases that consist of six, seven, and eight layers showed lower OA, IOU, and F1 score compared to those of the Sentinel-1 VV band image. The averaged OA, IOU, and F1 score of the Sentinel-1 VV band are 95.77, 80.35, and 88.85 respectively, whereas those of 'band combination 1358 (VV, Slope, PC, and TRI)' are 96.73, 85.42, and 92.08, showing improvements in all of the criteria for accuracy assessment. The degrees of improvement of the three criteria are 0.96, 5.07, and 3.23, respectively. The improved segmentation accuracy of 'band combination VV, Slope, PC, and TRI' showed a higher OA and F1 score compared to other Sentinel-1-based flood monitoring models [33] or deep learning-based flood monitoring models [17,30,49]. In addition, the averaged processing time, i.e., training and inference time for a Sentinel-1 image, of the 'band combination VV, Slope, PC, and TRI' is greatly shorter than that of [17,33].

Such improvement was clearer in the water body extraction results for the C-Chindwin river basin, which contains mountainous areas. For the image, quantitative evaluation of 'band combination 1358' (VV, Slope, PC, and TRI) showed an improvement in F1 score by 7.68 percent compared to the segmentation output of the Sentinel-1 VV band, and it was also demonstrated through visual interpretation. As segmentation accuracy was improved and the terrain effects in the Sentinel-1 images and output images were reduced, the results of the qualitative accuracy assessment through visual inspection is consistent with that of the quantitative accuracy assessment. To the best of our knowledge, this is the first study that demonstrates the synergistic use of geospatial data in deep learning-based water body extraction over wide areas.

#### *5.4. Novelty, Limitations, and Future Work*

The main purpose of this research is to present a reliable deep learning-based flood monitoring model that has better predictive ability by testing the effectiveness of geospatial data. For this research, the U-Net architecture was customised and optimised to utilise Sentinel-1 data and seven different types of geospatial layers. Through the research, it was demonstrated that the accuracy of deep learning-based water body extraction can be improved by using geospatial data, and based on the experiment, a new water body extraction model is presented for flood monitoring across the Southeast Asia region. While previous studies focused on producing more training data or advancing network architectures to improve image classification accuracy, we focused rather on utilising available flood data and flood-related geospatial data and demonstrated our assumption that deep learning-based water body extraction can be improved by using geospatial layers as additional input layers.

Although it was demonstrated that deep learning-based water body extraction can be improved by exploiting geospatial layers, it does not mean that classification performance is always improved by using geospatial layers, and the result of this research is applicable to other existing deep neural networks without testing its applicability and transferability. As per the research aim of this study, this research is confined to evaluating satellite data and available geospatial layers. To derive more reliable water body extraction models for flood monitoring, more geospatial layers and non-geospatial data need to be tested, and the possibility of reducing misclassification of other factors, such as roads and airports, needs to be verified to achieve better classification accuracy.

## 6. Conclusions

Floods occur unexpectedly and cause devastating damage over broad areas. Yet the timely monitoring and estimation of inundated areas using satellite data has been limited by satellite data acquisition and classification accuracy. Although deep learning is a promising method for satellite image classification, the synergistic use of geospatial data for water body extraction from Sentinel-1 data using deep learning and the applicability of existing deep learning models have not been thoroughly tested for operational flood monitoring. To fill the knowledge gap, a novel water body extraction model was presented based on a deep neural network that exploits Sentinel-1 data and flood-related geospatial datasets, including digital elevation model (DEM), Slope, Aspect, Profile Curvature (PC), Topographic Wetness Index (TWI), Terrain Ruggedness Index (TRI), and Buffer for the Southeast Asia region. For the model, the U-Net was customised and optimised to utilise Sentinel-1 data and other flood-related geospatial data in GeoTiff format for operational flood monitoring in the Southeast Asia region. The testing and validation of the water body extraction model was applied to three Sentinel-1 images for Vietnam, Myanmar, and Bangladesh. Model performance and segmentation accuracy for all of the 128 cases that the combination of stacked layers had determined were evaluated following the types of combined input layers.

Through this research, it was demonstrated that the accuracy of deep learning-based water body extraction can be improved up to 7.68 percent by using geospatial data, and based on the experiment, a new water body extraction model that is further verified through visual inspection and the evaluation of model performance, including training and inference time, is presented for operational flood monitoring across the Southeast Asia region. As per the research aim of this study, this research is confined to evaluating satellite data and available geospatial layers. To derive more reliable water body extraction models for operational flood monitoring, more geospatial layers and non-geospatial data need to be tested, and the possibility of reducing misclassification of other factors, such as roads and airports, needs to be verified to achieve better classification accuracy.

**Author Contributions:** Conceptualization, J.K.; methodology, J.K. and D.-j.K.; software, J.K., H.J. and H.K.; validation, J.K. and H.K.; formal analysis, J.K., H.K., H.J. and S.-H.J.; investigation, J.K. and H.K.; resources, J.K. and D.-j.K.; data curation, J.K.; writing—original draft preparation, J.K.; writing—review and editing, J.K., H.K., J.S., S.K.P.V. and D.-j.K.; visualization, J.K. and H.K.; supervision, D.-j.K.; project administration, D.-j.K.; funding acquisition, D.-j.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by a grant (20009742) from the Ministry-Cooperation R&D program of Disaster-Safety, funded by the Ministry of Interior and Safety (MOIS, Korea).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Acknowledgments:** We thank the United Nations Institute for Training and Research's (UNITAR) Operational Satellite Applications Programme (UNOSAT) for providing UNOSAT flood data. The Sentinel-1 data used in this study are provided by ESA through the Sentinel-1 Scientific Data Hub.

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

## References

1. Schumann, G.J.-P.; Moller, D.K. Microwave Remote Sensing of Flood Inundation. *Phys. Chem. Earth* **2015**, *83*, 84–95. [[CrossRef](#)]
2. Tien Bui, D.; Hoang, N.D.; Martínez-Álvarez, F.; Ngo, P.T.T.; Hoa, P.V.; Pham, T.D.; Samui, P.; Costache, R. A Novel Deep Learning Neural Network Approach for Predicting Flash Flood Susceptibility: A Case Study at a High Frequency Tropical Storm Area. *Sci. Total Environ.* **2020**, *701*, 134413. [[CrossRef](#)] [[PubMed](#)]
3. Yilmaz, K.K.; Adler, R.F.; Tian, Y.; Hong, Y.; Pierce, H.F. Evaluation of a Satellite-Based Global Flood Monitoring System. *Int. J. Remote Sens.* **2010**, *31*, 3763–3782. [[CrossRef](#)]
4. Guha-Sapir, D.; Vos, F.; Below, R.; Ponslerre, S. *Annual Disaster Statistical Review 2011: The Numbers and Trends*; Centre for Research on the Epidemiology of Disasters (CRED): Brussels, Belgium, 2012.



5. Sheng, Y.; Gong, P.; Xiao, Q. Quantitative Dynamic Flood Monitoring with NOAA AVHRR. *Int. J. Remote Sens.* **2001**, *22*, 1709–1724. [[CrossRef](#)]
6. Voigt, S.; Giulio-Tonolo, F.; Lyons, J.; Kuřcera, J.; Jones, B.; Schneiderhan, T.; Platzeck, G.; Kaku, K.; Hazarika, M.K.; Czarán, L.; et al. Global Trends in Satellite-Based Emergency Mapping. *Science* **2016**, *353*, 247–252. [[CrossRef](#)] [[PubMed](#)]
7. Brivio, P.A.; Colombo, R.; Maggi, M.; Tomasoni, R. Integration of Remote Sensing Data and GIS for Accurate Mapping of Flooded Areas. *Int. J. Remote Sens.* **2002**, *23*, 429–441. [[CrossRef](#)]
8. Changnon, S.A. Assessment of Flood Losses in the United States. *J. Contemp. Water Res. Educ.* **2008**, *138*, 38–44. [[CrossRef](#)]
9. Shen, X.; Wang, D.; Mao, K.; Anagnostou, E.; Hong, Y. Inundation Extent Mapping by Synthetic Aperture Radar: A Review. *Remote Sens.* **2019**, *11*, 879. [[CrossRef](#)]
10. Hess, L.L.; Melack, J.M.; Filoso, S.; Wang, Y. Delineation of Inundated Area and Vegetation along the Amazon Floodplain with the SIR-C Synthetic Aperture Radar. *IEEE Trans. Geosci. Remote Sens.* **1995**, *33*, 896–904. [[CrossRef](#)]
11. Hahmann, T.; Martinis, S.; Twele, A.; Roth, A.; Buchroithner, M. Extraction of Water and Flood Areas from SAR data. In Proceedings of the 7th European Conference on Synthetic Aperture Radar, Friedrichshafen, Germany, 2–5 June 2008; pp. 1–4.
12. Manavalan, R. SAR Image Analysis Techniques for Flood Area Mapping—Literature Survey. *Earth Sci. Inform.* **2017**, *10*, 1–14. [[CrossRef](#)]
13. Tsyganskaya, V.; Martinis, S.; Marzahn, P.; Ludwig, R. SAR-based Detection of Flooded Vegetation—A Review of Characteristics and Approaches. *Int. J. Remote Sens.* **2018**, *39*, 2255–2293. [[CrossRef](#)]
14. Greifeneder, F.; Wagner, W.; Sabel, D.; Naeimi, V. Suitability of SAR Imagery for Automatic Flood Mapping in the Lower Mekong Basin. *Int. J. Remote Sens.* **2014**, *35*, 2857–2874. [[CrossRef](#)]
15. Pulvirenti, L.; Pierdicca, N.; Chini, M.; Guerriero, L. An Algorithm for Operational Flood Mapping from Synthetic Aperture Radar (SAR) Data Based on the Fuzzy Logic. *Nat. Hazards Earth Syst. Sci.* **2011**, *11*, 529–540. [[CrossRef](#)]
16. Martinis, S.; Kuenzer, C.; Wendleder, A.; Huth, J.; Twele, A.; Roth, A.; Dech, S. Comparing Four Operational SAR-based Water and Flood Detection Approaches. *Int. J. Remote Sens.* **2015**, *36*, 3519–3543. [[CrossRef](#)]
17. Kang, W.; Xiang, Y.; Wang, F.; Wan, L.; You, H. Flood Detection in Gaofen-3 SAR Images via Fully Convolutional Networks. *Sensors* **2018**, *18*, 2915. [[CrossRef](#)]
18. Zhang, P.; Chen, L.; Li, Z.; Xing, J.; Xing, X.; Yuan, Z. Automatic Extraction of Water and Shadow from SAR Images Based on a Multi-resolution Dense Encoder and Decoder Network. *Sensors* **2019**, *19*, 3576. [[CrossRef](#)]
19. Refice, A.; Capolongo, D.; Pasquariello, G.; D’Addabbo, A.; Bovenga, F.; Nutricato, R.; Lovergine, F.P.; Pietranera, L. SAR and InSAR for Flood Monitoring: Examples with COSMO-SkyMed Data. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2014**, *7*, 2711–2722. [[CrossRef](#)]
20. D’Addabbo, A.; Refice, A.; Pasquariello, G.; Lovergine, F.P.; Capolongo, D.; Manfreda, S. A Bayesian Network for Flood Detection Combining SAR Imagery and Ancillary Data. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3612–3625. [[CrossRef](#)]
21. Feng, M.; Sexton, J.O.; Channan, S.; Townshend, J.R. A Global, High-Resolution (30-m) Inland Water Body Dataset for 2000: First Results of a Topographic–Spectral Classification Algorithm. *Int. J. Digit. Earth* **2016**, *9*, 113–133. [[CrossRef](#)]
22. Pekel, J.F.; Cottam, A.; Gorelick, N.; Belward, A.S. High-Resolution Mapping of Global Surface Water and its Long-Term Changes. *Nature* **2016**, *540*, 418–422. [[CrossRef](#)] [[PubMed](#)]
23. Bates, B.C.; Kundzewicz, Z.W.; Wu, S.; Palutikof, J.P. *Climate Change and Water: Technical Paper of the Intergovernmental Panel on Climate Change*; IPCC Secretariat: Geneva, Switzerland, 2009; p. 210.
24. Kundzewicz, Z.W.; Kanae, S.; Seneviratne, S.I.; Handmer, J.; Nicholls, N.; Peduzzi, P.; Mechler, R.; Bouwer, L.M.; Arnell, N.; Mach, K.; et al. Flood Risk and Climate Change: Global and Regional Perspectives. *Hydrol. Sci. J.* **2013**, *59*, 1–28. [[CrossRef](#)]
25. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Garcia-Rodriguez, J. A Review on Deep Learning Techniques Applied to Semantic Segmentation. *arXiv* **2017**, arXiv:1704.06857.
26. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [[CrossRef](#)]
27. Fawaz, H.I.; Lucas, B.; Forestier, G.; Pelletier, C.; Schmidt, D.F.; Weber, J.; Webb, G.I.; Idoumghar, L.; Muller, P.A.; Petitjean, F. InceptionTime: Finding AlexNet for Time Series Classification. *arXiv* **2019**, arXiv:1909.04939.
28. Pashaei, M.; Kamangir, H.; Starek, M.J.; Tissot, P. Review and Evaluation of Deep Learning Architectures for Efficient Land Cover Mapping with UAS Hyper-spatial Imagery: A Case Study over a Wetland. *Remote Sens.* **2020**, *12*, 959. [[CrossRef](#)]
29. Ngo, P.T.T.; Hoang, N.D.; Pradhan, B.; Nguyen, Q.K.; Tran, X.T.; Nguyen, Q.M.; Nguyen, V.N.; Samui, P.; Tien Bui, D. A Novel Hybrid Swarm Optimized Multilayer Neural Network for Spatial Prediction of Flash Floods in Tropical Areas Using Sentinel-1 SAR Imagery and Geospatial Data. *Sensors* **2018**, *18*, 3704. [[CrossRef](#)] [[PubMed](#)]
30. Wieland, M.; Martinis, S. A Modular Processing Chain for Automated Flood Monitoring from Multi-Spectral Satellite Data. *Remote Sens.* **2019**, *11*, 2330. [[CrossRef](#)]
31. UNOSAT. UNOSAT Flood Dataset. 2019. Available online: <http://floods.unosat.org/geoportal/catalog/main/home.page> (accessed on 7 September 2020).
32. Nemni, E.; Bullock, J.; Belabbes, S.; Bromley, L. Fully Convolutional Neural Network for Rapid Flood Segmentation in Synthetic Aperture Radar Imagery. *Remote Sens.* **2020**, *12*, 2532. [[CrossRef](#)]
33. Twele, A.; Cao, W.; Plank, S.; Martinis, S. Sentinel-1-based Flood Mapping: A Fully Automated Processing Chain. *Int. J. Remote Sens.* **2016**, *37*, 2990–3004. [[CrossRef](#)]

34. Tzavella, K.; Fekete, A.; Fiedrich, F. Opportunities Provided by Geographic Information Systems and Volunteered Geographic Information for a Timely Emergency Response during Flood Events in Cologne, Germany. *Nat. Hazards*. **2018**, *91*, 29–57. [[CrossRef](#)]
35. Kia, M.B.; Pirasteh, S.; Pradhan, B.; Mahmud, A.R.; Sulaiman, W.N.A.; Moradi, A. An Artificial Neural Network Model for Flood Simulation Using GIS: Johor River Basin, Malaysia. *Environ. Earth Sci.* **2012**, *67*, 251–264. [[CrossRef](#)]
36. Stefanidis, S.; Stathis, D. Assessment of Flood Hazard Based on Natural and Anthropogenic Factors Using Analytic Hierarchy Process (AHP). *Nat. Hazards* **2013**, *68*, 569–585. [[CrossRef](#)]
37. Tehrany, M.S.; Pradhan, B.; Jebur, M.N. Flood Susceptibility Mapping Using a Novel Ensemble Weights-of-evidence and Support Vector Machine Models in GIS. *J. Hydrol.* **2014**, *512*, 332–343. [[CrossRef](#)]
38. Al-Juaidi, A.E.; Nassar, A.M.; Al-Juaidi, O.E. Evaluation of Flood Susceptibility Mapping Using Logistic Regression and GIS Conditioning Factors. *Arab. J. Geosci.* **2018**, *11*, 765. [[CrossRef](#)]
39. Pradhan, B. Flood Susceptible Mapping and Risk Area Delineation Using Logistic Regression, GIS and Remote Sensing. *J. Spat. Hydrol.* **2010**, *9*, 1–18.
40. Riley, S.J.; DeGloria, S.D.; Elliot, R. Index that Quantifies Topographic Heterogeneity. *Intermt. J. Sci.* **1999**, *5*, 23–27.
41. Tarboton, D.G. A New Method for the Determination of Flow Directions and Upslope Areas in Grid Digital Elevation Models. *Water Resour. Res.* **1997**, *33*, 309–319. [[CrossRef](#)]
42. Beven, K.J.; Kirkby, M.J. A Physically Based, Variable Contributing Area Model of Basin Hydrology. *Hydrol. Sci. J.* **1979**, *24*, 43–69.
43. Sörensen, R.; Zinko, U.; Seibert, J. On the Calculation of the Topographic Wetness Index: Evaluation of Different Methods Based on Field Observations. *Hydrol. Earth Syst. Sci.* **2006**, *10*, 101–112. [[CrossRef](#)]
44. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
45. Wurm, M.; Stark, T.; Zhu, X.X.; Weigand, M.; Taubenböck, H. Semantic Segmentation of Slums in Satellite Images Using Transfer Learning on Fully Convolutional Neural Networks. *ISPRS J. Photogramm. Remote Sens.* **2019**, *150*, 59–69. [[CrossRef](#)]
46. Noh, H.; Hong, S.; Han, B. Learning Deconvolution Network for Semantic Segmentation. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1520–1528.
47. Wang, S.; Chen, W.; Xie, S.M.; Azzari, G.; Lobell, D.B. Weakly Supervised Deep Learning for Segmentation of Remote Sensing Imagery. *Remote Sens.* **2020**, *12*, 207. [[CrossRef](#)]
48. Du, L.; McCarty, G.W.; Zhang, X.; Lang, M.W.; Vanderhoof, M.K.; Li, X.; Huang, C.; Lee, S.; Zou, Z. Mapping Forested Wetland Inundation in the Delmarva Peninsula, USA Using Deep Convolutional Neural Networks. *Remote Sens.* **2020**, *12*, 644. [[CrossRef](#)]
49. Yang, M.D.; Tseng, H.H.; Hsu, Y.C.; Tsai, H.P. Semantic Segmentation Using Deep Learning with Vegetation Indices for Rice Lodging Identification in Multi-date UAV Visible Images. *Remote Sens.* **2020**, *12*, 633. [[CrossRef](#)]
50. Goutte, C.; Gaussier, E. A Probabilistic Interpretation of Precision, Recall and F-score, with Implication for Evaluation. In Proceedings of the European Conference on Information Retrieval, Santiago de Compostela, Spain, 21–23 March 2005; pp. 345–359.
51. Golik, P.; Doetsch, P.; Ney, H.; Cross-Entropy, vs. Squared Error Training: A Theoretical and Experimental Comparison. In Proceedings of the Interspeech, Lyon, France, 25–29 August 2013; pp. 1756–1760.