# Flooded Areas Segmentation Using SAR Images Based on Deep Learning Models

Mohaddeseh Mesvari, Reza Shah-Hosseini *

School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran, (m.mesvari, rshahosseini)@ut.ac.ir

**Keywords:** Flood detection, Deep learning, PSPNet, MANet, PAN, SAR imagery.

**Abstract**

Floods represent one of the most devastating natural disasters worldwide, causing substantial human and economic losses, particularly in vulnerable urban and agricultural areas. The advent of high-resolution satellite data and deep learning techniques has significantly improved flood detection and monitoring capabilities. This study explores the efficacy of three advanced deep learning models—PAN, PSPNet, and MANet—for semantic segmentation of flooded areas using Sentinel-1 synthetic aperture radar (SAR) imagery. The ETCI 2021 competition dataset, comprising VV and VH polarization SAR data and corresponding flood masks, was employed to train and evaluate the models. Comprehensive experiments revealed that VH polarization outperforms VV across all models, yielding higher accuracy, precision, recall, F1-score, and Intersection over Union (IoU). Among the tested architectures, MANet demonstrated superior performance with an IoU of 84.21% and F1-score of 91.11%, attributed to its multi-scale and mutual attention mechanisms. These findings affirm the value of SAR imagery, particularly VH polarization, combined with deep learning for accurate flood detection and underscore the potential for real-time disaster response applications.

## 1. Introduction

Floods, as one of the most destructive natural disasters, cause extensive human and financial losses worldwide every year, especially in urban and agricultural areas. With the intensification of climate change and the increase in the frequency of heavy rainfall, the need for intelligent systems for rapid identification and monitoring of floods is felt more than ever. In this regard, the combination of satellite images and deep learning algorithms has been proposed as a new and efficient tool in natural disaster management (Abdulaal et al., 2014, Feng et al., 2015, Yavari et al., 2020).

Satellite images initially had low spatial resolution and were limited to the visible and near-infrared spectral bands, but with the advancement of optical and radar technologies, the quality and accuracy of data increased dramatically. In recent decades, with the arrival of advanced satellites such as Sentinel-1 and Sentinel-2 belonging to the European Union's Copernicus program, remote sensing data with higher spatial, temporal and spectral resolution and free access have become available to researchers and disaster management organizations. Sentinel-1, using SAR (Synthetic Aperture Radar), has the ability to image the earth's surface in all weather conditions and at night, and is therefore very suitable for monitoring flooded areas. Sentinel-2 also provides useful information about land cover, soil moisture and flooding status with its multispectral images.

Satellite images have wide applications in flood detection. These images can accurately identify flood-affected areas by comparing the situation before and after heavy rainfall. Combining multi-temporal and multi-spectral data allows researchers to analyze flood spread, changes in watersheds, and vulnerability of different areas. In addition, radar images have the ability to penetrate cloud cover, which is a great advantage in critical situations. With the development of image processing algorithms and machine learning, satellite image analysis has also become much more advanced. These analyses can automatically detect flooded areas and play a key role in rapid decision-making for crisis management. In general, the evolution of satellite technology, along with advances in analytical algorithms, has provided humanity with a powerful tool for flood monitoring and management.

Previous studies on flood detection using satellite imagery and image analysis algorithms have shown that this field has made significant progress in recent decades. Many researchers have used optical satellite data such as Landsat, MODIS, Sentinel-2, as well as radar data such as Sentinel-1 and RADARSAT, to identify flood-affected areas with high accuracy using various methods.

Flood detection research has evolved from traditional thresholding and index-based techniques to more sophisticated machine learning and deep learning frameworks. Early flood detection relied on spectral indices such as NDWI and NDVI, as well as supervised classifiers like Maximum Likelihood Classification, but these methods were limited by cloud cover and atmospheric conditions. The introduction of Synthetic Aperture Radar (SAR) sensors, particularly Sentinel-1, revolutionized flood mapping by providing cloud- and illumination-independent data. Traditional machine learning models such as Support Vector Machines (SVM) and Random Forest (RF) improved classification accuracy through the inclusion of spectral, textural, and contextual features. More recently, convolutional neural networks (CNNs) such as U-Net, PSPNet, MANet, and DeepLab have achieved state-of-the-art results in flood detection, outperforming traditional models by automatically learning hierarchical spatial representations. Several studies have demonstrated the effectiveness of deep learning for flood segmentation using Sentinel-1 SAR data and highlighted the role of polarization (VV and VH) in influencing detection accuracy (Abbasi et al., 2023, Pech-May et al., 2023, Amitrano et al., 2024). Nonetheless, research challenges persist, including domain generalization, data imbalance, and the integration of multi-source observations. Recent reviews

* Corresponding author

emphasize the importance of robust, scalable architectures for cross-regional flood mapping (Ghosh et al., 2022, Toma et al., 2024)

In this regard, Andrei Tuel and colleagues (Twele et al., 2016) designed a series of automated processing systems based on thresholding and fuzzy logic for the real-time detection and monitoring of floods using Sentinel-1 images. The designed system achieved an accuracy ranging from 94% to 96.1% in the region from Greece to Turkey. The results of this research indicated that using VV polarization data instead of VH provided higher accuracy.

Marion Tangar et al in 2017 (Tanguy et al., 2017) combined high-resolution SAR images from the RADARSAT-2 satellite (with a resolution of 3 meters and HH polarization) with hydraulic data from a specific time period during the flood to create flood maps in urban and rural areas using fuzzy-based classification methods, achieving an accuracy of 87%.

In 2019, Li et al introduced a self-learning convolutional neural network (A-SLCNN) using TerraSAR-X satellite images, addressing the limitation of training data (Li et al., 2019). They demonstrated that using images from different time intervals played a crucial role in flood detection. In 2021, David Monz et al (Muñoz et al., 2021) examined the performance of convolutional neural networks and data fusion for generating flood maps using Landsat multispectral images with a spatial resolution of 30 meters, achieving an accuracy of 97%. In the same year, Lei et al (Lei et al., 2021) considered ten parameters affecting flooding as predictive variables, using convolutional and recurrent neural networks for flood mapping. Ultimately, the convolutional neural network model exhibited a better performance with an accuracy of 84% compared to other networks. One notable aspect of these studies is that they primarily focus on specific areas; therefore, the effectiveness of the methods used in each study cannot be definitively generalized to other regions. Additionally, some of these studies lacked sufficient satellite images for training neural networks, and the accuracy in several cases was relatively low. Given the importance of accurately detecting flood phenomena and segmenting affected areas, the precision of the employed methods should be as high as possible.

This research focuses on flood detection and assessment using satellite images, along with deep learning methods and algorithms aimed at swiftly identifying flood-affected areas for effective management of this natural disaster in vulnerable regions. This study employs a set of deep learning models that perform well in recognizing and classifying flood areas. The models demonstrate higher accuracy compared to existing research and has been developed using satellite images acquired from various parts of the world. Consequently, this allows the models to perform adequately on images from other regions that were not included in the training process. Moreover, this model can be utilized for identifying and segmenting flood phenomena in satellite images over different time periods and in the coming years.

## 2. Materials and Methods

### 2.1 Dataset

This study uses synthetic aperture radar (SAR) data from the Sentinel-1 satellite in VV and VH polarization. The data is extracted from the ETCI 2021 (EarthVision Challenge 2021: IEEE GRSS Data Fusion Contest on Flood Detection) dataset, which is being collected by NASA Impact, IEEE GRSS, and other partners. The objective of the competition is to develop accurate and reliable algorithms for detecting flood zones in different geographical regions using satellite imagery. The ETCI 2021 dataset consists of Sentinel-1 C-band GRD-level SAR images with VV polarization, presented as time pairs (pre-flood and post-flood). The images are collected extensively for different regions of India, covering a wide range of climatic and topographic variations. The spatial resolution of the data is 10 meters, and due to its radar nature, it is possible to image at night and in cloudy conditions, which is very crucial for flood detection.

In order to facilitate the training and evaluation process of the algorithms, a Ground Truth Mask is provided for each pair of images, which identifies flooded and non-flooded areas in binary form. These labels are prepared based on a combination of optical data, field reports, and expert analysis and have high accuracy. Due to the wide spatial coverage, high quality of the labels, and access to multi-temporal radar images, the use of this dataset provides a suitable platform for training deep learning models in automatic flood detection. The spatial distribution related to the acquisition of this dataset is shown in Figure (1).

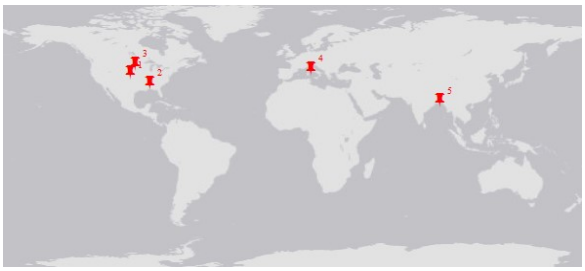


Figure 1. Location of the study areas

In this study, the training and validation sections of this dataset have been used to train and evaluate the models. This study leverages deep learning models and radar imagery in both VV and VH polarizations to accurately detect flooded areas and validate the results against ground-truth data.

The ETCI 2021 dataset includes a total of 66,810 image tiles, derived from multiple flood-prone regions such as Nebraska, North Alabama, Bangladesh, Red River North, and Florence. Each tile contains dual-polarization SAR bands (VV and VH) along with corresponding flood and water-body masks. Following the official ETCI 2021 data split, 44,062 tiles were used for training, 10,400 for validation, and 12,348 for testing. In this study, only the official training and validation subsets were used for model development and evaluation, while the test set was excluded since its ground-truth labels are not publicly available. To prepare the input data, the VV and VH channels were stacked together with their ratio to create a three-channel composite image, followed by per-channel normalization using statistics computed from the training set. Data augmentation techniques such as random rotation, horizontal flipping, and small-scale scaling were applied to improve generalization. This detailed dataset description ensures transparency and reproducibility of the experimental setup.

### 2.2 Methodology

In this study, two deep learning models, PSPNet and PAN, were employed to identify flood-prone areas from high-resolution satellite imagery. Although both models are built upon pyramid-based architectures for multi-scale context aggregation, their internal mechanisms differ. PSPNet relies solely on pyramid

pooling to integrate contextual cues without using explicit attention modules. In contrast, PAN enhances its segmentation capability by embedding attention mechanisms, which allow the model to focus more precisely on relevant features and edges. This makes PAN particularly effective in delineating fine details, especially along the boundaries of flood-affected areas, while PSPNet excels in capturing broader scene-level context.

In addition to these two models, we have also used the MANet model, which has high capabilities for semantic segmentation tasks. The main difference between this architecture and pyramid-based architectures like PSPNet and PAN lies in its approach to feature refinement. MANet adopts a multi-scale attention mechanism combined with mutual attention modules, enabling it to iteratively refine features at multiple levels. Unlike pyramid-based architectures that aggregate context in a more static and hierarchical manner, MANet dynamically recalibrates feature representations by modeling interdependencies across spatial and channel dimensions. This makes MANet especially effective at capturing both global structure and local details, improving its ability to accurately segment complex flood patterns in high-resolution imagery.

### 2.2.1 PAN

The PAN (Pyramid Attention Network) is a deep learning architecture utilized in computer vision tasks, particularly for semantic segmentation (Zhou et al., 2019, Mei et al., 2023). Its primary objective is to effectively extract and aggregate image features to facilitate accurate pixel-level classification. Unlike traditional methods that rely on intricate convolutional operations or complex decoder architectures, PAN employs attention mechanisms and a spatial pyramid structure to generate dense feature representations. PAN has demonstrated competitive performance in prominent semantic segmentation benchmarks such as PASCAL VOC 2012 and Cityscapes, outperforming several earlier approaches. The strength of the PAN architecture lies in its capacity to capture both global context and fine-grained local details, thereby enhancing the overall segmentation accuracy. By integrating features across multiple spatial scales, the model improves the semantic interpretation of images at both the macro and micro levels.

A key innovation in PAN is the Feature Pyramid Attention (FPA) module, which enhances feature representation by influencing high-level outputs and incorporating global contextual information. This module enables the model to selectively emphasize relevant spatial features, leading to improved performance in complex segmentation tasks. The attention structure of the FPA module is illustrated in Figure (2).

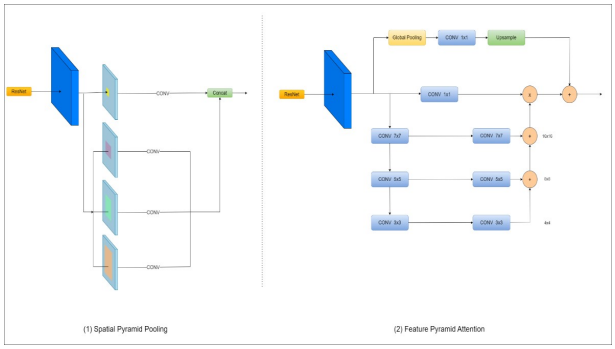


Figure 2. Attention module in PAN.

Another critical component of the PAN architecture is the Global Attention Upsample (GAU) module (Li et al., 2018), embedded in each decoder layer. This module refines the selection of spatial details by merging high-level semantic features with low-level visual cues. As depicted in the figure (3), the GAU module enables the network to model dependencies across different image regions, thereby boosting segmentation precision (Elhassan et al., 2021).

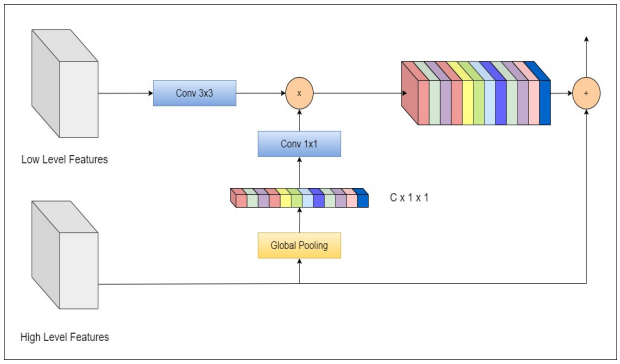


Figure 3. Global Attention Upsample module.

### 2.2.2 PSPNet

PSP-Net (Pyramid Scene Parsing Network) is a semantic segmentation architecture designed to enhance image segmentation tasks (Zhao et al., 2017, Liu et al., 2011), particularly within machine vision applications. To effectively capture contextual information from images, PSP-Net employs a pyramid pooling module that aggregates features from various spatial scales. This multi-scale aggregation improves the model's ability to generate reliable and detailed segmentation outputs. The architecture has been widely applied in diverse domains such as medical image segmentation, autonomous driving, and visual complication detection. Its strong performance in benchmark datasets like the Scene Parsing Challenge 2016 and PASCAL VOC 2012 has contributed to its adoption in numerous studies.

According to figure (4), PSP-Net begins by extracting high-level features from the input image using a backbone convolutional neural network, typically pre-trained on a large dataset. This feature extraction step reduces the spatial resolution to approximately one-eighth of the original image size. The resulting feature map is then processed by the pyramid pooling module, which operates at four different spatial levels. These levels capture information from the entire image, half-scale regions, and progressively smaller segments, allowing the model to understand both global and local context (Yang et al., 2020). The pooled features from all pyramid levels are upsampled and concatenated with the original feature map (He et al., 2015). This combined representation is then passed through a convolutional layer to produce the final semantic segmentation map. The architecture of PSP-Net is designed to support pixel-level classification with high accuracy.

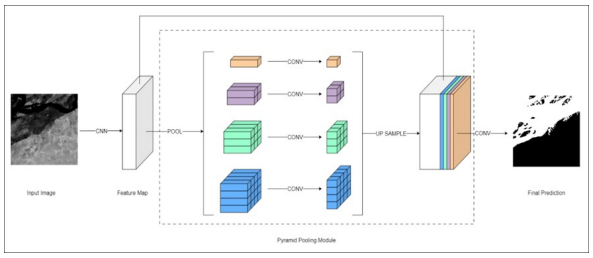


Figure 4. PSPNet architecture.

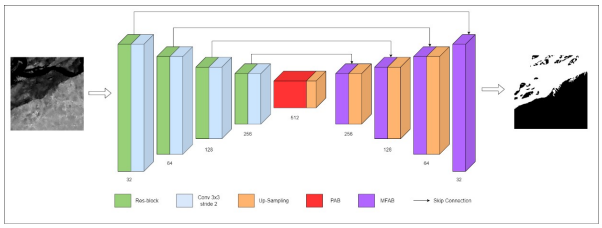


Figure 5. MANet architecture.

By leveraging its pyramid pooling strategy, PSP-Net effectively captures spatial hierarchies and improves segmentation accuracy across varying image content. Comparative evaluations have shown that it outperforms conventional semantic segmentation models such as FCN, U-Net, DeepLab, and DilatedNet, especially in complex scene parsing tasks.

### 2.2.3 MANet

MANet (Li et al., 2022, Fan et al., 2020) is a deep learning architecture developed for the automatic detection of liver tumors in medical imaging. Its primary objective is to accurately identify the location and extent of liver tumors, which is essential for supporting radiologists in clinical diagnosis.

Residual Block (Res-block):

The network incorporates three 3×3 convolutional layers combined with residual connections to effectively extract high-dimensional feature representations. Additionally, 1×1 convolution is employed to regulate the number of input feature channels. To address memory constraints and maintain performance, group normalization is utilized instead of traditional batch normalization.

Location-Based Attention Block:

This module is designed to capture spatial dependencies within the local feature maps. By focusing on the spatial relationships among features, the location-based attention block enhances the model's ability to extract relevant spatial information from the input data.

Multi-Scale Fusion Attention (MFA) Block:

Inspired by the human visual system, the attention mechanism in this block enables the model to prioritize relevant information across different feature maps. The MFA block facilitates the fusion of both high-level and low-level features, learning inter-band dependencies without increasing spatial dimensionality. It selectively enhances informative features while suppressing less relevant ones, thereby improving the model's representational capacity. The complete architecture of MANet, including these key modules, is illustrated in Figure (5).

High-level features capture the semantic content of an image, while low-level features transferred through skip connections retain fine-grained edge and texture information essential for reconstructing image details. In the MANet model, the Mean Squared Error (MSE) loss function is employed. This function penalizes larger prediction errors more heavily by squaring the difference between predicted and actual values. Consequently, MSE is particularly suitable when the dataset exhibits relatively small error margins, ensuring stable and precise optimization.

All models were trained using satellite images, with annotated flood extents serving as ground truth labels. The dataset was divided into training, validation, and testing subsets. To prevent overfitting and improve generalization, data augmentation techniques such as random rotation, scaling, and horizontal flipping were applied. Training was conducted using the MSE loss function. A learning rate scheduler dynamically adjusted the learning rate to facilitate convergence.

To enhance the performance of the proposed segmentation framework without introducing new network architectures, we focused on optimizing two critical methodological components: the encoder backbone and the learning-rate scheduling strategy. Several pre-trained encoders were evaluated, including ResNet-50, ResNet-101, ResNet-152, and EfficientNet-B7, all integrated within the same decoder structure to ensure fair comparison. This analysis aimed to determine which feature extractor provides the most robust representation of SAR imagery for flood mapping. EfficientNet-B7, owing to its compound scaling mechanism, was expected to capture multi-scale spatial features more effectively, leading to improved segmentation accuracy.

In addition, we employed the torch-lr-finder library to automatically estimate an optimal learning rate by examining the relationship between training loss and learning rate across a logarithmic scale. The optimal learning rate, corresponding to the region of steepest loss decline, was identified as $1.35 \times 10^{-4}$. This value was adopted for subsequent training sessions to promote faster convergence and improved model stability. Together, these methodological refinements ensure a more reliable and efficient training process while maintaining architectural consistency across experiments.

All models in this study were implemented using the PyTorch 1.13 framework and trained on an NVIDIA RTX 3090 GPU with 24 GB of memory. The input SAR tiles were resized to 256 × 256 pixels and normalized before being fed into the networks. A batch size of 16 was used for all experiments to balance memory efficiency and convergence stability. The Adam optimizer was employed with the initial learning rate obtained with the torch-lr-finder tool. Each model was trained for a maximum of 500 epochs with early stopping applied if the validation loss failed to improve for 20 epochs. These consistent training configurations ensure a fair comparison among models and enhance the reproducibility of the experimental results.

## 3. Results and Discussion

In this section, we present and compare the results of the three deep learning models. The outcomes of PAN model are illustrated in Figure (6). These visual results were derived from the dataset acquired over Alabama, using VV and VH polarizations. As demonstrated in Figure (6), the model achieved higher accuracy with the VH polarization dataset, indicating superior performance in flood prediction under this configuration.

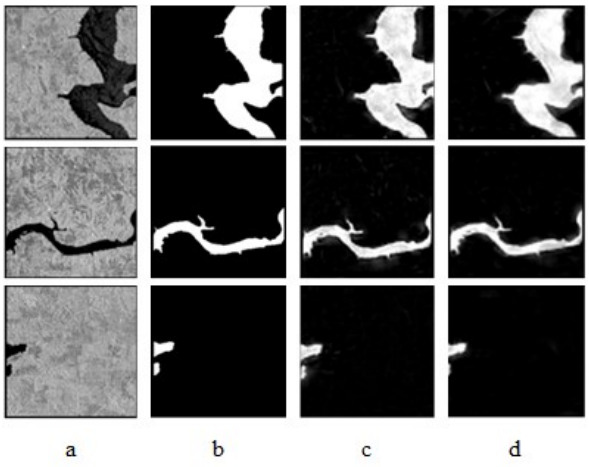


Figure 6. Results from the PAN model, (a) Radar images, (b) Reality map, (c) VH output and (d) VV output

This paper identifies several evaluation metrics resulting from the approach used, which include: Precision, Recall (Buckland and Gey, 1994), IoU (Rosebrock, 2016), F1 Score(Chicco and Jurman, 2020), and Accuracy (Maxwell et al., 2021). Table 1 presents the evaluation metrics of the model for both VV and VH polarizations. As evident from the results, the values of the evaluation criteria are consistently higher for the VH polarization compared to VV, indicating that the model performs more effectively on the dataset with VH polarization.

| Metrics | VV | VH |
|---|---|---|
| Accuracy | 99.42 | 99.49 |
| Precision | 90.42 | 88.84 |
| IoU | 67.58 | 70.24 |
| F1-Score | 77.55 | 80.57 |
| Recall | 72.61 | 76.63 |

Table 1. Values of evaluation criteria in the PAN model

The overall accuracy metric measures the ratio of correctly classified pixels to the total number of pixels. While this provides a general indication of performance, it is highly sensitive to class imbalance. In datasets such as ETCI 2021, where the number of non-flooded pixels greatly exceeds that of flooded ones, the accuracy metric tends to be dominated by the majority (non-flooded) class. Consequently, high accuracy values may not fully reflect segmentation performance for the minority (flooded) class.

The performance of the PSPNet model is illustrated in Figure (7). The results were generated using radar datasets with two polarization modes: VV and VH. A comparative visual assessment indicates that the model achieved superior performance on the VH polarization dataset compared to the VV polarization dataset.

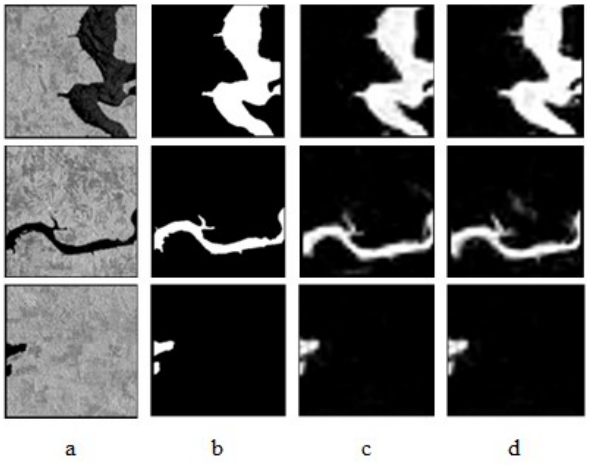


Figure 7. Results from the PSPNet model, (a) Radar images, (b) Reality map, (c) VH output and (d) VV output

Table (2) presents the model evaluation metrics for both VV and VH polarization cases. Analysis of these values enables a comparative assessment of the model's performance under each polarization. The results indicate that the PSPNet model delivers more accurate outcomes when applied to the VH polarization dataset.

| Metrics | VV | VH |
|---|---|---|
| Accuracy | 99.37 | 99.40 |
| Precision | 84.93 | 86.02 |
| IoU | 69.53 | 70.46 |
| F1-Score | 79.55 | 81.50 |
| Recall | 78.05 | 77.94 |

Table 2. Values of evaluation criteria in the PSPNet model

As previously described, this model is a neural network architecture that incorporates external memory modules to improve learning and reasoning capabilities. It demonstrates strong performance across a range of tasks, particularly those involving complex reasoning, long-term dependencies, and sequential data patterns. Similar to other deep learning approaches, this network employs the EfficientNet-B7 encoder, which was selected through empirical evaluation for its compatibility and effectiveness. The outcomes of the MANet model are presented in Figure (8). Visually, the results indicate higher accuracy in the VH polarization compared to the VV polarization. Overall, the model performs well in delineating flood-affected areas.

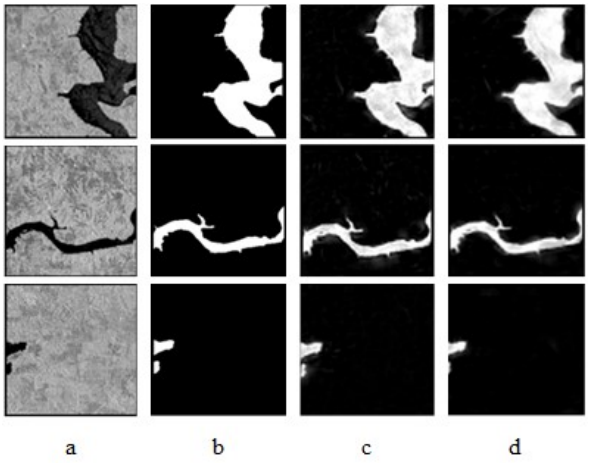


Figure 8. Results from the MANet model, (a) Radar images, (b) Reality map, (c) VH output and (d) VV output

The evaluation metrics are summarized in the table below. As evident from the results, the metrics exhibit higher values for VH polarization compared to VV polarization, indicating that the MANet model performs more effectively on the radar dataset with VH polarization.

| Metrics | VV | VH |
|---|---|---|
| Accuracy | 99.63 | 99.67 |
| Precision | 91.78 | 92.96 |
| IoU | 82.36 | 84.21 |
| F1-Score | 89.56 | 91.11 |
| Recall | 88.21 | 89.53 |

Table 3. Values of evaluation criteria in the MANet model

This study demonstrates the effectiveness of deep learning models in leveraging Sentinel-1 SAR data for accurate flood detection, revealing notable differences in performance based on polarization modes and model architectures. The VH polarization consistently outperformed VV across key metrics such as Intersection over Union (IoU), F1-score, and Recall, suggesting its greater sensitivity to surface water changes. This enhanced sensitivity is likely due to VH's stronger interaction with vegetated and rough surfaces, which are prevalent in flood-affected areas. This pattern was evident across all three models examined: PAN, PSPNet, and MANet.

Among the tested models, MANet achieved the best overall results, with an IoU of 84.21% and an F1-score of 91.11% when using VH polarization. These strong results are attributable to MANet's multi-scale attention and mutual attention mechanisms, which allow for more accurate representation of complex spatial patterns and boundaries typical of flooded regions. Its capacity to integrate both detailed local features and broader contextual information makes it particularly well-suited for flood delineation. While PAN and PSPNet also performed well, they were slightly less effective than MANet. The PAN architecture, which includes Feature Pyramid Attention (FPA) and Global Attention Upsample (GAU) modules, successfully captured multi-scale semantic and localized features, resulting in high-quality segmentation. PSPNet, although lacking explicit attention modules, remained competitive due to its robust pyramid pooling structure, which enhances the model's contextual awareness. All three models achieved high overall accuracy (above 99%), supporting their potential for real-time flood monitoring applications. The overall accuracy values are relatively high due to the low target-to-non-target ratio in several test tiles, where flooded pixels represent a small fraction of the total area. This imbalance leads to inflated accuracy values dominated by correctly identified non-flooded regions. Therefore, we complement overall accuracy with Precision, Recall, and F1-score metrics to provide a more balanced and class-sensitive evaluation. However, variations in precision, particularly in the PAN model, indicate a trade-off between high sensitivity (recall) and the occurrence of false positives, especially when using VV polarization. This underscores the need for careful selection of polarization and tailored model tuning for specific deployment scenarios. The findings affirm the value of integrating SAR data with advanced deep learning architectures for flood detection. The clear advantage of VH polarization and the strong performance of MANet point toward promising directions for further development. Future research could investigate ensemble techniques or the fusion of SAR with optical imagery to enhance performance under diverse environmental conditions.

Figure (9) illustrates that encoder architecture plays a crucial role in determining the convergence dynamics and segmentation performance. Among the evaluated models, the ResNet-based encoders demonstrated progressively improved stability with increasing depth, while EfficientNet-B7 consistently achieved the lowest and most stable loss curve across all epochs. This indicates that EfficientNet's compound scaling strategy provides richer multi-scale feature representations, which are particularly effective for delineating flood boundaries in heterogeneous SAR imagery. Quantitatively, the EfficientNet-B7 encoder reduced the final loss by approximately 40% compared to ResNet-50, highlighting its superior representational power.

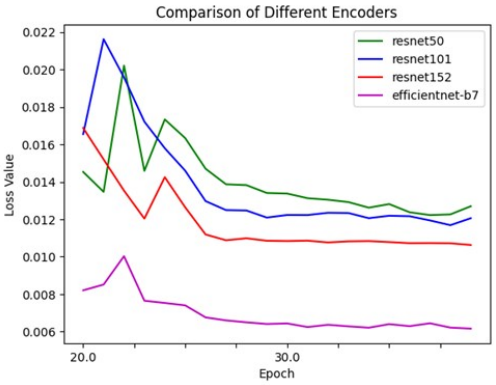


Figure 9. Comparison of loss curves for different encoder backbones

The learning-rate optimization experiment using the torch-lr-finder tool (Figure (10)) further refined the training process. The automatically suggested optimal learning rate ($1.35 \times 10^{-4}$) corresponded to the point of steepest loss descent, providing a balance between rapid convergence and stable generalization. Adopting this learning rate led to faster convergence in early epochs and improved final accuracy across all models.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-W8-2025-501-2026 | © Author(s) 2026. CC BY 4.0 License.

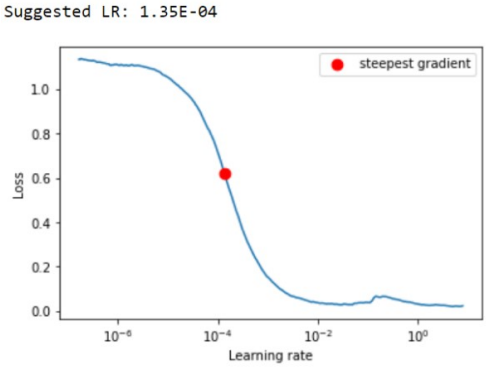


Figure 10. Learning-rate-finder analysis using the *torch-lr-finder* tool

Together, the encoder and learning-rate optimization analyses demonstrate that careful architectural and hyperparameter selection can yield substantial performance gains, representing a methodological refinement that enhances the robustness and efficiency of the segmentation framework without altering its core structure.

## 4. Conclusion

This study demonstrates that combining high-resolution Sentinel-1 SAR imagery with advanced deep learning models offers an effective approach for accurate flood detection. Through rigorous comparison, it was found that VH polarization consistently provided better performance than VV polarization across all tested architectures. Among PAN, PSPNet, and MANet, the latter achieved the highest accuracy and segmentation quality, thanks to its robust attention-based architecture capable of capturing complex spatial patterns. The results highlight MANet's suitability for operational flood mapping, particularly under challenging environmental conditions. These findings pave the way for more responsive disaster management systems and suggest future work in model fusion or SAR-optical data integration to further enhance flood monitoring capabilities. Furthermore, the study underscores the critical role of advanced deep learning architectures in interpreting complex SAR data for natural disaster applications. As climate change continues to amplify the frequency and intensity of extreme weather events, the ability to detect floods rapidly and accurately becomes increasingly vital. The demonstrated superiority of VH polarization emphasizes the importance of selecting appropriate radar configurations in operational systems. Future extensions of this work could involve deploying these models in real-time flood monitoring platforms, integrating auxiliary data such as topography or rainfall, and conducting cross-regional generalization studies. Ultimately, the insights gained from this research contribute to building more resilient societies through improved early warning systems and informed disaster risk reduction strategies.

## References

ABBASI, M., SHAH-HOSSEINI, R., AGHDAMI-NIA, M. Sentinel-1 polarization comparison for flood segmentation using deep learning. Proceedings, 2023. MDPI, 14.

ABDULAAL, M., ALGARNI, M., SHAMIM, A., CLAUDEL, C. Unmanned Aerial Vehicle based flash flood monitoring using Lagrangian trackers. International Workshop on Robotic Sensor Networks. Cyber-physical systems, 2014.

AMITRANO, D., DI MARTINO, G., DI SIMONE, A., IMPERATORE, P. J. R. S. 2024: Flood detection with SAR: A review of techniques and datasets. 16**,** 656.

BUCKLAND, M., GEY, F. 1994: The relationship between recall and precision. *Journal of the American society for information science,* 45**,** 12-19.

CHICCO, D., JURMAN, G. 2020: The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics,* 21**,** 6.

ELHASSAN, M. A., CHEN, Y., CHEN, Y., HUANG, C., YANG, J., YAO, X., YANG, C., CHENG, Y. 2021: PPANet: Point-Wise Pyramid Attention Network for Semantic Segmentation. *Wireless Communications and Mobile Computing,* 2021**,** 5563875.

FAN, T., WANG, G., LI, Y., WANG, H. 2020: Ma-net: A multi-scale attention network for liver and tumor segmentation. *Ieee Access,* 8**,** 179656-179665.

FENG, Q., LIU, J., GONG, J. 2015: Urban flood mapping based on unmanned aerial vehicle remote sensing and random forest classifier—A case of Yuyao, China. *Water,* 7**,** 1437-1455.

\GHOSH, B., GARG, S., MOTAGH, M. J. I. A. O. T. P., REMOTE SENSING & SCIENCES, S. I. 2022: Automatic flood detection from sentinel-1 data using deep learning architectures. 3**,** 201-208.

HE, K., ZHANG, X., REN, S., SUN, J. 2015: Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence,* 37**,** 1904-1916.

LEI, X., CHEN, W., PANAHI, M., FALAH, F., RAHMATI, O., UUEMAA, E., KALANTARI, Z., FERREIRA, C. S. S., REZAIE, F., TIEFENBACHER, J. P. 2021: Urban flood modeling using deep-learning approaches in Seoul, South Korea. *Journal of Hydrology,* 601**,** 126684.

LI, H., XIONG, P., AN, J., WANG, L. 2018: Pyramid attention network for semantic segmentation. *arXiv preprint arXiv:1805.10180*.

LI, M., ZHONG, B., MA, K.-K. MA-NET: Multi-scale attention-aware network for optical flow estimation. ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022. IEEE, 2844-2848.

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 Journal of Photogrammetry and Remote Sensing,* 152**,** 178-191.

LIU, C., YUEN, J., TORRALBA, A. 2011: Nonparametric scene parsing via label transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 33**,** 2368-2382.

MAXWELL, A. E., WARNER, T. A., GUILLÉN, L. A. 2021: Accuracy assessment in convolutional neural network-based deep learning remote sensing studies—Part 1: Literature review. *Remote Sensing,* 13**,** 2450.

MEI, Y., FAN, Y., ZHANG, Y., YU, J., ZHOU, Y., LIU, D., FU, Y., HUANG, T. S., SHI, H. 2023: Pyramid attention network for image restoration. *International Journal of Computer Vision,* 131**,** 3207-3225.

MUÑOZ, D. F., MUÑOZ, P., MOFTAKHARI, H., MORADKHANI, H. 2021: From local to regional compound flood mapping with deep learning and data fusion techniques. *Science of the Total Environment,* 782**,** 146927.

PECH-MAY, F., AQUINO-SANTOS, R., DELGADILLO-PARTIDA, J. J. R. S. 2023: Sentinel-1 SAR images and deep learning for water body mapping. 15**,** 3009.

ROSEBROCK, A. 2016: Intersection over Union (IoU) for object detection. *P y I mage S earch. com, https://www. pyimagesearch. com/2016/11/07/intersectio n-over-union-iou-for-object-detection/(accessed May 18, 2021)*.

TANGUY, M., CHOKMANI, K., BERNIER, M., POULIN, J., RAYMOND, S. 2017: River flood mapping in urban areas combining Radarsat-2 data and flood return period data. *Remote Sensing of Environment,* 198**,** 442-459.

TOMA, A., ȘANDRIC, I., MIHAI, B.-A. J. E. J. O. R. S. 2024: Flooded area detection and mapping from Sentinel-1 imagery. Complementary approaches and comparative performance evaluation. 57**,** 2414004.

TWELE, A., CAO, W., PLANK, S., MARTINIS, S. 2016: Sentinel-1-based flood mapping: a fully automated processing chain. *International Journal of Remote Sensing,* 37**,** 2990-3004.

YANG, Z., SU, Y., GARG, H., XIE, X., WU, S. 2020: Data-driven fuzzy multiple criteria decision making and its potential applications. *Mathematical Problems in Engineering*.

YAVARI, A., HOMAYOUNI, S., OUBENNACEUR, K., CHOKMANI, K. 2020: Flood inundation modeling in ungauged basins using Unmanned Aerial Vehicles imagery. *Earth Observation and Geomatics Engineering,* 4**,** 44-55.

ZHAO, H., SHI, J., QI, X., WANG, X., JIA, J. Pyramid scene parsing network. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017. 2881-2890.

ZHOU, Z., SIDDIQUEE, M. M. R., TAJBAKHSH, N., LIANG, J. 2019: Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging,* 39**,** 1856-1867.