ORIGINAL ARTICLE

Automatic Flood Detection from Sentinel-1 Data Using a Nested UNet Model and a NASA Benchmark Dataset

Binayak Ghosh[1] · Shagun Garg[1,2] · Mahdi Motagh[1,3] · Sandro Martinis[4]

Abstract

During flood events near real-time, synthetic aperture radar (SAR) satellite imagery has proven to be an efficient management tool for disaster management authorities. However, one of the challenges is accurate classification and segmentation of flooded water. A common method of SAR-based flood mapping is binary segmentation by thresholding, but this method is limited due to the effects of backscatter, geographical area, and surface characterstics. Recent advancements in deep learning algorithms for image segmentation have demonstrated excellent potential for improving flood detection. In this paper, we present a deep learning approach with a nested UNet architecture based on a backbone of EfficientNet-B7 by leveraging a publicly available Sentinel-1 dataset provided jointly by NASA and the IEEE GRSS Committee. The performance of the nested UNet model was compared with several other UNet-based convolutional neural network architectures. The models were trained on flood events from Nebraska and North Alabama in the USA, Bangladesh, and Florence, Italy. Finally, the generalization capacity of the trained nested UNet model was compared to the other architectures by testing on Sentinel-1 data from flood events of varied geographical regions such as Spain, India, and Vietnam. The impact of using different polarization band combinations of input data on the segmentation capabilities of the nested UNet and other models is also evaluated using Shapley scores. The results of these experiments show that the UNet model architectures perform comparably to the UNet++ with EfficientNet-B7 backbone for both the NASA dataset as well as the other test cases. Therefore, it can be inferred that these models can be trained on certain flood events provided in the dataset and used for flood detection in other geographical areas, thus proving the transferability of these models. However, the effect of polarization still varies across different test cases from around the world in terms of performance; the model trained with the combinations of individual bands, VV and VH, and polarization ratios gives the best results.

Keywords Flood detection · NASA · Synthetic aperture radar · Transfer learning · Deep learning · UNet++ · UNet

✉ Binayak Ghosh
binghosh@gmail.com

Shagun Garg
sg2009@cam.ac.uk

Mahdi Motagh
motagh@gfz-potsdam.de

Sandro Martinis
sandro.martinis@dlr.de

1   Department of Geodesy, Section of Remote Sensing and Geoinformatics, GFZ German Research Centre for Geosciences, 14473 Potsdam, Germany

2   Future Infrastructure and Built Environment, Department of Engineering, University of Cambridge, Cambridge, UK

3   Institute of Photogrammetry and Geoinformation (IPI), Leibniz University Hannover, 30167 Hannover, Germany

4   German Aerospace Centre (DLR), Köln, Germany

1 Introduction

Flooding is a widespread and dramatic natural disaster that affects lives, infrastructures, economics, and local ecosystems all over the world. Floods often cause loss of life and substantial property damage. Moreover, the economic ramifications of flood damage disproportionately impact the most vulnerable members of society. Due to their imaging capabilities that allow data acquisition regardless of illumination and weather conditions, satellite synthetic aperture radar (SAR) data have become the most widely used Earth observation (EO) data for operational flood monitoring (Martinis et al. 2015). For flood mapping using SAR amplitude data, the specular reflection occurring on smooth water surfaces results in most cases in a dark tone in SAR data, making floodwater distinguishable from other land

surfaces. In the case of urban areas, flooding results in an increase of the backscatter, which results in brighter images.

Some of the challenges in this regard include accurate classification and segmentation of flooded water as well as availability of annotated datasets for training models and subsequent comparison. One of the prevalent methods is binary segmentation using thresholding methods like Otsu (1975). In the paper Landuyt et al. (2019), the authors describe their comprehensive study of established pixel-based flood mapping approaches like global and enhanced thresholding, active contour modeling, and change detection on SAR images of different flooded areas. However, the generalization ability of these methods is limited due to the effects of backscatter, geographical area, and time of image collection. Another area of approach for flood detection using SAR images includes rule-based classification methods like M et al. (2017). In the paper Pradhan et al. (2016), the authors also include Taguchi optimization techniques along with rule-based classification for mapping flood extent. Few studies have utilized texture information for mapping flood using a single SAR image Dasgupta et al. (2018); Ritushree et al. (2023). Ouled Sghaier et al. (2018) introduces a texture analysis method to map flood extent from a time series of SAR images. With the development of the Google Earth Engine, web-based applications for flood mapping like those described in Tripathy and Malladi (2022) are also being introduced, which involves using a long stretch of SAR images before and after the flood and calculating the mean, standard deviation, and Z-score of each pixel.

Over the past few years, machine learning methods have been in prevalence for flood mapping. Techniques like Bayesian network fusion (Li et al. 2019b), self-organized maps (Skakun 2010), and support vector machines (Insom et al. 2015) have been applied for extraction of flooded areas from optical as well as SAR satellite images, although there have been limited studies in this domain due to the lack of large-scale labeled flood event datasets. Deep learning methods represented by convolutional neural networks have proven to be effective in the field of flood damage assessment (Bai et al. 2021; Ghosh et al. 2020) and have enabled development of new methods for automated extraction of flood extent from SAR images (Zhang et al. 2019). In the paper Muñoz et al. (2021), the authors combined multispectral Landsat imagery and dual-polarized synthetic aperture radar imagery to evaluate the performance of integrating convolutional neural network and a data fusion framework for generating compound flood mapping. These studies show that deep learning algorithms play an important role in enhancing flood classification. The recent release of the large-scale open-source Sen1Floods11 dataset (Bonafilia et al. 2020) has boosted research into using deep learning algorithms for water-type detection in flood disasters, as was shown in Konapala

and Kumar (2021). Similarly, in the Bai et al. (2021), the authors have utilized this dataset for training and validation as well as testing the model on a previously unseen subset from the same dataset. The paper Katiyar et al. (2021) also uses off-the-shelf models like baseline Segnet and UNet for flood mapping and comparison of different results based on different labeling techniques. Fully convolutional networks (FCN) have been quite prevalent for flood mapping, as described in the papers Li et al. (2019a); Nemni et al. (2020); Rudner et al. (2019). All the previous studies have shown that flood detection is possible using rule-based as well as deep learning methods. However, it is also important to compare and quantify how some of these state-of-the-art deep learning models perform on a global scale, and whether these deep learning models can enhance automatically generated flood maps. A higher quality of detection also helps with the timeliness of rapid mapping of floods and in providing critical information for disaster relief management services.

In this work, we aim to design and train a deep learning model on the labeled flood data from some specific geographical regions and then test the performance of the trained models on real Sentinel-1 data from other geographical regions. This is to test whether it is possible to detect floods from SAR data from certain parts of the world, even though the model has been trained on a training dataset from a completely different geographical area. We utilize the publicly available Sentinel-1 dataset provided jointly by the NASA Interagency Implementation and Advanced Concepts Team and the Institute of Electrical and Electronics Engineers Geoscience and Remote Sensing Society (IEEE- GRSS) technical committee. The model trained on this dataset was then applied to flood events from five varying geographical areas to assess the performance of the model. For our analysis, we have implemented the EfficientNet-B7 architecture of the encoder in a nested UNet model to assess its performance on flood segmentation from Sentinel-1 images. The nested UNet has previously been used for medical image segmentation (Zhou et al. (2018)). In our previous paper, Ghosh et al. (2022), we presented an initial evaluation of two deep learning models, a baseline UNet and a feature pyramid network (FPN), on only the Florence subset of the NASA dataset. In this study, for evaluation, the performance of the UNet++ model has been compared to other convolutional neural network models like Resnet34, InceptionV3, and Efficient net-B7, all having the same baseline UNet architecture. This is to evaluate whether the nested UNet with EfficientNet-B7 backbone performs significantly better than other baseline UNet architectures with backbones of Resnet34, InceptionV3, and EfficientNet-B7. We chose UNet because it is a common performance baseline for image segmentation. In the paper Helleis et al. (2022), the authors utilize a similar method

of comparing many deep learning models like ResNet-34 and different variations of the UNet like AlbuNet for flood mapping.

However, one of the main differences between this work and the work described in Helleis et al. (2022) is the choice of the training dataset. In Helleis et al. (2022), the authors train the models on the Sen1Floods11 dataset, which has already been mentioned. Our choice of dataset is the dataset provided jointly by the NASA Interagency Implementation and Advanced Concepts Team and the IEEE- GRSS technical committee. While many of the publications on flood mapping mentioned before used the Sen1Floods11 dataset for training, the dataset from NASA has not been so widely used to date. In that regard, we believe it will be interesting to see how models trained on this dataset perform on flood case studies from different parts of the world.

Another aspect of evaluation in this work is the impact of using different polarization band combinations of input data on the segmentation capabilities of the nested UNet and other models. The nested UNet, along with the other models, is trained on single bands like VV and VH as well as band combinations to assess the feature importance of these bands for flood segmentation, similar to the paper Helleis et al. (2022). The feature importance of the bands is then measured using the Shapley score.

2　Materials and Methods

2.1　Training Dataset

The dataset was derived from the flood detection challenge organized by the NASA Disaster Team in collaboration with the Alaska SAR Facility – Distributed Active Archive Centers (ASF-DAAC). The dataset was labeled by the NASA IMPACT Machine Learning Team (NASA 2021) to generate the flood extent maps. The dataset is quite diverse and more representative of the different variations of geographical areas that were affected by flood, including agricultural land and urban settings. The dataset covered four flood events from Nebraska, North Alabama, Bangladesh, and Florence. A total of 54 Geotiff images were converted into tiles of 256×256 pixels. More comprehensive details about the flood events are detailed in Table 1.

Each tile is generated from Sentinel-1 C-band synthetic aperture radar (SAR) imagery data acquired in interferometric wide swath mode in $5\,\text{m} \times 20\,\text{m}$ resolution, with dual polarization mode i.e., "VV" and "VH". Subsequent additional steps are applied for further processing the images using the Sentinel-1 Toolbox (Sentinel-1 2015), like applying border noise correction, speckle filtering and radiometric terrain normalization using a digital elevation model (DEM) at a resolution of $30\,\text{m}$. The final terrain-corrected values are converted to decibels via log scaling ($10 \times \log10(x)$). The
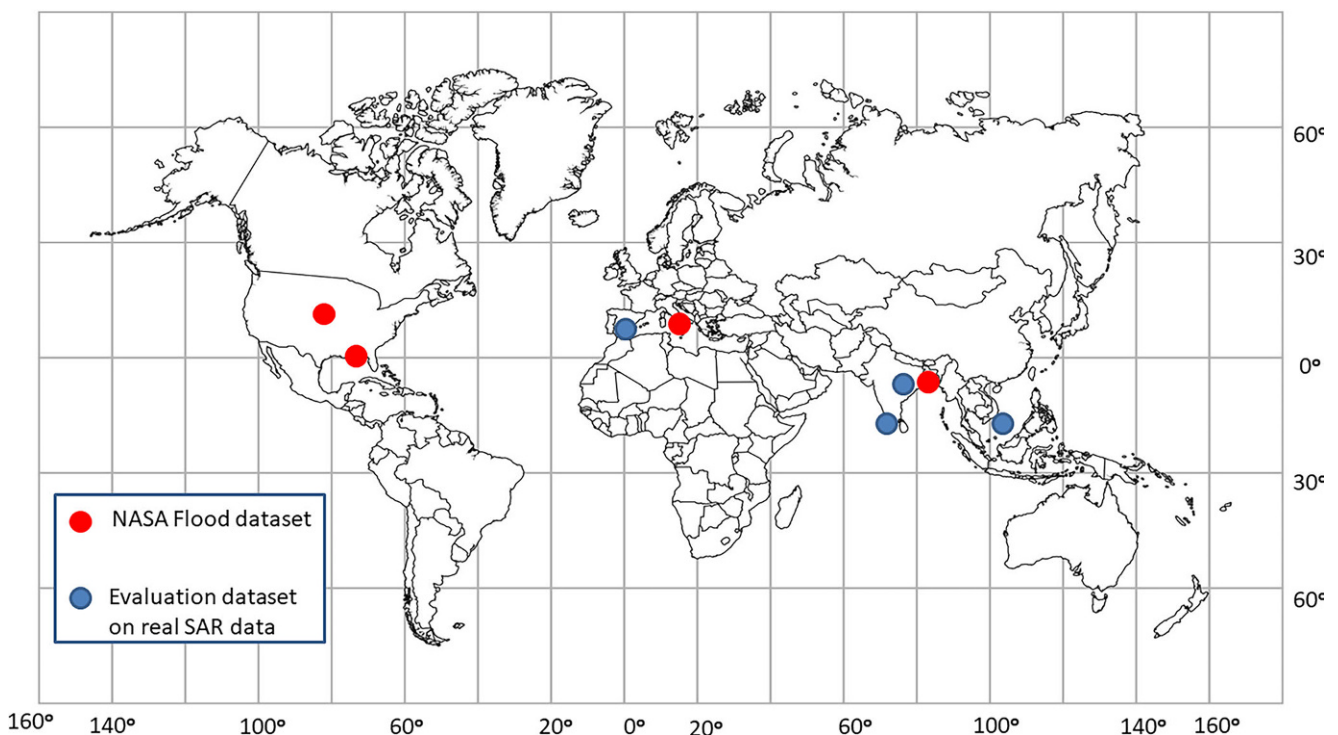

Fig. 1 Overview of the study sites. Colored dots represent the tile locations, red dots represent NASA flood dataset sites, and blue dots represent the sites of the real SAR data used for evaluation

Table 1 Table showing the different flood events covered in the dataset

Area name	Total area covered	Flood start date	Flood end date
Bangladesh	7,150 km^2	14.03.2017	12.07.2017
Nebraska	1,741 km^2	14.03.2019	01.04.2019
North Alabama	13,789 km^2	02.02.2019	27.03.2019
Florence	7,197 km^2	01.10.2018	05.10.2018

imagery is then converted to a 0–255 grayscale image. All these steps were also followed while preparing the evaluation dataset from different geographical areas. The whole NASA dataset consists of approximately 66.000 tiled images from these various geographic locations. In addition to the two polarization bands, VV and VH, an additional channel containing the value of $(VV + VH)/(VV - VH)$ for every pixel was added to all the images. The corresponding flood label was treated as the ground truth.

The Bangladesh geographic area, which is predominant in the dataset, is primarily an agricultural hub, and recently harvested fields can look similar to floods due to low backscatter in both VV and VH polarizations. Similarly, the validation dataset from Florence has a primarily urban setting. Such varying backscatter is relevant for performance optimizations and generalizability to test imagery, as our ultimate aim is to apply the model for flood extent detection on Sentinel-1 images from varying geographic locations.

For our evaluation, we keep the entire Florence dataset as our validation set while training our model. A total of 8382 images from the Florence dataset are separated from the rest of the data, and the remaining approximately 25,000 image tiles are used for training. The dataset is imbalanced, i.e., the proportion of images with some flood region presence is lower than that of the images without it; thus, during training, it is ensured that each batch contains at least 50% samples with some amount of flood region present through stratified sampling. Also, a data augmentation technique is applied whereby each tile is rotated by 90°, 180°, and 270° to reduce the data imbalance problem. This method is shown in Fig. 2.
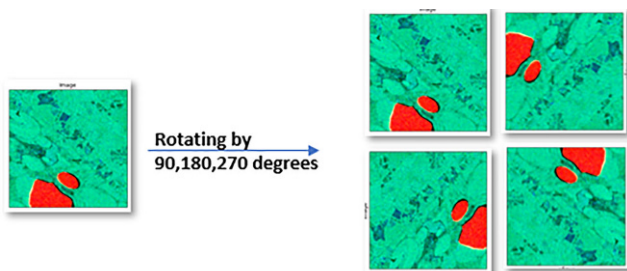

Fig. 2 Each tile is rotated by 90°, 180°, and 270° to reduce the data imbalance problem

2.2 Evaluation Dataset From Different Geographical Areas

The trained models were then tested on Sentinel-1 flood data from five different geographical areas. The flood events covered in these dataset described below:

2.2.1 Spain Floods, 2019

From 11 to 14 September 2019, torrential rainfall (recorded 296.4 mm of rain in 24 hours (spa 2019) caused major flooding in southeastern Spain (Valencia, Alicante, Murcia, Albacete, and Almería). By 12 September, several rivers, including the Segura and Cànyoles rivers, overflowed their banks and caused major flooding and material damage. To analyze this flooding, VV and VH pairs of Sentinel-1 images were acquired from the ESA Sentinel Hub (Hub 2015) for 11 and 16 September, 2019, as before and after flood images.

2.2.2 Kerala Floods, India, 2018

In August 2018, severe floods affected the south Indian state of Kerala due to unusually high rainfall during the monsoon season, resulting in dams filling to their maximum capacities. Almost all dams had to be opened, since the water level had risen close to overflow level, thus flooding local low-lying areas (Lal et al. 2020). To analyze this flooding, VV and VH pairs of Sentinel-1 images were acquired from the ESA Sentinel Hub (Hub 2015) for 4 July and 27 August, 2018, as before and after flood images.

2.2.3 Bihar Floods, India, 2021

In the state of Bihar in the northern Indian plains, due to excessive rainfall between June 1 to July 23, 2021, several rivers submerged many villages in their path (Downtoearth 2021). Other than that, these floods were also attributed to the Farakka Barrage on the river Ganga in West Bengal, Bihar's eastern neighbour. The Farakka Barrage in West Bengal, which regulates the flow of the Ganga, led to sediment deposition upstream of Farakka, which, in turn, led to a rise in the height of the river bed and, hence, the record flood levels in the state. To analyze this flooding, VV and VH pairs of Sentinel-1 images were acquired from the ESA Sentinel Hub (Hub 2015) for 12 June and 6 July, 2021, as before and after flood images.

2.2.4 Vietnam Floods, 2020

In October 2020, the central region of Vietnam experienced a number of inclement tropical storms, including Linfa and Nangka (Reliefweb 2020), which brought heavy rain-

fall (around 2000 mm between 5 to 20 October 2020) and caused water levels in rivers to rise rapidly. Several rivers, especially in the Dong Hoi and Dong Ha regions, reached historically high levels and caused heavy flooding. To analyze these floods, VV and VH pairs of Sentinel-1 images of both Dong Hoi and Dong Ha regions were acquired from the ESA Sentinel Hub (Hub 2015) for 12 June and 6 July, 2021, as before and after flood images.

These five different datasets represent different types of floods, due to heavy and short rainfall, sustained rainfall over a period of days, severe tropical storms, and floods due to opening of dams. Other than that, the flood events were also from different geographical regions of the world. All the Sentinel-1 data were prepossessed using the same steps as mentioned in Sect. 2.1 to match the training data. The additional channel containing the value of $(VV + VH)/(VV - VH)$ for every pixel was also added to each test image. The Sentinel-1 images were divided into tiles of 256×256. The trained models were then applied to detect the water pixels in both the before-flood and after-flood images in each of the datasets. The labeled reference water masks for these datasets for evaluation were created by hand labeling using Sentinel-2 data from the exact same dates as reference. The resulting segmentation maps were then recombined to

the original scene size using a tapered cosine function as described by Wieland and Martinis (2019) to reduce the prediction errors close to the tile borders.

2.3 Model Architecture

The state-of-the-art models for image segmentation are variants of the encoder – decoder architecture like U-Net (Ronneberger et al. 2015). However, it has been shown in (Zhou et al. 2018) that more effective image segmentation architectures like nested UNet or UNet++ with nested dense convolutional blocks and dense skip connections can more effectively capture fine-grained details. The main idea behind UNet++ is to bridge the semantic gap between the feature maps of the encoder and decoder prior to fusion. To enable deep supervision, a 1×1 convolutional layer followed by a sigmoid activation function was appended to each of the target nodes. As a result, UNet++ generates four segmentation maps given an input image, which will be further averaged to generate the final segmentation map. Similar to (Zhou et al. 2018), we set the pruning level to L^3.

In this study, we use EfficientNet-B7 as the encoder architecture for the UNet++ segmentation network. As men-

Fig. 3 UNet++ based segmentation model with L^3 pruning investigated in the study (Zhou et al. 2018)
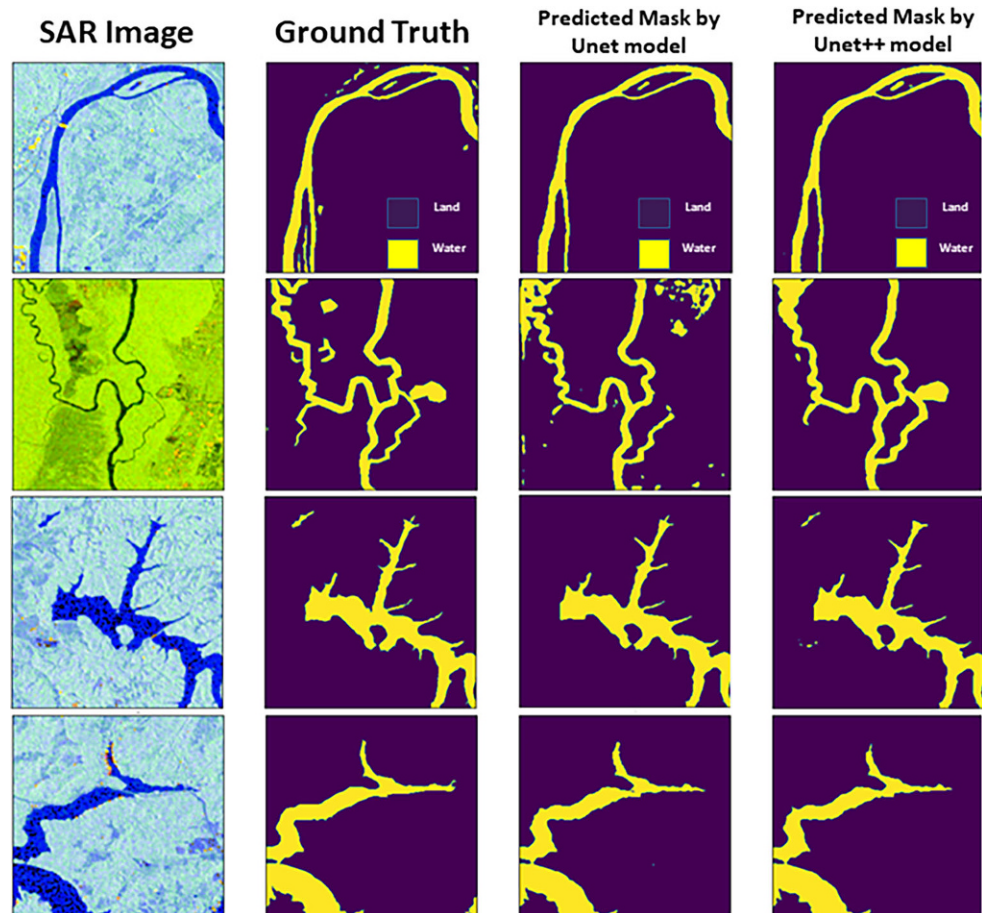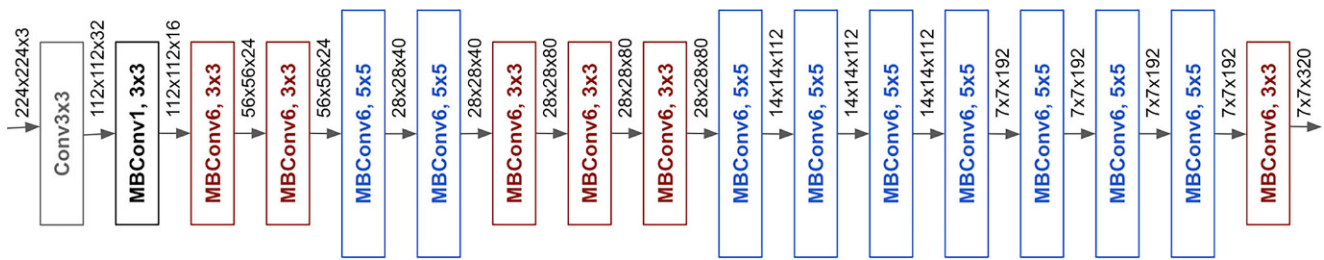
Fig. 4 Network architecture of EfficientNet-B0 (Tan and Le 2020)

tioned in (Tan and Le 2020), to maximize the model accuracy for any given resource constraints, an optimization problem can be designed with parameters w, d, and r, which are coefficients for scaling network width, depth, and resolution, respectively. The authors of the paper (Tan and Le 2020) also use neural architecture search to design a new baseline network and scale it up to obtain a family of models, called EfficientNets, which achieve much better accuracy and efficiency than previous convolutional neural networks.

In (Tan and Le 2020), the authors introduce a new compound scaling method which uses a compound coefficient ϕ to uniformly scale network width (β^{ϕ}), depth (α^{ϕ}), and resolution (γ^{ϕ}), where α, β, and γ are constants that can be determined by a small grid search. Intuitively, ϕ is a user-specified coefficient that controls how many more resources are available for model scaling, while α, β, and γ specify how to assign these extra resources to network width, depth, and resolution, respectively.

Fig. 4 shows the architecture of EfficientNet-B0. Its main building block is mobile inverted bottleneck MBConv from the paper (Sandler et al. 2019), to which squeeze-and-excitation optimization, as mentioned in (Hu et al. 2019), is added. Starting from the baseline EfficientNet-B0, a compound scaling method is added to scale it up with two steps:

- $\phi = 1$ is first fixed, assuming twice more resources available, and a small grid search of α, β, and γ is done.
- α, β, and γ are fixed as constants and the baseline network is scaled up with different ϕ, to obtain EfficientNet-B1 to B7.

In particular, EfficientNet-B7 has achieved state-of-the-art 84.3% top-1 accuracy on the ImageNet dataset (Deng et al. 2009), while being 8.4x smaller and 6.1x faster on inference than the best existing convolution network. The EfficientNet also transfers well and achieves state-of-the-art accuracy on CIFAR-100 (91.7%) (Krizhe-vsky 2009) dataset.

2.4 Model Training Details

For our analysis, we use EfficientNet-B7 as the encoder backbone for the nested UNet. The other models for comparison are Resnet34, InceptionV3, and Efficient net-B7, all having the same baseline Unet architecture. The models are retrained on our filtered dataset and the weights of the pre-trained networks are also fine-tuned by continuing the back-propagation. For the entire study, the mini-batch size was selected as 64 and iterated over the whole dataset for 100 epochs, for all the models. The Adam optimizer, as introduced in (Kingma and Ba 2017), was used for training optimization, with a learning rate of 3e-4. As mentioned in (Zhou et al. 2018), a combination of binary cross-entropy and dice coefficient as the loss function for each of the four semantic levels was added.

2.5 Model Evaluation Metrics

For our model evaluation, metrics like accuracy, precision, recall, F1 score, intersection over union (IoU), and kappa have been used. The mIoU is the average between the IoU of the segmented objects over all the images of the dataset.

For an image to be classified accurately, both precision and recall should be high. For this purpose, F1 score and mIoU are often used as a tradeoff metric to quantify both over- and under-segmentation into one measure. While training, a sum of F1 score and mIoU, which is the model evaluation score, is used as a metric for evaluating the model while training.

A modified K-fold cross validation approach was used to evaluate the performance of the models. The filtered dataset was randomly divided into 10 equal subsets. For each round, the models were trained on 9 subsets randomly and validated on the remaining subset. The process was repeated for k=10 by randomly selecting 9 subsets for training and the remaining one for validation. The model with the highest model evaluation score was selected and its performance was evaluated on the validation dataset. This was repeated for all the models with different architectures. The training was performed on a server with three nVidia GP104GL (Quadro P4000) GPUs, with driver NVIDIA UNIX x86.64

Kernel Module 460.56. The whole model development and training process was performed using Tensorflow along with the Keras library in Python.

2.6 Importance of Polarization Band Combinations Using Shapley Score

Shapley values are a metric for feature importances for machine learning models (Lundberg and Lee 2017). This method requires retraining the model on all feature subsets $S \subseteq F$, where F is the set of all features. It assigns an importance value to each feature that represents the effect on the model prediction of including that feature. To compute this effect, a model $f_{\cup i}$ is trained with that feature present, and another model f_S is trained with the feature withheld. Predictions from the two models are then compared on the current input $f_{\cup i}(x_{\cup i}) - f_S(x_S)$, where x_S represents the values of the input features in the set S. Since the effect of withholding a feature depends on other features in the model, the preceding differences are computed for all possible subsets $S \subseteq F$ (Lundberg and Lee 2017). The Shapley values are then computed and used as feature attributions. They are a weighted average of all possible differences:

$$\phi_i = \sum_{S \subseteq F} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_{S \cup i}(x_{S \cup i}) - f_S(x_S)] \quad (1)$$

In this paper, the Shapley scores of each model have been calculated for all possible polarization band combinations:

- VV, VH, (VV + VH)/(VV − VH)
- only VV
- only VH
- only (VV + VH)/(VV − VH)
- VV, VH

- VV, (VV + VH)/(VV − VH)
- VH, (VV + VH)/(VV − VH)

The scores for each of these band combinations denote how much contribution each band combination has towards flood segmentation using the given model.

3 Results

3.1 Training Results

The progression of training and validation loss and training and validation IoU, for the best UNet++ model with EfficientNet-B7 as encoder, over 15 epochs after K-fold cross-validation, is depicted in Fig. 5.

3.2 Results on the NASA Dataset

The results of the performances of the best models of the baseline UNet with Resnet34, InceptionV3, and EfficientNet-B7, and the UNet++ with EfficientNet-B7, after K-fold cross-validation, on a few of the validation images from Florence are shown in Fig. 6.

Based on the labels from the validation data, metrics like accuracy, precision, recall, F1 score, intersection over union (IoU), and kappa were calculated and the results are depicted in Table 2.

To assess whether the performances of the models are significantly different with each other, the McNemar's test was used (McNemar 1947). McNemar's test is used to compare the predictive accuracy between two models based on a contingency table of the two model's predictions. The test is performed taking into account exactly which cases the first model predicted correctly where the second model predicted wrong and vice versa. The p-values were calcu-

Fig. 5 **a** shows the progression of training and validation loss and **b** shows the progression of training and validation IoU for the UNet model with EfficientNet-B7 as encoder over 15 epochs
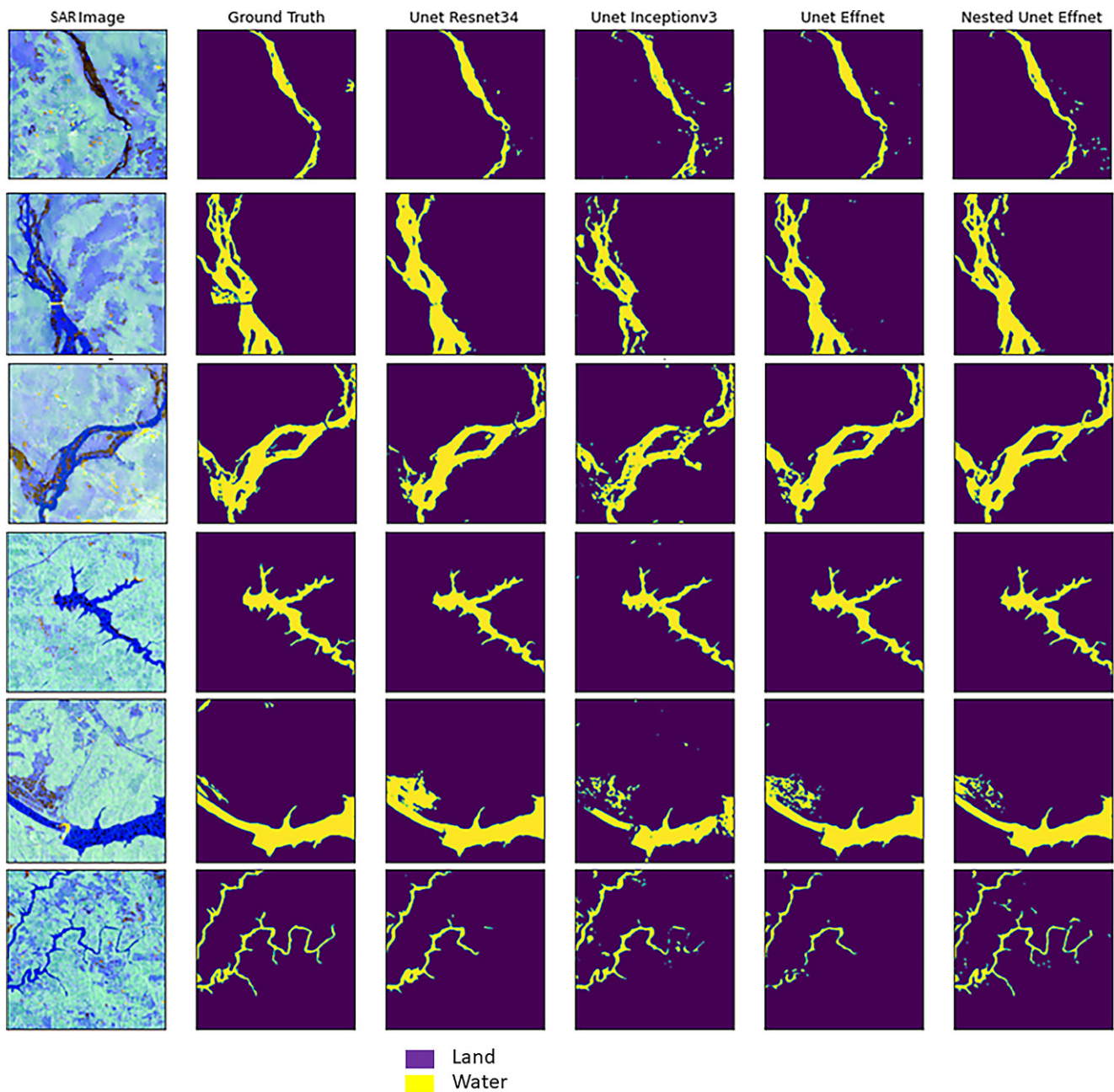
Fig. 6 The SAR image, the corresponding ground truth, and the prediction results for the trained baseline UNet models with Resnet34, InceptionV3, and EfficientNet-B7, and the UNet++ with EfficientNet-B7 for the validation set from Florence flood event

Table 2 Comparison of performance metrics of the baseline UNet with Resnet34, InceptionV3, and Efficientnet-B7, and the UNet++ with EfficientNet-B7 on the validation data from Florence

Performance metric	UNet with Resnet34	UNet with InceptionV3	UNet with EfficientNet-B7	UNet++ with EfficientNet-B7
Accuracy	98.1%	97.9%	98.3%	98.8%
Precision	85.1%	80.05%	87.02%	89.5%
Recall	88.25%	87.8%	88.9%	89.1%
F1	86.6%	83.74%	87.94%	89.3%
IoU	73.1%	71.7%	75.06%	75.76%
Kappa	79%	74.5%	80.7%	81.6%

Table 3 Comparison of pairwise p-values among the baseline UNet with Resnet34, InceptionV3, and EfficientNet-B7, and the UNet++ with EfficientNet-B7 on the validation data from Florence

Resnet34 (UNet)-InceptionV3 (UNet)	Resnet34 (UNet)-EfficientNet-B7 (UNet)	Resnet34 (UNet)-EfficientNet-B7 (UNet++)	InceptionV3 (UNet)-Efficient-Net-B7 (UNet)	InceptionV3 (UNet)-EfficientNet-B7 (UNet++)	EfficientNet-B7 (UNet)-EfficientNet-B7 (UNet++)
0.32	0.45	0.13	0.15	0.07	0.5

Fig. 7 Figure showing the IoU scores for each model trained with the individual bands VV, VH, $(VV + VH)/(VV - VH)$, and all possible combinations for the Florence dataset
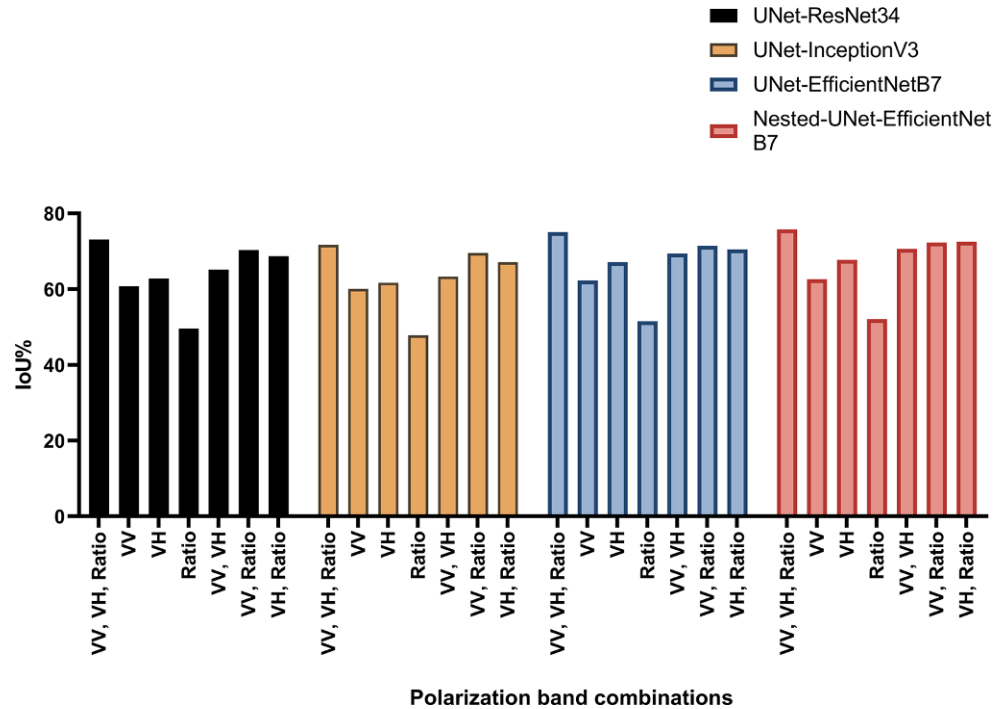
Fig. 8 Figure showing the absolute Shapley (SHAP) score for VV, VH, $(VV+VH)/(VV-VH)$, and their combinations for each of the models for the validation set from Florence
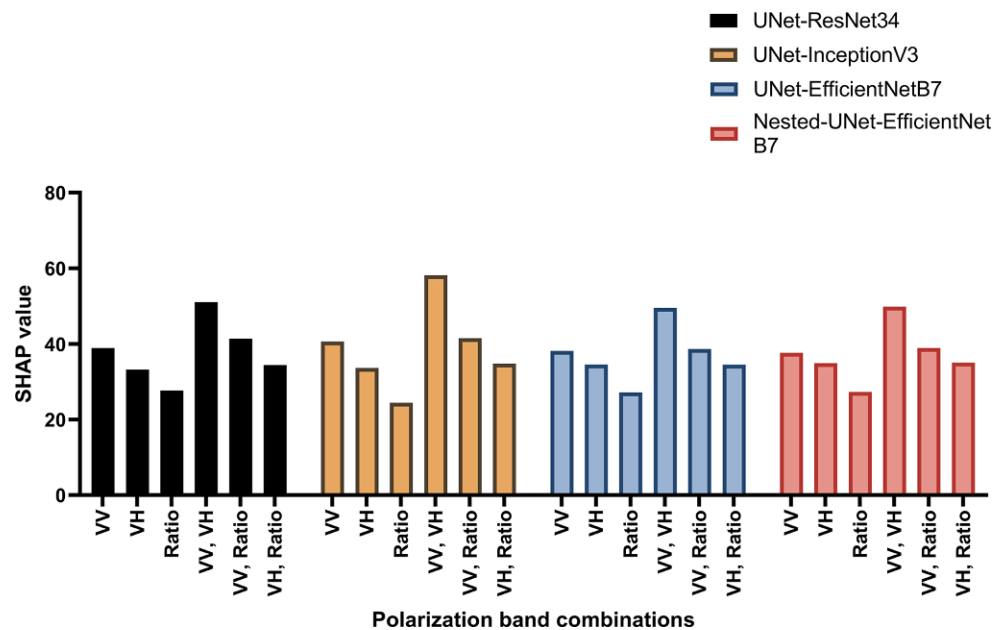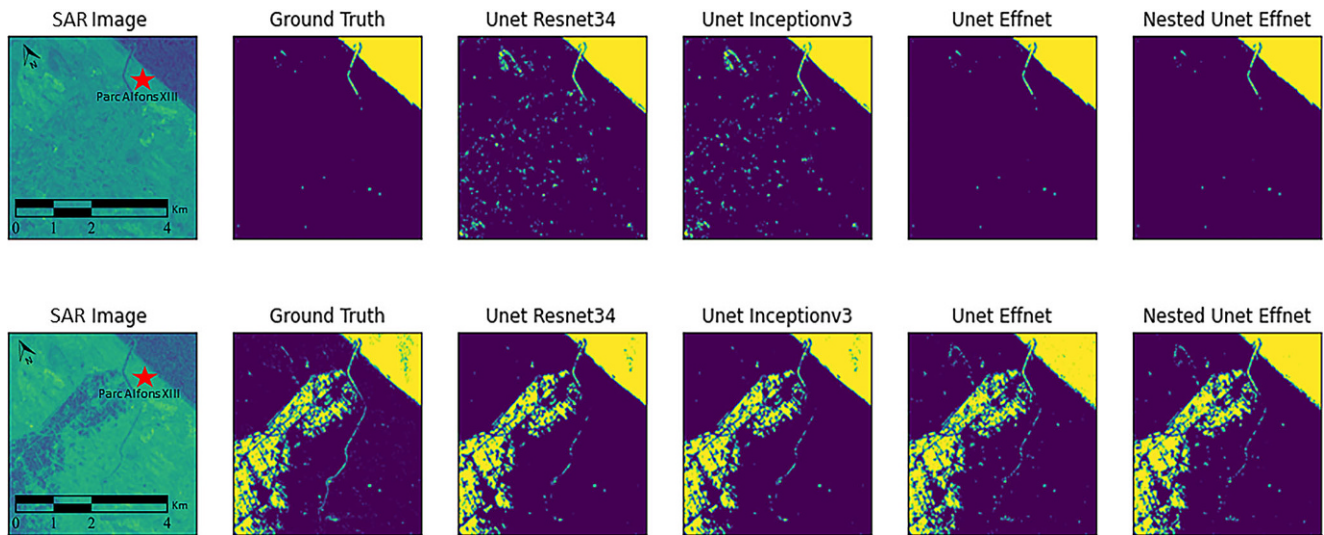
Fig. 9 The SAR image, the corresponding ground truth, and the prediction results for the trained baseline UNet with Resnet34, InceptionV3, and EfficientNet-B7 models, and the UNet++ with Efficientnet-B7 model, for the pre-flood and post-flood Sentinel-1 image from the case study in Spain 2019. The first row consists of the pre-flood results while the second row highlights the post-flood results
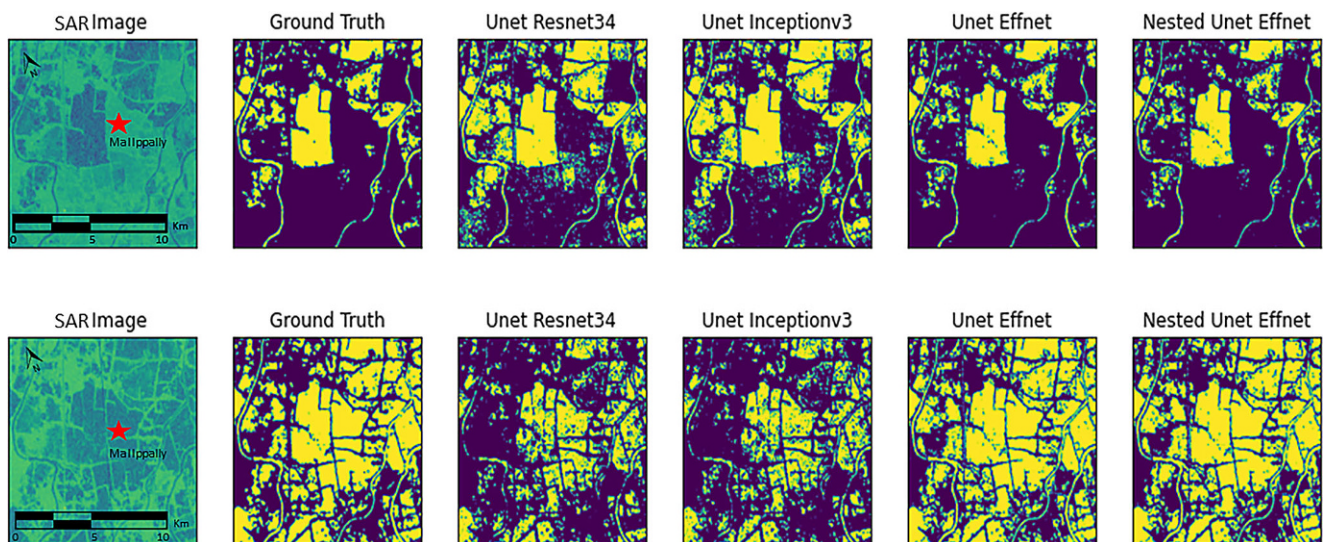

Fig. 10 The SAR image, the corresponding ground truth, and the prediction results for the trained baseline UNet with Resnet34, InceptionV3, and EfficientNet-B7 models, and the UNet++ with EfficientNet-B7 model, for the pre-flood and post-flood Sentinel-1 image from the case study in Kerala 2018. The first row consists of the pre-flood results while the second row highlights the post-flood results

lated for the pairwise McNemar test for each model, with the significance level kept at a standard value of 0.05. The pairwise p-values are shown in table. From the Table 3, it can be seen that in all cases, the p-value is larger than the assumed significance threshold of 0.05. Therefore, it can be concluded that there are no significant differences among the different models.

To check how the IoU scores vary depending on the combinations of bands used, the models were trained on the training set repeatedly, each time using a different band combination, as mentioned in Sect. 2.6. In each case, the

trained models were then used to predict the flood map of the Florence validation set and the IoU was calculated as shown in Fig. 7. From Fig. 7, it can be seen that for all four models, the model trained on the data having all three bands performs the best, and the model trained on only VH band outperforms the model with only the VV band. Also, it can be seen that the addition of the ratio $(VV + VH)/(VV - VH)$ band enhances the performance of the model.

As mentioned in Sect. 2.6, the absolute Shapley scores were calculated to get an idea of the marginal contribution of each of the polarization bands and their combinations to
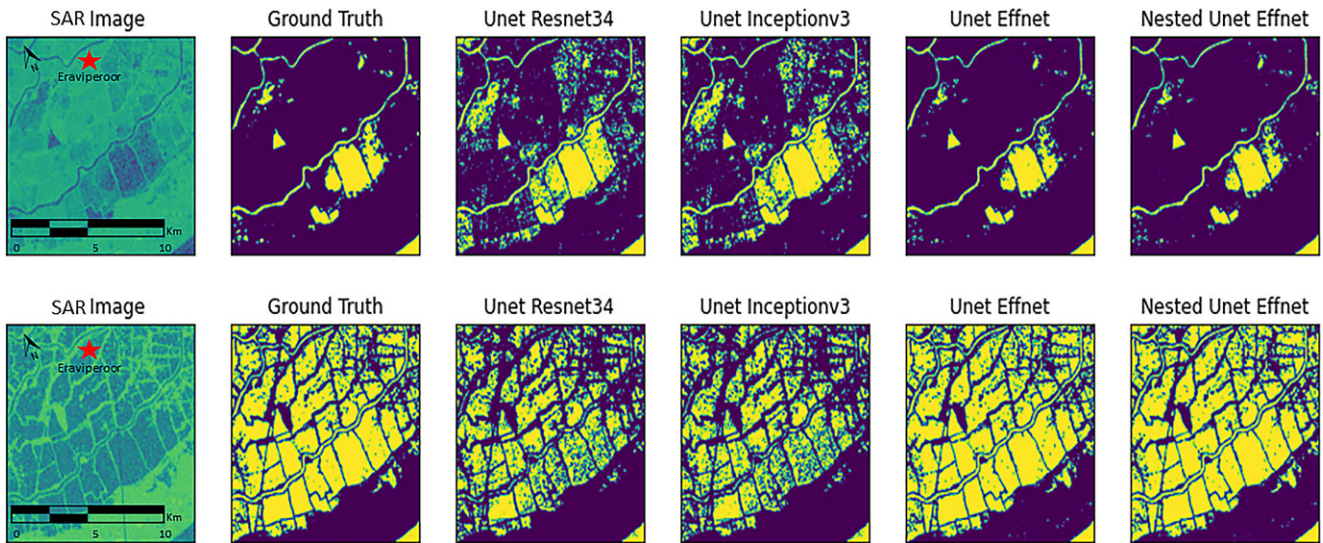
Fig. 11 The SAR image, the corresponding ground truth, and the prediction results for the trained baseline UNet with Resnet34, InceptionV3, and EfficientNet-B7 models, and the UNet++ with EfficientNet-B7 model, for another pre-flood and post-flood Sentinel-1 image from the case study in Kerala 2018. The first row consists of the pre-flood results while the second row highlights the post-flood results
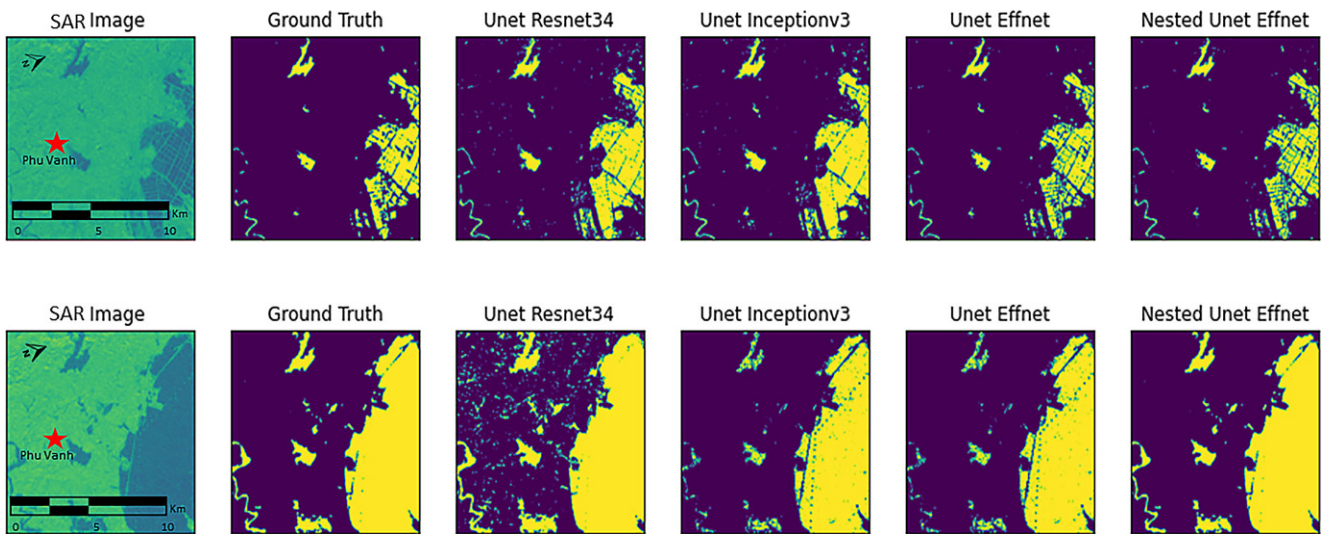

Fig. 12 The SAR image, the corresponding ground truth, and the prediction results for the trained baseline UNet with Resnet34, InceptionV3, and EfficientNet-B7 models, and the UNet++ with EfficientNet-B7 model, for the pre-flood and post-flood Sentinel-1 image from the case study in Vietnam 2020. The first row consists of the pre-flood results while the second row highlights the post-flood results

the final prediction and shown in Fig. 8. As can be seen from Fig. 8, for all of the models, the highest marginal contribution for each of the models is from the combination of VV and VH bands, and the individual contribution of the VV band is higher than that of the VH band in all cases. The inclusion of the ratio $(VV + VH)/(VV - VH)$ with VV or VH bands contributes more than the VV or VH bands individually. In case of the model having the backbone of the UNet-Inceptionv3, the contribution of the VV, VH band combination is the highest, compared to the other models for this specific dataset.

3.3 Results on the Real Dataset

The results of the performances of the baseline UNet with Resnet34, InceptionV3, and EfficientNet-B7 models, and the UNet++ with EfficientNet-B7, on some of the pre-flood and post-flood datasets from Spain, Kerala, and Vietnam are shown in Figs. 9–13.

Based on the reference masks hand-labeled from Sentinel-2 data, metrics like accuracy, precision, recall, F1 score, intersection over union (IoU), and kappa were cal-
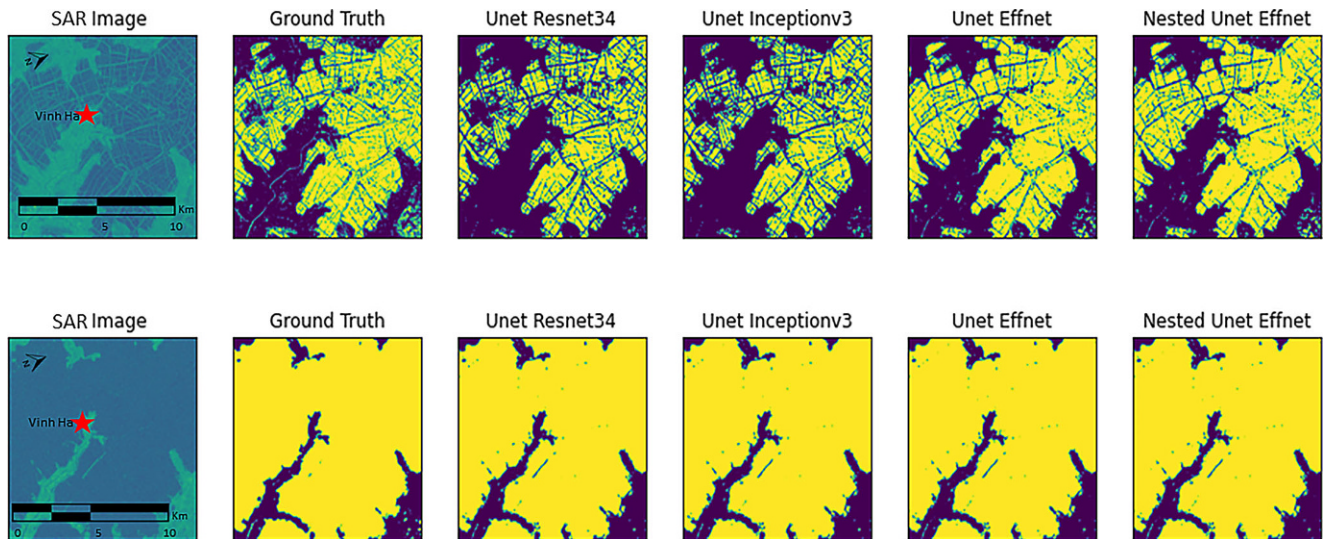
Fig. 13 The SAR image, the corresponding ground truth, and the prediction results for the trained baseline UNet with Resnet34, InceptionV3, and EfficientNet-B7 models, and the UNet++ with EfficientNet-B7 model, for another pre-flood and post-flood Sentinel-1 image from the case study in Vietnam 2020. The first row consists of the pre-flood results while the second row highlights the post-flood results

culated for all the datasets and the results are depicted in Tables 4–7 respectively.

To assess whether the performances of the models are significantly different from each other, the McNemar's test was used as before. The pairwise *p*-values for the four different datasets are shown in Table 8. From the Table 8, it can be seen that in all cases, the *p*-value is larger than the assumed significance threshold of 0.05. Therefore, it can be concluded that there are no significant differences among the different models.

3.3.1 Comparison of IoU Scores for Models with Different Band Combinations

As mentioned in Sect. 3.2, to check how the IoU scores vary depending on the combinations of bands used, the models were trained on the training set repeatedly, each time using a different band combination. In each case, the trained models were then used to predict the flood map of the test datasets from Spain, Kerala, Bihar, and Vietnam. The IoU

Table 4 Comparison of performance metrics of the baseline UNet with Resnet34, InceptionV3, and EfficientNet-B7, and the UNet++ with EfficientNet-B7, on the Sentinel-1 data from Spain

Performance metric	UNet with Resnet34	UNet with InceptionV3	UNet with EfficientNet-B7	UNet++ with EfficientNet-B7
Accuracy	98.6%	98.1%	98.7%	98.8%
Precision	81.6%	79.3%	82.4%	82.8%
Recall	85.3%	83.4%	86.1%	86.4%
F1	83.4%	81.3%	84.2%	84.5%
IoU	71.5%	71.1%	72.9%	73.0%
Kappa	79.8%	78.7%	80.4%	80.5%

Table 5 Comparison of performance metrics of the baseline UNet with Resnet34, InceptionV3, and EfficientNet-B7, and the UNet++ with EfficientNet-B7, on the Sentinel-1 data from Kerala

Performance metric	UNet with Resnet34	UNet with InceptionV3	UNet with EfficientNet-B7	UNet++ with EfficientNet-B7
Accuracy	96.8%	96.3%	97.1%	97.7%
Precision	82.3%	80.1%	83.7%	84.0%
Recall	83.8%	80.5%	87.0%	87.3%
F1	83%	80.3%	85.3%	85.6%
IoU	69.7%	66.8%	73.5%	74.1%
Kappa	78.5%	75.6%	80.3%	80.8%

Table 6 Comparison of performance metrics of the baseline UNet with Resnet34, InceptionV3, and EfficientNet-B7, and the UNet++ with EfficientNet-B7, on the Sentinel-1 data from Bihar

Performance metric	UNet with Resnet34	UNet with InceptionV3	UNet with EfficientNet-B7	UNet++ with EfficientNet-B7
Accuracy	97.7%	97.5%	98.3%	98.5%
Precision	85.5%	81.3%	88.5%	89.7%
Recall	88.4%	87.6%	89.1%	89.4%
F1	86.9%	84.3%	88.8%	89.5%
IoU	71.7%	70.2%	73.9%	74.7%
Kappa	78.7%	77.5%	79.8%	80.3%

Table 7 Comparison of performance metrics of the baseline UNet with Resnet34, InceptionV3, and EfficientNet-B7, and the UNet++ with EfficientNet-B7 on the Sentinel-1 data from Vietnam

Performance metric	UNet with Resnet34	UNet with InceptionV3	UNet with EfficientNet-B7	UNet++ with EfficientNet-B7
Accuracy	98.1%	97.7%	98.7%	98.9%
Precision	83.4%	78.9%	84.0%	86.2%
Recall	84.9%	81.4%	85.7%	86.2%
F1	84.1%	80.1%	84.8%	86.2%
IoU	71.7%	70.1%	73.7%	74.0%
Kappa	80.7%	78.7%	81.2%	81.6%

Table 8 Comparison of pairwise p-values among the baseline UNet with Resnet34, InceptionV3, and EfficientNet-B7, and the UNet++ with EfficientNet-B7, on the four different test datasets

Dataset	Resnet34 (UNet)-InceptionV3 (UNet)	Resnet34 (UNet)-EfficientNet-B7 (UNet)	Resnet34 (UNet)-EfficientNet-B7 (UNet++)	InceptionV3 (UNet)-EfficientNet-B7 (UNet)	InceptionV3 (UNet)-EfficientNet-B7 (UNet++)	EfficientNet-B7 (UNet)-EfficientNet-B7 (UNet++)
Spain	0.33	0.26	0.18	0.22	0.16	0.47
Kerala	0.39	0.34	0.15	0.29	0.07	0.41
Bihar	0.38	0.32	0.12	0.21	0.11	0.45
Vietnam	0.28	0.36	0.23	0.19	0.08	0.42

scores were calculated for each case study separately and shown in Fig. 14.

From Fig. 14, it can be seen that in all cases, the model trained on the data with all three bands performs the best.

The Shapley values were then calculated using Eq. (1) and shown in Fig. 15. As can be seen from Fig. 15, the highest marginal contribution for each of the models is from the combination of VV and VH bands in all cases.

4 Discussion

In this work, the main objective was to leverage the huge amount of publicly available Sentinel-1 data to delineate open water bodies which can be further used in flood extent mapping in varied geographical areas of the world. In the literature study shown in Sect. 1, almost all of the related works dealt mostly with the techniques applied but not with the generalizability of the methods in different varied geographical and topographical conditions.

4.1 Model Architecture

The UNet models with Resnet34, InceptionV3, and EfficientNet-B7 architectures perform comparably to the UNet++ with EfficientNet-B7 backbone, as is visible from the results on both the NASA dataset as well as on the real test cases. In both instances, there were no significant differences among the results from the four different architectures. One reason behind this may be that deeper encoders with more layers may not necessarily improve the predictive capabilities of the model. A visual comparison of the segmentation results in Figs. 6, 9–13 suggests that there may not be any significant differences among the results of the UNet models with Resnet34, InceptionV3, and EfficientNet-B7 architectures; however, the UNet with InceptionV3 architecture tends to produce noisy predictions in some fine-edge cases. This can be aligned with the results from Bai et al. (2021), where it was also reported that deep learning architectures like UNets tend to produce similar results to convolutional neural networks (CNN). One argument behind the UNet++ not producing signifi-
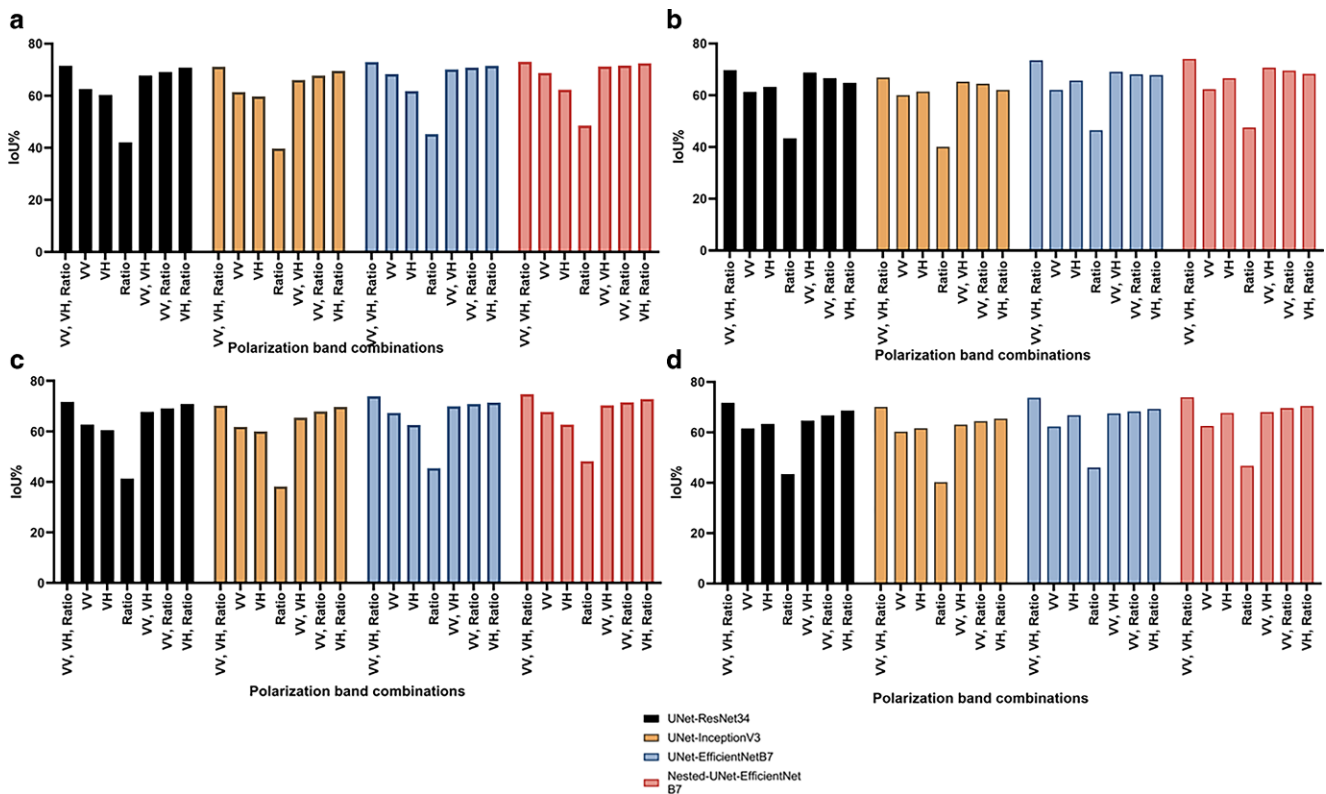
Fig. 14 a–d Show the IoU scores for each model trained with the individual bands VV, VH, $(VV+VH)/(VV-VH)$, and all possible combinations tested on the real datasets from Spain, Kerala, Bihar, and Vietnam, respectively

cantly better predictions than the UNet models can be that although the UNet++ employs the multidepth encoders and multiscale feature map fusion which may help with correct classification of areas prone to misclassification, the inherent multiscale information flow in UNet architectures, that is intrinsic to any encoder, is sufficient for this task. Hence, using a dedicated architecture focusing on multiscale fusion does not improve the results significantly. This also highlights another point, namely that the model architecture may not be the limiting parameter for obtaining better segmentation results at this stage. To further improve the performance of the models, adding additional training data, with a focus on hard examples, may be more beneficial.

4.2 Polarization

The segmentation results also depend strongly on the choice of polarization used for training and inference. It has been shown that using both polarizations, VV and VH, is necessary for improved detection of flooded areas. Additionally, adding a third band as a ratio of two VV+VH and VV-VH can add information in certain scenarios. A visual inspection of a scene from the Florence set in the NASA dataset gives an idea regarding how the models can per-

form differently after being trained and tested on different combinations of polarization bands, as shown in Fig. 16.

In Fig. 16, for the first scene (first row) the combination of the VV, VH and the $(VV + VH)/(VV - VH)$ bands perform superiorly compared to the model with only the VH band. The combination of the VH and VH bands as well as the just the VV band performs worse. In the second case (second row), the models with only the VH band, both VV, VH bands together and using the combination of the VV, VH and the $(VV + VH)/(VV - VH)$ bands perform almost comparably and better than the model with only the VV band.

The effect of polarization varies across different test cases from around the world, as is evident from the IoU score of the different models from four different real test cases with different polarization band combinations.

- In Fig. 14a, in the case of the test dataset from Spain, the models trained on only the VV band have a higher IoU% than the models trained on only the VH band. The combination of the VH and the ratio band performs better than combination of the VV and the ratio band as well as the combination of VV and VH bands.
- In Fig. 14b, in the case of the test dataset from Kerala, the models trained on only the VH band have a higher
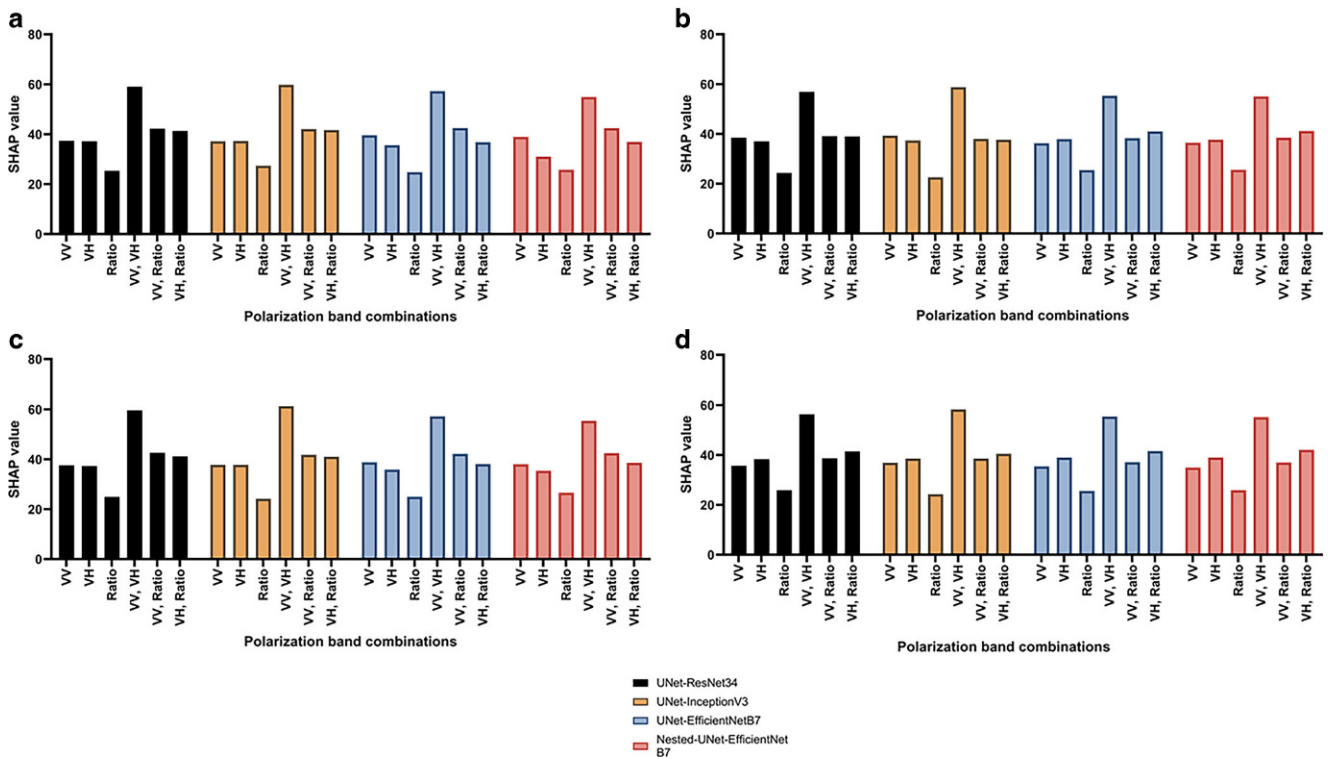
Fig. 15 **a–d** Show the Shapley (SHAP) scores for each model trained with the individual bands VV, VH, $(VV + VH)/(VV - VH)$, and all possible combinations tested on the real datasets from Spain, Kerala, Bihar, and Vietnam respectively

IoU% than the models trained on only the VV band. The combination of the VV and VH bands performs better than combination of the ratio band with either just the VV or the VH bands.

- In Fig. 14c, in the case of the test dataset from Bihar, the models trained on only the VV band have a higher IoU% than the models trained on only the VH band. The combination of the VH and the ratio band performs better than combination of the VV and the ratio band as well as the combination of VV and VH bands.
- In Fig. 14d, in the case of the test dataset from Vietnam, the models trained on only the VH band have a higher IoU% than the models trained on only the VV band. The combination of the VH and the ratio band performs better than combination of the VV and the ratio band as well as the combination of VV and VH bands.

4.3 Marginal Contribution From Each Polarization Band to the Final Prediction

Shapley values are an important metric for measuring the marginal contribution of each of the polarization band combinations towards the final prediction. Similar to the polarization bands, the marginal contribution of the these bands

towards the final prediction also varies across different test cases from around the world.

- In Fig. 15a, in the case of the test dataset from Spain, the contributions of the VV band and the VH band are almost equal for the UNet-ResNet34 and UNet-InceptionV3 models, and the contribution of the combination of the VV and ratio bands is marginally higher than that of the combination of the VH and ratio bands. For the UNet-EfficientNetB7 and Nested-UNet-EfficientNetB7 models, the marginal contribution of the VV band is higher than that of the VH band and the contribution of the combination of VV and ratio bands is higher than that of the combination of VH and ratio bands.
- In Fig. 15b, in the case of the test dataset from Kerala, the contribution of the VV band is marginally higher than that of the VH band for the UNet-ResNet34 and UNet-InceptionV3 models, and the contribution of the combination of the VV and ratio bands is almost equal to that of the combination of the VH and ratio bands. For the UNet-EfficientNetB7 and Nested-UNet-EfficientNetB7 models, the marginal contribution of the VH band is higher than that of the VV band and the contribution of the combination of VH and ratio bands is higher than that of the combination of VV and ratio bands.

Fig. 16 The SAR image, the corresponding ground truth, and the prediction results for the UNet++ with EfficientNet-B7 model on two scenes from Florence in the NASA dataset using only the VV band, only the VH band, both VV and VH bands together, and using a combination of the VV, VH, and the $(VV + VH)/(VV - VH)$ bands

5 Conclusion
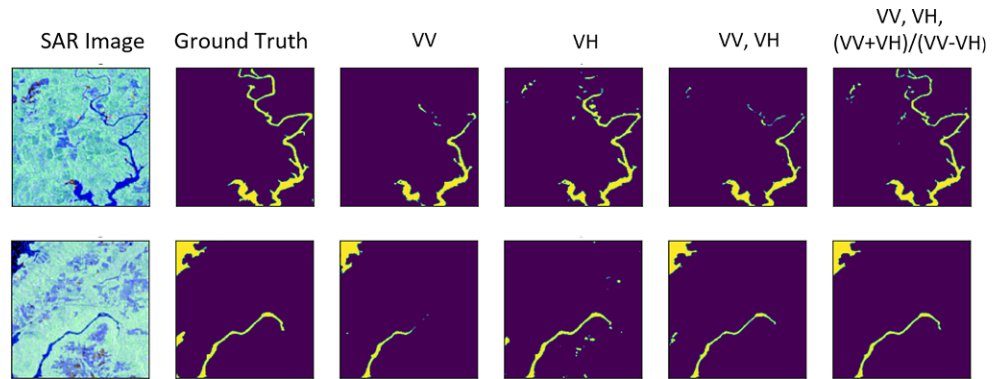
- In Fig. 15c, in the case of the test dataset from Bihar, the contributions of the VV band and the VH band are almost equal for the UNet-ResNet34 and UNet-InceptionV3 models, and the contribution of the combination of the VV and ratio bands is marginally higher than the combination of the VH and ratio bands. For the UNet-EfficientNetB7 and Nested-UNet-EfficientNetB7 models, the marginal contribution of the VV band is higher than that of the VH band and the contribution of the combination of VV and ratio bands is higher than that of the combination of VH and ratio bands.

- In Fig. 15d, in the case of the test dataset from Vietnam, for all models, the contribution of the VH band is marginally higher than that of the VV band for the UNet-ResNet34 and UNet-InceptionV3 models, and the contribution of the combination of the VH and ratio bands is almost equal to the combination of the VV and ratio bands.

  Previous studies (Baghdadi et al. 2001; Henry et al. 2006a) have shown that VV polarization is the most efficient channel for flood detection compared to the VH polarization signal due to the double bounce property (Ferro et al. 2011). This is because the backscatter values given by VV polarization are highly distinct in wet areas. However, results obtained through VH polarization are quite different for different weather conditions, as VH polarization is most sensitive to surface roughness conditions (Henry et al. 2006b), such as the roughness of water surfaces due to wind conditions. With regards to the comparison between only VV and VH polarization bands for flood detection, our results show that neither VV nor VH singularly can be considered as the preferred input for the purpose of flood mapping using deep learning models. Similar to the observations by Katiyar et al. (2021) and Helleis et al. (2022), the combination of VV and VH polarized data can be seen to perform superiorly compared to only one of the bands alone. This is also shown to hold true for more complex cases like flood mapping in urban areas Pelich et al. (2022).

In this study, the performance of the UNet++ with EfficientNet-B7 encoder is compared with three other state-of-the-art UNet-based segmentation models in the task of flood mapping based on SAR satellite images. Moreover, the impact of polarization band combinations on the performance of the trained models was also assessed. Overall, we can summarize our conclusions from our experiments with these points:

- The UNet models with Resnet34, InceptionV3, and EfficientNet-B7 architectures perform comparably to the UNet++ with EfficientNet-B7 backbone, as is visible from the results on both the NASA dataset as well as on the real test cases. In both cases, there were no significant differences among the results from the four different architectures.

- The segmentation results also depend strongly on the choice of polarization used for training and inference. However, the effect of polarization still varies across different test cases from around the world in terms of performance; the model trained with the combinations of the individual bands VV and VH and $(VV + VH)/(VV - VH)$ gives the best results.

- The highest marginal contribution for each of the models, as is evident from the Shapley values, is from the combination of VV and VH bands in all cases. Thus, calculation of the Shapley values has proven to be an effective measure of the marginal contribution of individual polarization bands as well as band combinations for flood detection using deep learning architectures.

- We have also shown how the flood dataset provided by the NASA Interagency Implementation and Advanced Concepts Team can be used as a benchmark dataset for training deep learning models. One of the main highlights is that these models can be trained on certain flood events provided in the dataset and used for flood detection in other geographical areas, thus proving the transferability of these models.

In this work, although only Sentinel-1 data have been used, these methods should also transfer to other SAR sensors as well. To handle problems with more complex urban, arid, and mountainous environments, further research may be carried out to check whether additional information, like slope or land cover information, or even extending the training set with more example scenes from these environments could help improve the results and increase the robustness of the models.

Availability of Data and Material The whole training dataset was obtained by the authors of this study from the IMPACT 2021 ETCI Competition on Flooding Detection GitHub page NASA (2021).

Conflict of interest The authors agree that there are no conflicts of interests.

References

(2019) Floods in southeastern spain. https://www.efas.eu/en/news/floods-southeastern-spain-september-2019. Accessed on 2023-05-06

Baghdadi N, Bernier M, Gauthier R, Neeson I (2001) Evaluation of c-band sar data for wetlands mapping. Int J Remote Sens 22(1):71–88. https://doi.org/10.1080/014311601750038857

Bai Y, Wu W, Yang Z, Yu J, Zhao B, Liu X, Yang H, Mas E, Koshimura S (2021) Enhancement of detecting permanent water and temporary water in flood disasters by fusing sentinel-1 and sentinel-2 imagery using deep learning algorithms: demonstration of Sen1floods11 benchmark Datasets. Remote Sens 13(11):2220. https://doi.org/10.3390/rs13112220

Bonafilia D, Tellman B, Anderson T, Issenberg E (2020) Sen1Floods11: a georeferenced dataset to train and test deep learning flood algorithms for Sentinel-1. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp 835–845 https://doi.org/10.1109/CVPRW50498.2020.00113

Dasgupta A, Grimaldi S, Ramsankaran R, Pauwels VR, Walker JP (2018) Towards operational sar-based flood mapping using neuro-fuzzy texture-based approaches. Remote Sens Environ 215:313–329. https://doi.org/10.1016/j.rse.2018.06.019 (https://www.sciencedirect.com/science/article/pii/S0034425718302979)

Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp 248–255 https://doi.org/10.1109/CVPR.2009.5206848

Downtoearth (2021) Not just climate change farakka also to blame for 2021 bihar floods. https://www.downtoearth.org.in/news/climate-change/not-just-climate-change-farakka-also-to-blame-for-2021-bihar-floods. Accessed on 2023-05-06

Ferro A, Brunner D, Bruzzone L, Lemoine G (2011) On the relationship between double bounce and the orientation of buildings in vhr sar images. IEEE Geosci Remote Sens Lett 8(4):612–616. https://doi.org/10.1109/LGRS.2010.2097580

Garg GSB, Motagh M, Haghshenas Haghighi M, Maghsudi S (2020) Automatic flood monitoring based on sar intensity and interferometric coherence using machine learning. In: EGU General Assembly 2020 EGU2020–12954, 4–8 May 2020. https://doi.org/10.5194/egusphere-egu2020-12954

Ghosh B, Garg S, Motagh M (2022) Automatic flood detection from sentinel-1 data using deep learning architectures. ISPRS Ann Photogramm Remote Sens Spatial Inf Sci 3:201–208. https://doi.org/10.5194/isprs-annals-v-3-2022-201-2022

Helleis M, Wieland M, Krullikowski C, Martinis S, Plank S (2022) Sentinel-1-based water and flood mapping: Benchmarking convolutional neural networks against an operational rule-based processing chain. IEEE J Sel Top Appl Earth Observations Remote Sensing 15:2023–2036. https://doi.org/10.1109/jstars.2022.3152127

Henry J, Chastanet P, Fellah K, Desnos Y (2006) Envisat multi-polarized asar data for flood mapping. Int J Remote Sens 27(10):1921–1929. https://doi.org/10.1080/01431160500486724

Henry J, Chastanet P, Fellah K, Desnos Y (2006) Envisat multi-polarized asar data for flood mapping. Int J Remote Sens 27(10):1921–1929. https://doi.org/10.1080/01431160500486724

Hu J, Shen L, Albanie S, Sun G, Wu E (2019) Squeeze-and-Excitation Networks. arXiv:170901507

Hub OA (2015) Open access hub. https://scihub.copernicus.eu/. Accessed on 2023-05-05

Insom P, Cao C, Boonsrimuang P, Liu D, Saokarn A, Yomwan P, Xu Y (2015) A support vector machine-based particle filter method for improved flooding classification. IEEE Geosci Remote Sens Lett 12(9):1943–1947. https://doi.org/10.1109/LGRS.2015.2439575

Katiyar V, Tamkuan N, Nagai M (2021) Near-real-time flood mapping using off-the-shelf models with sar imagery and deep learning. Remote Sens 13(12):2334. https://doi.org/10.3390/rs13122334

Kingma DP, Ba J (2017) Adam: A Method for Stochastic Optimization. arXiv:14126980

Konapala G, Kumar S (2021) Exploring Sentinel-1 and Sentinel-2 diversity for Flood inundation mapping using deep learning. Copernicus Meetings. Tech. Rep. EGU21-10445. https://doi.org/10.5194/egusphere-egu21-10445 (https://meetingorganizer.copernicus.org/EGU21/EGU21-10445.html, conference Name: EGU21)

Krizhe-vsky A (2009) Learning multiple layers of features from tiny images. Technical Report TR-2009. University of Toronto, Toronto, p 60

Lal P, Prakash A, Kumar A, Srivastava PK, Saikia P, Pandey A, Srivastava P, Khan M (2020) Evaluating the 2018 extreme flood hazard

events in kerala, india. Remote Sens Lett 11(5):436–445. https://doi.org/10.1080/2150704x.2020.1730468

Landuyt L, Van Wesemael A, Schumann GJP, Hostache R, Verhoest NEC, Van Coillie FMB (2019) Flood mapping based on synthetic aperture radar: An assessment of established approaches. IEEE Trans Geosci Remote Sens 57(2):722–739. https://doi.org/10.1109/TGRS.2018.2860054

Li Y, Martinis S, Wieland M (2019) Urban flood mapping with an active self-learning convolutional neural network based on terrasar-x intensity and interferometric coherence. ISPRS J Photogramm Remote Sens 152:178–191. https://doi.org/10.1016/j.isprsjprs.2019.04.014

Li Y, Martinis S, Wieland M, Schlaffer S, Natsuaki R (2019) Urban flood mapping using SAR intensity and Interferometric coherence via Bayesian network fusion. Remote Sens 11(19):2231. https://doi.org/10.3390/rs11192231 (https://www.mdpi.com/2072-4292/11/19/2231, number: 19 Publisher: Multidisciplinary Digital Publishing Institute)

Lundberg S, Lee SI (2017) A unified approach to interpreting model predictions

Martinis S, Kersten J, Twele A (2015) A fully automated terraSAR-X based flood service. Isprs J Photogramm Remote Sens 104: 203–212. https://doi.org/10.1016/j.isprsjprs.2014.07.014 (https://linkinghub.elsevier.com/retrieve/pii/S092427161-4001981)

McNemar Q (1947) Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 12(2):153–157

Muñoz DF, Muñoz P, Moftakhari H, Moradkhani H (2021) From local to regional compound flood mapping with deep learning and data fusion techniques. Sci Total Environ 782:146927. https://doi.org/10.1016/j.scitotenv.2021.146927 (https://www.sciencedirect.com/science/article/pii/S0048-969721019975)

Nanthini J et al (2017) An efficient detection of flood extent from satellite images using contextual features and optimized classification. https://www.ijsrd.com/Article.php?manuscript=IJSRDV4I110043. Accessed on 2022-08-08

NASA (2021) NASA's IMPACT Collaborates on Global Flood Detection Challenge | Earthdata. https://earthdata.nasa.gov/learn/articles/impact-flood-competition/. Accessed on 2021-10-09

Nemni E, Bullock J, Belabbes S, Bromley L (2020) Fully convolutional neural network for rapid flood segmentation in synthetic aperture radar imagery https://doi.org/10.3390/rs12162532

Otsu N (1975) A threshold selection method from gray-level histograms. IEEE Trans Syst, Man, Cybern 9:5

Ouled Sghaier M, Hammami I, Foucher S, Lepage R (2018) Flood extent mapping from time-series sar images based on texture analysis and data fusion. Remote Sens. https://doi.org/10.3390/rs10020237 (https://www.mdpi.com/2072-4292/10/2/237)

Pelich R, Chini M, Hostache R, Matgen P, Pulvirenti L, Pierdicca N (2022) Mapping floods in urban areas from dual-polarization insar

coherence data. IEEE Geosci Remote Sens Lett 19:1–5. https://doi.org/10.1109/lgrs.2021.3110132

Pradhan B, Tehrany MS, Jebur MN (2016) A new semiautomated detection mapping of flood extent from terrasar-x satellite image using rule-based classification and taguchi optimization techniques. Ieee Trans Geosci Remote Sens 54(7):4331–4342. https://doi.org/10.1109/TGRS.2016.2539957

Reliefweb (2020) Vietnam floods emergency appeal. https://reliefweb.int/report/viet-nam/vietnam-floods-emergency-appeal-n-mdrvn020-operation-update. Asssessed on 2022-03-05

Ritushree D, Garg S, Dasgupta A, Martinis S, Selvakumaran S, Motagh M (2023) Improving sar-based flood detection in arid regions using texture features. In: 2023 International Conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS), vol 1, pp 1–4 https://doi.org/10.1109/MIGARS57353.2023.10064526

Ronneberger O, Fischer P, Brox T (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:150504597

Rudner TG, Rußwurm M, Fil J, Pelich R, Bischke B, Kopačková V, Biliński P (2019) Multi3net: Segmenting flooded buildings via fusion of multiresolution, multisensor, and multitemporal satellite imagery. Proc AAAI Conf Artif Intell 33(01):702–709. https://doi.org/10.1609/aaai.v33i01.3301702

Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC (2019) MobilenetV2: inverted residuals and linear bottlenecks. arXiv:180104381

Sentinel-1 (2015) Sentinel-1 toolbox. https://sentinel.esa.int/web/sentinel/toolboxes/sentinel-1. Accessed on 2021-10-04

Skakun S (2010) A neural network approach to flood mapping using satellite imagery. Comput Inform 29:1013–1024

Tan M, Le QV (2020) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv:190511946

Tripathy P, Malladi T (2022) Global flood mapper: a novel google earth engine application for rapid flood mapping using sentinel-1 sar. Nat Hazards 114(2):1341–1363. https://doi.org/10.1007/s11069-022-05428-2

Wieland M, Martinis S (2019) A modular processing chain for automated flood monitoring from multi-spectral satellite data. Remote Sens 11(19):2330. https://doi.org/10.3390/rs11192330

Zhang P, Chen L, Li Z, Xing J, Xing X, Yuan Z (2019) Automatic extraction of water and shadow from SAR images based on a multi-resolution dense encoder and decoder network. Sensors 19(16):E3576. https://doi.org/10.3390/s19163576

Zhou Z, Siddiquee RMM, Tajbakhsh N, Liang J (2018) Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp 3–11 https://doi.org/10.1007/978-3-030-00889-5-1