University of Tehran

Earth Observation and Geomatics Engineering

Online ISSN: 2588-4360

Homepage: https://eoge.ut.ac.ir/

Comparing CNN and GAN Based Models for Flood Extent Mapping

(Selected Paper in the 8th ISPRS Geospatial Conference 2025, University of Tehran, Iran)

Ehsan Haghighi Gashti [1], and **Mahmoud Reza Delavar** [2✉]

1. GIS Dept., School of Surveying and Geospatial Eng., College of Engineering, University of Tehran, Tehran, Iran. E-mail: ehsanhaghighi77@ut.ac.ir
2. Corresponding author, Centre of Excellence in Geomatic Eng. in Disaster Management, and Land Administration in Smart City Lab., School of Surveying and Geospatial Eng., College of Engineering, University of Tehran, Tehran, Iran. E-mail: mdelavar@ut.ac.ir

Article Info

Article type:
Research Article

Article history:
Received 2026-01-19
Received in revised 2026-02-02
Accepted 2026-02-03
Available 2026-05-12

Keywords:
Flood Extent Mapping,
CNN,
GAN,
Sentinel-1,
Sentinel-2.

ABSTRACT

Floods are among the most destructive natural hazards worldwide, demanding accurate and timely flood extent mapping for effective disaster management. With the increasing availability of satellite observations, deep learning methods have become vital approaches to improve flood detection. This study systematically compares convolutional neural network (CNN) architectures (U-Net, DeepLabV3 and HRNet) with a conditional generative adversarial network (GAN) (Pix2Pix) for flood extent mapping using the Sen1Floods11 dataset, which provides paired Sentinel-1 synthetic aperture radar (SAR) and Sentinel-2 optical imagery with annotated flood masks. Models were trained separately on SAR and optical data to assess modality-specific performance and evaluated using Overall accuracy, Intersection over Union (IoU), Precision, Recall and F1-score show that SAR-based models consistently outperform optical-based counterparts, highlighting the robustness of radar data under adverse weather conditions. Among CNNs, DeepLabV3 achieved the highest IoU (84.56%) and F1-score (92.35%), offering precise delineation of flooded areas. In contrast, Pix2Pix achieved the highest overall accuracy (97.45%) and Recall (93.79%), capturing broader flood extents but with a tendency to overestimate flood extent. These findings indicate that CNNs are well-suited for high-precision applications such as risk assessment, while GAN-based models may be advantageous for rapid emergency response where minimizing missed detections is critical. The study underscores the complementary strengths of CNNs and GANs and points toward multimodal fusion and real-time deployment as promising future directions for operational flood monitoring.

Cite this article: Haghighi Gashti, & H., Delavar, M.R. (2025). Comparing CNN and GAN Based Models for Flood Extent Mapping, *Earth Observation and Geomatics Engineering*, Volume 9, Issue 1, Pages 126-136. http//doi.org/ 10.22059/eoge.2026.409803.1208

Publisher: University of Tehran.

DOI: http//doi.org/ 10.22059/eoge.2026.409803.1208

1. Introduction

Floods are among the most frequent and devastating natural hazards, responsible for widespread damage to infrastructure, economic losses, displacement of populations, and loss of life each year (Kundzewicz et al., 2014). The impacts of floods are intensifying due to climate change, rapid urbanization and environmental degradation, which together increase both the frequency and severity of flood events worldwide (Tedla et al., 2025). Therefore, effective flood monitoring and mapping are essential to support disaster preparedness, emergency response, recovery planning and long-term risk management. Accurate flood extent information can guide emergency services, inform evacuation strategies and facilitate the design of resilient infrastructure in flood-prone regions (Vamvakeridou-Lyroudia et al., 2020). Satellite-based remote sensing has become a cornerstone of flood monitoring because of its ability to provide timely, large-scale and repeated observations across diverse geographic regions (Schumann et al., 2023). Among available satellite missions, Sentinel-1 synthetic aperture radar (SAR) and Sentinel-2 multispectral optical imagery are particularly valuable for flood mapping. Sentinel-1 SAR offers the advantage of day-and-night, all-weather imaging, enabling the detection of flooded areas even under cloud cover and heavy rainfall conditions that frequently accompany floods (Misra et al., 2025). Sentinel-2, on the other hand, provides rich spectral and spatial information that enhances the discrimination of land cover classes and water bodies. Together, these complementary data sources offer an opportunity for more robust and accurate flood detection (Kurapati et al., 2020).

Traditional methods for flood extent mapping relied on thresholding (Long et al., 2014), change detection (Li et al., 2018) or spectral indices derived from SAR and optical data. While effective in controlled settings, these approaches often struggle in heterogeneous environments, under mixed land cover conditions, or when noise and speckle are present in SAR imagery (Girisha et al., 2025). Over the past decade, deep learning (DL) techniques have revolutionized remote sensing analysis, offering state-of-the-art performance for semantic segmentation (Haghighi Gashti et al., 2025) tasks. Convolutional neural networks (CNNs) in particular have been widely adopted for flood mapping due to their ability to capture spatial hierarchies and learn discriminative features directly from raw imagery. Architectures such as U-Net (Ronneberger et al., 2015), DeepLabV3 (Chen et al., 2017) and HRNet (Jingdong Wang et al., 2020) have become benchmarks for semantic segmentation (Haghighi Gashti et al., 2024), demonstrating robustness and scalability in extracting flooded areas from multisensory data. Despite their strong performance, CNN-based approaches have some limitations. They typically require large amounts of annotated training data, can be sensitive to domain shifts across regions or sensors and may struggle to capture complex contextual relationships or generate realistic fine-grained structures in the output maps (Bentivoglio et al., 2022). In parallel, Generative Adversarial Networks (GANs) have emerged as powerful models for image-to-image translation (Jinyu Wang et al., 2024) and cross-domain adaptation. Conditional GANs (cGANs) such as Pix2Pix (Isola et al., 2017) have been applied in segmentation tasks by treating label prediction as a translation problem from input images to target masks. Building on this, Pix2PixHD (T.-C. Wang et al., 2018) extends the original Pix2Pix framework by incorporating a coarse-to-fine generator and multi-scale discriminators, enabling the synthesis of high-resolution realistic outputs.

Despite the rapid progress of deep learning-based flood mapping, the comparative role of different model families remains insufficiently understood. Most existing studies focus on optimizing or proposing individual CNN architectures for flood segmentation, while the potential of conditional generative adversarial networks (cGANs) has primarily been explored in the context of domain adaptation or data augmentation rather than direct operational flood extent mapping. To the best of our knowledge, a systematic and fair comparison between representative CNN-based semantic segmentation models and cGAN-based approaches for flood mapping, using a large-scale standardized dataset and consistent evaluation protocol, is still lacking. This gap limits our understanding of whether GAN-based models can offer complementary advantages over conventional CNNs, particularly in terms of spatial coherence, boundary delineation and generalization across diverse flood scenarios.

In this study, we address this gap by conducting a comprehensive comparison between convolutional neural network (CNN) and conditional generative adversarial network (cGAN) architectures for flood extent mapping using the Sen1Floods11 dataset. Specifically, we benchmark three widely used CNN models—U-Net, DeepLabV3 and HRNet—against the Pix2Pix cGAN framework under identical training and evaluation settings. The main contributions of this work are threefold: (1) providing a systematic assessment of CNN- and GAN-based models for flood mapping on a large, multi-regional dataset; (2) analyzing quantitative performance differences as well as qualitative characteristics of the generated flood extent maps; and (3) discussing the strengths and limitations of each model family in the context of operational flood monitoring. The findings of this study offer practical insights into model selection for flood mapping applications based on accuracy, robustness and output characteristics.

Section 2 reports on some major previous researches that have been undertaken regarding the use of CNN and GAN model for flood mapping. In section 3, the dataset that has been used in this research is explained. Section 4 explains the research methodology. Section 5 presents the results and section 6 concludes the paper and suggests some future research directions.

2. Related Work

Sarker et al. (2019) proposed a fully convolutional neural network (F-CNN) model for mapping flood extent using Landsat satellite imagery (Sarker et al., 2019). A total of 64 models were developed and trained, varying the neighborhood size of training samples and the number of convolutional filters. To assess the generalizability of the model, it was evaluated on a separate set of Landsat images covering flood-affected regions across Australia. Comparative analysis with a conventional support vector machine (SVM) classifier demonstrated that the F-CNN model more accurately identified flooded areas. Notably, the F-CNN achieved a maximum precision of 76.7%, significantly outperforming the 45.27% achieved by the SVM model.

Seydi et al. (2022) introduced a framework for flood area mapping using diverse remote sensing datasets (Seydi et al., 2022). The framework integrated synthetic aperture radar (SAR), optical imagery and altimetry data in order to train a CNN model. To evaluate the model effectiveness, the method was applied to large-scale datasets from Golestan and Khuzestan Provinces in the north and south-west of Iran. The findings demonstrated that the framework achieved high accuracy in identifying flooded regions. Both visual inspections and numerical results confirm its capability, with overall accuracy exceeding 98% across both study sites.

In another study, Tavus et al. (2022) implemented a U-Net architecture for multi-class segmentation of flooded regions and flooded vegetation, using Sentinel-1 SAR imagery combined with elevation data (Tavus et al., 2022). Training samples were generated through an automatic thresholding method (OTSU) in Sardoba, Uzbekistan and Sagaing, Myanmar and the model was validated in Ordu, Turkey and the Ca River, Vietnam, by visual comparison with existing flood maps. The results demonstrate that CNNs are highly effective in distinguishing flooded areas and flooded vegetation, even when trained on geographically distinct regions. The model achieved F1-scores of 91% for flood areas and 85% for flooded vegetation.

Sirsant et al. (2025) introduced a multi-stage framework that integrated topographic feature extraction, CNN-based flood classification and regression modelling to enhance flood mapping accuracy (Sirsant et al., 2025). The methodology was evaluated using manually collected flood data from Selva Nagar, Chennai, India, during the December 2023 flood. The results demonstrated strong performance, achieving a R^2 of 0.93 and a RMSE of 25.01%. Comparative analysis with established CNN architectures, including ResNet50 and InceptionV3, confirmed the superiority of the proposed model.

In case of using GAN models for flood extent mapping, Sekine et al. (2024) improved the precision of SAR-based flood mapping by integrating optical information (Sekine et al., 2024). They employed a publicly available SAR–optical paired dataset from flood-affected regions to train GAN model that convert SAR images into an optical-like style. After that, they developed a flood segmentation CNN model using both the original SAR data and the translated images. The results demonstrated that GANs produced high-quality translations and that segmentation model leveraging both SAR and optical data outperformed those trained on single-source inputs.

Dubey and Katarya (2024) proposed a super-resolution approach for SAR imagery using a Satellite Super Resolution-based Generative Adversarial Network (SSR-GAN). To evaluate performance, metrics such as Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Multiscale Structural Similarity (MSSIM) and Mean Squared Error (MSE) were used. The results demonstrated that SSR-GAN outperformed existing methods, achieving higher PSNR, SSIM and MSSIM scores while producing lower MSE values.

3. Dataset

We used the Sen1Floods11 dataset (Figure 1), a publicly available benchmark for flood detection and segmentation. It includes paired Sentinel-1 SAR and Sentinel-2 optical imagery from flood events between 2015 and 2019 across 11 countries (Bolivia, Colombia, Ghana, India, Myanmar, Nigeria, Pakistan, Paraguay, Somalia, Spain and USA). Ground truth labels for flooded and non-flooded areas were derived from high-resolution optical imagery and validated by experts. The dataset multi-modal design enables the use of both SAR and optical data, facilitating more robust model development. To ensure geographic independence, we split the dataset at the country level: seven countries (Bolivia, India, Myanmar, Nigeria, Paraguay, Spain and USA) for training, three (Colombia, Ghana and Pakistan) for validation, and one (Somalia) reserved for testing. This strategy allows rigorous evaluation of model generalization to unseen flood scenarios. The geographic distribution of the dataset across the training, validation, and test countries is illustrated in Figure 1.

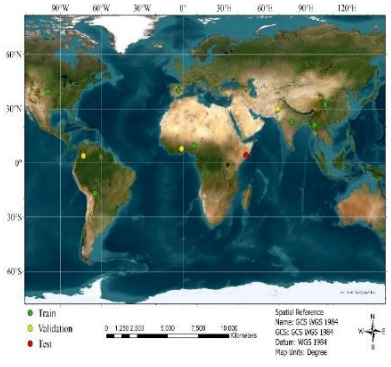


Figure 1. Training, validation and test data distribution for the study areas

Each image patch has a spatial resolution of 512 × 512 pixels. The dataset initially had 4,106 patches for training,

483 for validation, and 242 for testing. To increase the diversity of samples and reduce overfitting, several data augmentation techniques were applied including random horizontal and vertical flips, rotations, scaling and brightness adjustments. Table 1 summarizes the dataset split and augmentation results.

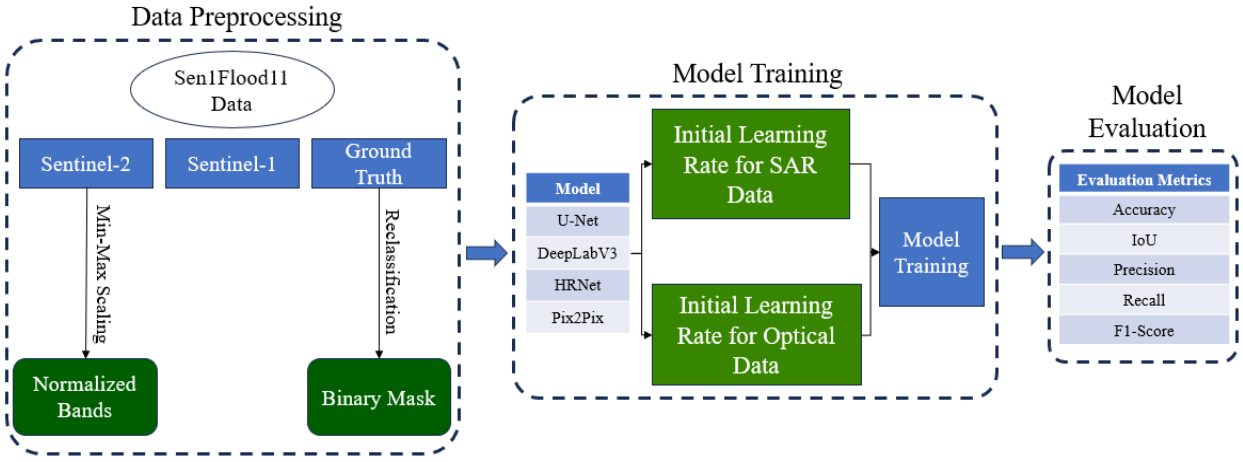


Figure 2. Conceptual model of the study

Table 1. Dataset split and augmentation summary

Dataset split	Original number of patches	Dataset split
Training	4106	16424
Validation	483	1932
Test	242	968

4. Methodology

The conceptual model for this study is shown in Figure 2. In this study, we implemented a comparative framework to evaluate CNNs and cGANs for flood extent mapping. The methodology comprises five components: (1) data preprocessing, (2) model architectures, (3) calculating initial learning rate from learning curve (4) training configuration, and (5) evaluation metrics.

4.1. Data preprocessing

All Sentinel-1 SAR and Sentinel-2 patches from Sen1Floods11 were pre-processed independently per modality. SAR inputs consist of two channels (VV and VH polarizations) which were treated as two-channel images. Sentinel-2 inputs use the 12 spectral channels provided in the dataset. Each channel was normalized independently using min–max scaling to the range [0,1]. Ground-truth flood masks in Sen1Floods11 contain three values including flooded areas (1), non-flooded areas (0) and permanent water bodies (–1). For consistency with binary segmentation tasks, permanent water pixels were reclassified from –1 to 0, effectively treating them as non-flooded. In Figure 3 some examples of the training data are shown.

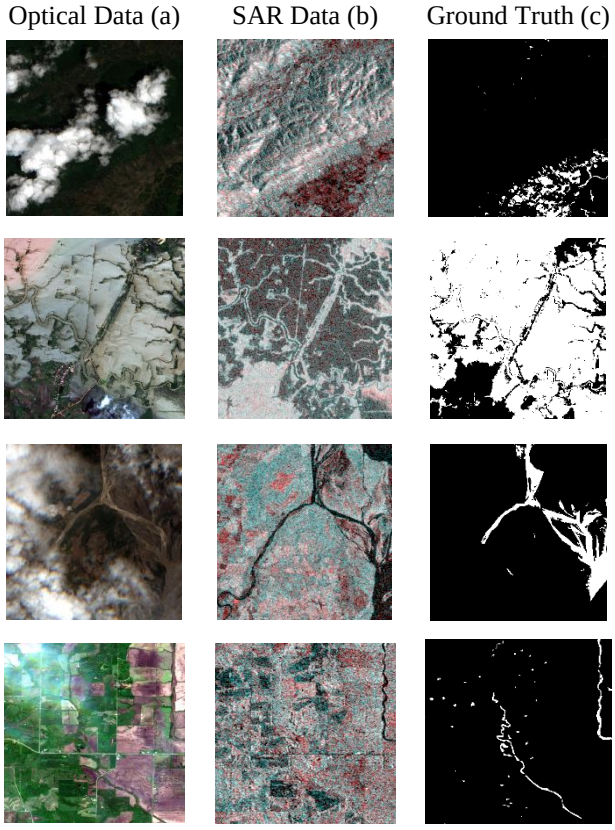


Figure 3. Optical (a) and SAR (b) data and their corresponding flood mask (c)

4.2. Model architectures and training specifications

To ensure a meaningful and representative comparison, the selected models were chosen to reflect different architectural paradigms commonly used in deep learning-based semantic segmentation. Specifically, U-Net represents encoder–decoder architectures with skip connections optimized for precise pixel-level prediction,

DeepLabV3 exemplifies context-aware models that leverage dilated convolutions for large receptive fields, and HRNet maintains high-resolution representations to preserve fine spatial details throughout the network. In contrast, Pix2Pix represents a conditional generative adversarial network (cGAN) framework that formulates flood mapping as an image-to-image translation problem, introducing adversarial learning to encourage spatial realism and sharper boundaries. Together, these models provide a comprehensive basis for evaluating how different design principles influence flood extent mapping performance.

- U-Net is an encoder–decoder architecture with symmetric skip connections between corresponding layers, originally designed for biomedical image segmentation and widely adopted in remote sensing due to its effectiveness in learning from limited labeled data. The skip connections allow low-level spatial details from the encoder to be fused with high-level semantic features in the decoder, enabling precise pixel-level delineation of flooded areas, particularly along complex boundaries such as riverbanks or narrow channels. In this study, two U-Net models were trained separately: one with Sentinel-1 SAR inputs (2 channels: VV, VH) and another with Sentinel-2 optical inputs (12 spectral bands).
- DeepLabV3 is a context-aware semantic segmentation architecture that leverages dilated convolutions to enlarge the receptive field without reducing spatial resolution. This design enables the model to capture long-range contextual information, which is crucial for flood mapping in heterogeneous landscapes where flooded areas may exhibit spectral similarities with surrounding surfaces such as wet soil or shadowed terrain. As with U-Net, separate DeepLabV3 models were trained for SAR and optical modalities.
- High-Resolution Network (HRNet) differs from conventional encoder–decoder architectures by maintaining high-resolution feature maps throughout the network while exchanging information across parallel multi-resolution branches. This property is particularly advantageous for flood mapping, as it improves boundary accuracy and reduces spatial artifacts when detecting thin flood patterns and sharp land–water transitions.
- Pix2Pix is a conditional generative adversarial network (cGAN) that reformulates semantic segmentation as an image-to-image translation task. By combining an adversarial loss with a pixel-wise reconstruction loss, Pix2Pix encourages the generation of spatially coherent and visually realistic flood extent maps with sharper boundaries. This makes it a suitable representative of GAN-based approaches for evaluating whether adversarial learning can complement or surpass conventional CNN-based segmentation in flood mapping applications.

4.3. Calculating initial learning rate from learning curve

The learning rate finder is a method for selecting an effective learning rate (Lr) before training a deep learning (DL) model. It works by training on a small data subset while gradually increasing the learning rate, usually on a logarithmic scale and recording the loss. Plotting loss against learning rate produces a curve with three phases: slow decrease at low rates, sharp decline at the optimum rates and divergence at high rates (Smith, 2017). The optimum learning rate is selected from the steepest loss reduction region, just before the loss rises. This approach speeds up convergence, avoids instability and improves training reliability. We applied the process introduced by (Smith, 2017) to determine our learning rate. The process starts with a very small learning rate Lr_{Start} (here 10^{-7}) and a maximum Lr_{End} (here 10^{-1}). The learning rate is increased exponentially over N iterations using Equation 1 (Smith, 2017) to ensure a smooth logarithmic increase:

$$Lr_i = Lr_{Start} \times \left(\frac{Lr_{End}}{Lr_{Start}}\right)^{\frac{i}{N}} \quad (1)$$

At each iteration, a forward and backward pass is performed with a mini-batch, model parameters are updated with Lr_i and the loss L_i is recorded. To reduce noise from mini-batch training, the loss can be smoothed with an exponential moving average using Equation 2 (Smith, 2017):

$$L_i^{Smooth} = \beta \times L_{i-1}^{Smooth} + (1-\beta) \times L_i \quad (2)$$

where, β is close to 1. To prevent divergence, training stops if the loss exceeds four times the minimum observed loss. Finally, the learning rate and loss values are plotted (log scale) and the optimum learning rate is selected from the region of the steepest decline. In Figure 4 the learning curve and the corresponding "best initial learning rate" for each model are illustrated. Given that models were trained separately for Sentinel-1 and Sentinel-2 data, each model had 2 different learning rates corresponding to each dataset including U-Net (SAR) and U-Net (Optical) models. For U-Net, the optimum rates were 0.00022908 when trained on SAR data and 0.00019054 when trained on optical data. DeepLabV3 achieved its best performance with learning rates of 0.00013182 for SAR data and 0.00007585 for optical data. HRNet obtained optimum rates of 0.00006309 for SAR data and 0.00015848 for optical data. Finally, Pix2Pix the best learning rates was 0.00010964 for SAR data and 0.00019054 for optical data.

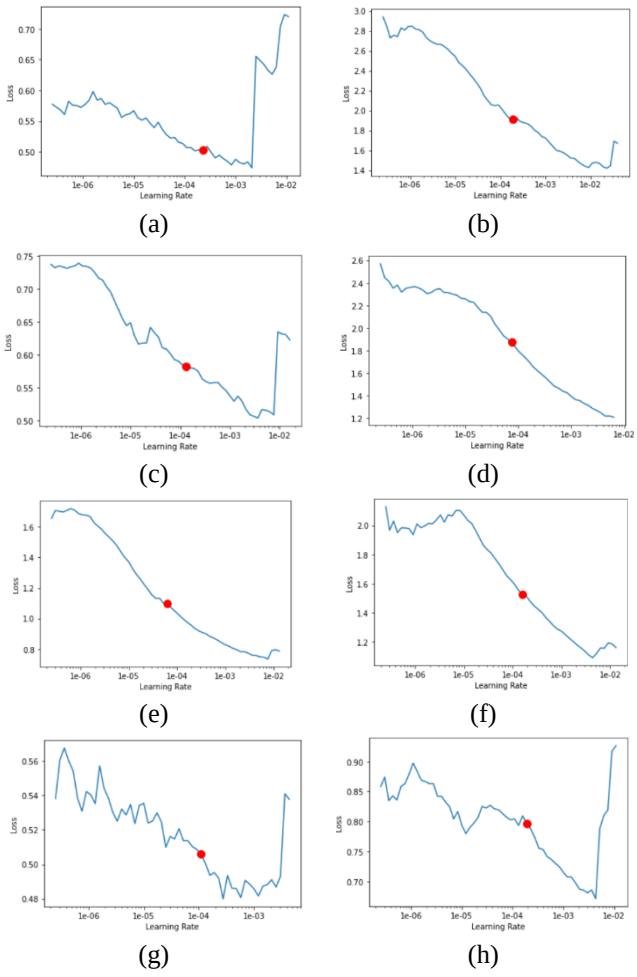


Figure 4. Learning curve for all models. (a) U-Net (SAR), (b) U-Net (Optical), (c) DeepLabV3 (SAR), (d) DeepLabV3 (Optical), (e) HRNet (SAR), (f) HRNet (Optical), (g) Pix2Pix (SAR), (h) Pix2Pix (Optical)

4.4. Training Configuration

All models were implemented in TensorFlow and trained on a NVIDIA GeForce RTX 4070 Ti GPU with an Intel Core i7-13700K CPU and the Adam optimizer was used for all the models. To improve convergence, we employed the "ReduceLROnPlateau" callback function, which adaptively reduces the learning rate when validation performance plateaued. The batch size was set to 8, and training was conducted for up to 100 epochs with early stopping based on validation loss to prevent overfitting. For CNN-based models, the loss function was binary cross-entropy (BCE) using Equation 3 (Guo et al., 2022):

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i.\log(\hat{y}_i) + (1-y_i).\log(1-\hat{y}_i)] \quad (3)$$

where, y_i is the ground-truth label of pixel i (1 = flood, 0 = non-flood), $\hat{y}_i$ is the predicted probability of flooding, and N is the total number of pixels in the batch. For Pix2Pix models, we used the default loss functions as defined in the original implementations. Pix2Pix combines adversarial loss with an L1 reconstruction term (Equations 4-6) (Isola et al., 2017):

$$L_{Pix2Pix} = L_{GAN}(G,D) + \lambda L_{L1}(G) \quad (4)$$

$$L_{GAN}(G,D) = E_{x,y}[\log D(x,y)] + E_x[\log(1-D(x,G(x))]] \quad (5)$$

$$L_{L1}(G) = E_{x,y}[\|y - G(x)\|] \quad (6)$$

where, x is the input, y is the ground-truth mask, G is the generator and D is the discriminator. The hyperparameter λ (in Pix2Pix model it is set to 100) controls the trade-off between adversarial sharpness and pixel-level accuracy.

4.5. Evaluation Metrics

The performance of all models was evaluated on the held-out country-level test set from Sen1Floods11 (see Section 3). To provide a comprehensive assessment, we used five widely adopted metrics in semantic segmentation including Overall Accuracy, Intersection over Union (IoU), Precision, Recall and F1-score presented using Equation 7 to 11 (He et al., 2022):

$$\text{Overall Accuracy} = \frac{\text{TP}}{\text{TP+TN+FP+FN}} \quad (7)$$

$$IoU = \frac{TP}{TP+FP+FN} \quad (8)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (9)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (10)$$

$$\text{F1} - \text{score} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \quad (11)$$

5. Results

5.1. Training time analysis

First, we evaluated the computational efficiency of the models by recording their total training times. As summarized in Table 2, all architectures trained considerably faster on the SAR data compared to optical data. This difference arises mainly from the higher dimensionality and spectral complexity of Sentinel-2 optical imagery, which increases the computational load.

Table 2. Comparison of training time for different models

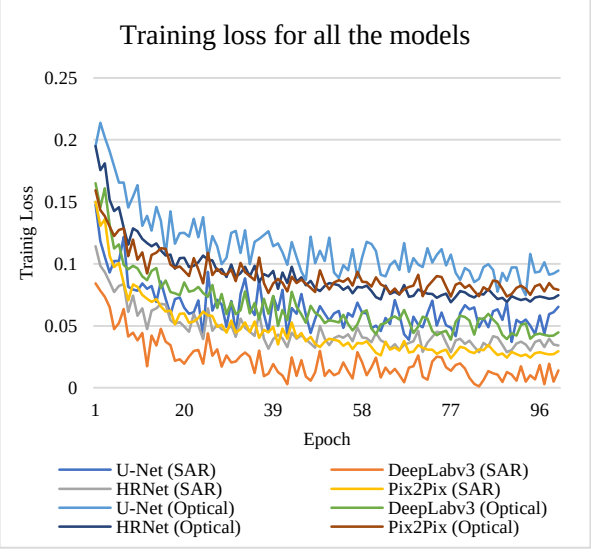
Model	Training time with SAR data (hh:mm)	Training time with optical data (hh:mm)
U-Net	6:40	16:50
DeepLabV3	8:30	24:10
HRNet	9:20	27:40
Pix2Pix	11:26	30:20

The analysis of training times shows that models trained

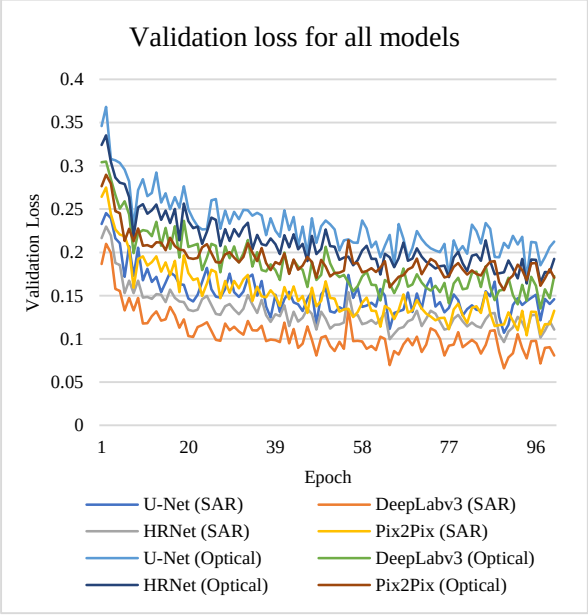
with Sentinel-1 SAR data consistently required less time to converge compared to those trained with Sentinel-2 optical imagery. U-Net completed training in 6 hours and 40 minutes on SAR data, whereas training on optical data took 16 hours and 50 minutes. DeepLabV3 required 8 hours and 30 minutes on SAR and 24 hours and 10 minutes on optical data. Pix2Pix had the longest training times, with 11 hours and 26 minutes for SAR and 30 hours and 20 minutes for optical data. These results indicate that training with optical imagery is nearly two to three times more time-consuming than with SAR imagery across all tested architectures.

5.2. Models' evaluation

Figure 5 presents the training and validation loss curves for U-Net, DeepLabV3, HRNet and Pix2Pix models using both Sentinel-1 SAR and Sentinel-2 optical datasets. All the employed models demonstrate a clear downward trend in training loss, indicating effective optimization. models trained on SAR data achieved lower training and validation losses than those trained on optical data. This highlights the strength of SAR imagery for flood extent mapping, as it is less affected by atmospheric conditions and provides stronger contrast between water and non-water surfaces. In contrast, optical data introduced greater variability, leading to slower convergence and higher final loss values.



(a)

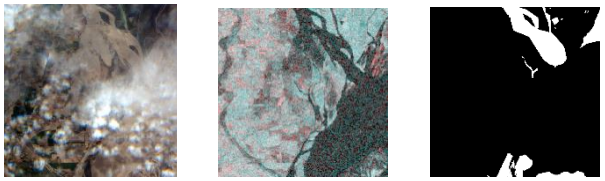


(b)

Figure 5. Training (a) and validation (b) loss for all models

Among the models, DeepLabV3 showed the lowest and most stable loss values for both SAR and optical datasets, while U-Net exhibited stronger fluctuations, especially with optical data. HRNet and Pix2Pix converged to intermediate levels but demonstrated clear improvements when trained with SAR inputs compared to optical inputs.

The close alignment between training and validation curves across all the experiments indicates limited overfitting and good generalization capability.



(a) (b) (c)

(d) (e) (f)

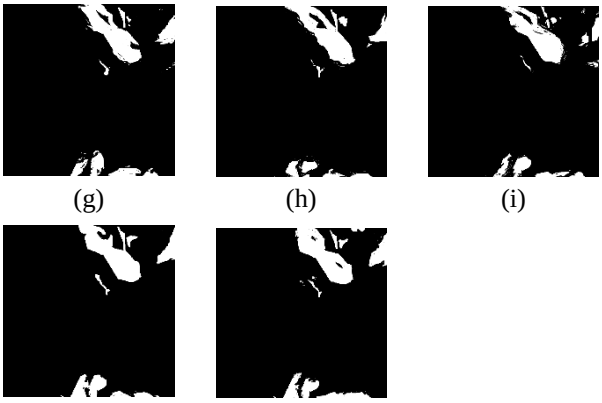


Figure 6. Comparison of input data, ground truth, and model outputs. (a) Optical image, (b) SAR image, (c) Ground truth, (d) U-Net (SAR), (e) U-Net (Optical), (f) DeepLabV3 (SAR), (g) DeepLabV3 (Optical), (h) HRNet (SAR), (i) HRNet (Optical), (j) Pix2Pix (SAR), (k) Pix2Pix (Optical)

Table 3 summarizes the performance of U-Net, DeepLabV3, HRNet and Pix2Pix on the test set using both Sentinel-1 SAR and Sentinel-2 optical datasets. Evaluation metrics include Overall Accuracy, IoU, Precision, Recall and F1-score. The results indicate that SAR-based models outperformed optical-based models across all architectures, consistent with the training and validation loss trends. The shorter training time for SAR data when considered alongside the higher segmentation accuracy achieved by SAR-based models, highlights the dual advantage of SAR data, which not only provides superior mapping performance but also allows for more computationally efficient model training. Among the CNN architectures, DeepLabV3 achieved the best overall balance, delivering the highest IoU (84.56%), Precision (94.63%) and F1-scores (92.35%). This demonstrates its strong ability to delineate flooded regions accurately while minimizing false positives. HRNet ranked second, with slightly lower IoU and Precision but competitive Recall values. U-Net showed the weakest performance, consistent with its instability observed during training.

Pix2Pix achieved the highest Overall Accuracy (97.45%) and Recall (93.79%) among all models, indicating its strong ability to detect most flooded areas. This is likely due to its generative architecture, which encourages capturing broader spatial patterns and producing more inclusive predictions. Consequently, Pix2Pix tends to over-predict flooded regions, resulting in lower Precision and IoU compared to DeepLabV3, but ensuring that fewer flooded areas are missed.

In summary, DeepLabV3 provides the most precise segmentation maps, while Pix2Pix offers more comprehensive coverage of flooded regions. The choice between the two depends on the application: Pix2Pix may be preferred for rapid emergency response where missing flooded areas is costly, whereas DeepLabV3 is better suited for high-precision flood risk assessment.

Table 3. Evaluation metrics summary

Data	Model	Metrics				
		Overall Accuracy (%)	IoU (%)	Precision (%)	Recall (%)	F1-Score (%)
SAR	U-Net	85.87	73.11	81.43	83.78	82.58
	DeepLabV3	95.76	**84.56**	**94.63**	90.18	**92.35**
	HRNet	90.35	80.23	87.19	89.45	88.30
	Pix2Pix	**97.45**	81.34	89.71	**93.79**	91.70
Optical	U-Net	83.23	70.36	80.49	83.47	81.95
	DeepLabV3	94.78	80.29	92.52	89.22	90.84
	HRNet	89.64	79.64	86.01	88.18	87.08
	Pix2Pix	95.06	78.48	89.52	90.10	89.80

6. Conclusion and Future Research Directions

6.1. Advantages and Limitations of CNN and GAN-based Architectures for Flood Mapping

The experimental results highlight clear strengths and weaknesses associated with both CNN- and GAN-based approaches for flood extent mapping. Among the evaluated CNN architectures, DeepLabV3 consistently achieved the highest Intersection over Union (IoU), Precision and F1-score, indicating its strong capability for accurately delineating flooded regions while minimizing false positives. This performance can be attributed to its use of dilated convolutions, which enable effective capture of long-range contextual information without sacrificing spatial resolution. HRNet also demonstrated competitive performance, particularly in preserving fine-scale flood boundaries, due to its high-resolution feature representation. However, CNN-based models generally exhibited lower Recall compared to Pix2Pix, suggesting a tendency to under-detect some flooded pixels, especially in ambiguous or transitional areas.

In contrast, the Pix2Pix conditional GAN exhibited the highest Overall Accuracy and Recall across both SAR and optical datasets, indicating its effectiveness in detecting the majority of flooded areas. The adversarial training strategy encourages spatially coherent and inclusive flood predictions, which is advantageous in applications where missing flooded regions could have severe consequences, such as rapid disaster response. Nevertheless, this strength comes at the cost of increased overprediction, as reflected by lower Precision and IoU values compared to DeepLabV3. Additionally, Pix2Pix required longer training times and demonstrated higher computational complexity, particularly when trained on multi-spectral optical data.

These findings suggest a trade-off between precision-oriented CNN models and recall-oriented GAN models in flood extent mapping. CNN-based architectures are more

suitable for applications requiring high spatial accuracy and reliable boundary delineation, such as flood risk assessment and damage estimation. Conversely, GAN-based approaches like Pix2Pix may be more appropriate for emergency response scenarios where maximizing flood detection is prioritized over strict boundary accuracy. Understanding these advantages and limitations is essential for selecting an appropriate model architecture based on specific operational requirements.

6.2. Conclusion

The comparative analysis of CNN- and GAN-based architectures provides important insights into their respective strengths and limitations for flood extent mapping. The results demonstrate that CNN-based models, particularly DeepLabV3, consistently achieved higher Intersection over Union (IoU), Precision and F1-scores, underscoring their effectiveness in delineating flooded areas with high spatial accuracy. These findings are in agreement with numerous recent studies that report the strong performance of CNN architectures, especially context-aware models, for flood extent extraction from SAR and optical imagery (Gao et al., 2024). The ability of CNNs to capture multi-scale contextual information and maintain spatial consistency has been widely recognized as a key factor in their success for semantic segmentation tasks in remote sensing.

At the same time, the performance of the Pix2Pix conditional GAN highlights the potential of adversarial learning for flood mapping applications. The Pix2Pix model achieved the highest Overall Accuracy and Recall, indicating its effectiveness in capturing broader spatial flood patterns and reducing omission errors. This behavior is consistent with previous research that has shown GAN-based models tend to produce more inclusive segmentation outputs due to adversarial training, albeit often at the cost of reduced Precision and increased overprediction. Our findings therefore align with existing observations that GANs may be better suited for applications where minimizing missed detections is prioritized over strict boundary accuracy.

The superiority of SAR-based models across all evaluated architectures further reinforces conclusions from earlier flood mapping studies that emphasize the robustness of SAR data under adverse weather conditions. Unlike optical imagery, which remains sensitive to atmospheric effects, cloud cover and surface reflectance variability, SAR imagery consistently provided clearer contrast between flooded and non-flooded areas. This confirms the widely reported suitability of SAR as a primary data source for operational flood monitoring. Nevertheless, optical imagery retains value by providing complementary spectral information, and the observed limitations suggest that multimodal approaches may yield more balanced and robust flood mapping performance than single-modality models.

Overall, the results indicate that no single deep learning architecture can be considered universally optimal for flood extent mapping. Instead, the choice of model should be guided by application-specific requirements. CNN-based architectures are more appropriate for precision-demanding tasks such as flood risk assessment and damage estimation, whereas GAN-based models may be preferable for rapid emergency response scenarios where minimizing missed flooded areas is critical.

6.3. Future Research Directions

The outcomes of this study point to several avenues for further research. A natural extension is the development of multimodal fusion frameworks that simultaneously leverage SAR and optical imagery, as well as other geospatial data such as digital elevation models (DEMs) or land cover information, to provide richer contextual information for flood segmentation. Advances in attention mechanisms (Saleh et al., 2024) and transformer-based models (Song et al., 2025) offer promising opportunities for more effective cross-modal integration. Another important direction for future research involves improving model generalization across geographic domains. Domain adaptation techniques including unsupervised style transfer and cycle-consistent adversarial networks (CycleGAN) (Zhu et al., 2017), may help overcome regional biases in training data and support the transfer of models trained on benchmark datasets to operational settings. Equally important is the adaptation of these methods for real-time applications. This would require optimizing computational efficiency, incorporating uncertainty quantification, and ensuring interpretability, which are essential for their deployment in decision-support systems used by emergency agencies.

References

Bentivoglio, R., Isufi, E., Jonkman, S. N., and Taormina, R. 2022. Deep learning methods for flood mapping: a review of existing applications and future research directions. Hydrology and Earth System Sciences Discussion, 2022, 1-50. https://doi.org/10.5194/hess-26-4345-2022

Chen, L.-C., Papandreou, G., Schroff, F., and Adam, H. 2017. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587. https://doi.org/10.48550/arXiv.1706.05587

Dubey, V., and Katarya, R. 2024. SSR-GAN: super resolution-based generative adversarial networks model for flood image enhancement. Signal, Image and Video Processing, 18(8), 5763-5773. https://doi.org/10.1007/s11760-024-03269-z

Gao, W., Liao, Y., Chen, Y., Lai, C., He, S., and Wang, Z. 2024. Enhancing transparency in data-driven urban pluvial flood prediction using an explainable CNN model. Journal of Hydrology, 645, 132228. https://doi.org/10.1016/j.jhydrol.2024.132228

Girisha, S., Savitha, G., and Sughosh, P. 2025. Flood Extent Mapping in SAR Images using Semi-Supervised Approach. Results in Engineering, 105304. https://doi.org/10.1016/j.rineng.2025.105304

Guo, C., Chen, X., Chen, Y., and Yu, C. 2022. Multi-stage attentive network for motion deblurring via binary cross-entropy loss. Entropy, 24(10), 1414. https://doi.org/10.3390/e24101414

Haghighi Gashti, E., Bahiraei, H., Valadan Zoej, M. J., and Ghaderpour, E. 2025. Fusion of Aerial and Satellite Images for Automatic Extraction of Building Footprint Information Using Deep Neural Networks. Information, 16(5), 380. https://doi.org/10.3390/info16050380

Haghighi Gashti, E., Delavar, M. R., Guan, H., & Li, J. (2024). Semantic Segmentation Uncertainty Assessment of Different U-net Architectures for Extracting Building Footprints. ISPRS annals of the photogrammetry, remote sensing and spatial information sciences, 10, 141-148. https://doi.org/10.5194/isprs-annals-X-4-2024-141-2024

He, H., Jiang, Z., Gao, K., Narges Fatholahi, S., Tan, W., Hu, B. and Li, J. 2022. Waterloo building dataset: A city-scale vector building dataset for mapping building footprints using aerial orthoimagery. Geomatica, 75(3), 99-115. https://doi.org/10.1139/geomat-2021-0006

Isola, P., Zhu, J. -Y., Zhou, T. and Efros, A. A. 2017. Image-to-Image Translation with Conditional Adversarial Networks, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 5967-5976. https://doi.org/10.1109/CVPR.2017.632.

Kundzewicz, Z. W., Kanae, S., Seneviratne, S. I., Handmer, J., Nicholls, N., Peduzzi, P., Mach, K. 2014. Flood risk and climate change: global and regional perspectives. Hydrological Sciences Journal, 59(1), 1-28. https://doi.org/10.1080/02626667.2013.857411

Kurapati, P. V., Babu, A., Rajosarimalala, S. T., & Pyla, K. R. 2020. Flood mapping and damage assessment using sentinel-1 & 2 in Google Earth Engine of Port Berge & Mampikony districts, Sophia region, Madagascar. In Proceeding of 41st Asian conference on remote sensing (ACRS) (pp. 1-9).

Li, Y., Martinis, S., Plank, S., and Ludwig, R. 2018. An automatic change detection approach for rapid flood mapping in Sentinel-1 SAR data. International Journal of Applied Earth Observation and Geoinformation, 73, 123-135. https://doi.org/10.1016/j.jag.2018.05.023

Long, S., Fatoyinbo, T. E., and Policelli, F. 2014. Flood extent mapping for Namibia using change detection and thresholding with SAR. Environmental Research Letters, 9(3), 035002. https://doi.org/ 10.1088/1748-9326/9/3/035002

Misra, A., White, K., Nsutezo, S. F., Straka III, W., and Lavista, J. 2025. Mapping global floods with 10 years of satellite radar data. Nature Communications, 16(1), 5762. https://doi.org/10.1038/s41467-025-60973-1

Ronneberger, O., Fischer, P., and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Cham: Springer international publishing. https://doi.org/10.1007/978-3-319-24574-4_28

Saleh, T., Holail, S., Zahran, M., Xiao, X. and Xia, G. -S. 2024. LiST-Net: Enhanced Flood Mapping with Lightweight SAR Transformer Network and Dimension-Wise Attention, in IEEE Transactions on Geoscience and Remote Sensing, vol. 62, pp. 1-17, 2024, Art no. 5211817. https://doi.org/10.1109/TGRS.2024.3397797

Sarker, C., Mejias, L., Maire, F., and Woodley, A. 2019. Flood mapping with convolutional neural networks using spatio-contextual pixel information. Remote Sensing, 11(19), 2331. https://doi.org/10.3390/rs11192331

Schumann, G., Giustarini, L., Tarpanelli, A., Jarihani, B., and Martinis, S. 2023. Flood modeling and prediction using earth observation data. Surveys in Geophysics, 44(5), 1553-1578. https://doi.org/10.1007/s10712-022-09751-y

Sekine, T., Yamanaka, A., Eda, T., Udagawa, T. and Busto, M. 2024. A GAN-Based SAR-Optical Data Fusion Approach for Enhanced Flood Mapping, IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium, Athens, Greece, 2024, pp. 8958-8962. https://doi.org/10.1109/IGARSS53475.2024.10641093

Seydi, S. T., Saeidi, V., Kalantar, B., Ueda, N., van Genderen, J. L., Maskouni, F. H., and Aria, F. A. 2022. Fusion of the multisource datasets for flood extent mapping based on ensemble convolutional neural network (CNN) model. Journal of sensors, 2022(1), 2887502. https://doi.org/10.1155/2022/2887502

Sirsant, S., Hinge, G., Singh, H., and Hamouda, M. A. 2025. A hybrid convolutional neural network model coupled with AdaBoost regressor for flood mapping using geotagged flood photographs. Natural Hazards, 121(5), 5799-5819. https://doi.org/10.1007/s11069-024-07041-x

Smith, L. N. 2017. Cyclical Learning Rates for Training Neural Networks, 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, USA, 2017, pp. 464-472. https://doi.org/10.1109/WACV.2017.58

Song, W., Guan, M., and Yu, D. 2025. SwinFlood: A hybrid CNN-Swin Transformer model for rapid spatiotemporal flood simulation. Journal of Hydrology, 660, 133280. https://doi.org/10.1016/j.jhydrol.2025.133280

Tavus, B., Can, R., and Kocaman, S. 2022. A cnn-based flood mapping approach using sentinel-1 data. ISPRS annals of the photogrammetry, remote sensing and spatial information sciences, V-3-2022, 549–556, https://doi.org/10.5194/isprs-annals-V-3-2022-549-2022

Tedla, M. G., Cho, Y., and Rasmy, M. 2025. Exploring the dynamics of flood and drought hazards and adaptation strategies in the horn of Africa. Water International, 50(3-4), 225-234. https://doi.org/10.1080/02508060.2025.2463797

Vamvakeridou-Lyroudia, L., Chen, A., Khoury, M., Gibson, M., Kostaridis, A., Stewart, D. and Savic, D. 2020. Assessing and visualising hazard impacts to enhance the resilience of Critical Infrastructures to urban flooding. Science of the Total Environment, 707, 136078. https://doi.org/10.1016/j.scitotenv.2019.136078

Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y. and Wang, X. 2020. Deep high-resolution representation learning for visual recognition. IEEE transactions on pattern analysis and machine intelligence, vol. 43, no. 10, pp. 3349-3364, 1 Oct. 2021. https://doi.org/10.1109/TPAMI.2020.2983686

Wang, J., Yang, H., He, Y., Zheng, F., Liu, Z. and Chen, H. 2024. An unpaired SAR-to-optical image translation method based on Schrödinger bridge network and multi-scale feature fusion. Scientific Reports, 14(1), 27047. https://doi.org/10.1038/s41598-024-75762-x

Wang, T. -C., Liu, M. -Y., Zhu, J. -Y., Tao, A., Kautz, J. and Catanzaro, B. 2018. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 8798-8807. https://doi.org/10.1109/CVPR.2018.00917

Zhu, J. -Y., Park, T., Isola, P. and Efros, A. A. 2017. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks, 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 2242-2251. https://doi.org/10.1109/ICCV.2017.244