ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 05, May 2026

Deep Learning for Sentinel-1 Flood Monitoring: A Review of Architectural Advances and Performance Benchmarks

[1] Shishir Mishra, [2] Shashank Kumar, [3] Sneha Kishore, [4] Amit Kumar

[1] [2] [3] [4] Department of Artificial Intelligence, Galgotias College of Engineering and Technology, Gautam Buddh Nagar, Uttar Pradesh, India

Corresponding Author Email: [1] shishir.22gecbai132@galgotiacollege.edu

***Abstract*— - Floods are one of the most destructive natural hazards in South Asia, and it is difficult to map during the monsoon season because cloud cover blocks optical satellites days at a time. Sentinel-1 SAR avoids this: it can work behind clouds, during the day or night, at 10-meter resolution and with repeat cycles of 6-12 days. The paper is a review of the various ways deep learning architectures to flood segmentation of Sentinel-1 have evolved, starting with fully convolutional networks, and moving on to modern vision transformers. We compare three families of models: UNet++ (CNN), Swin-UNet (transformer) and SegFormer-B0 (efficient transformer) on a regionally stratified dataset with 55% South Asian coverage using Sen1Floods11, Kuro Siwo and UrbanSARFloods as benchmarks. With only 3.7M parameters, SegFormer-B0 has an IoU of 0.6404, which is very small but still provides a benefit over other models that are both smaller and larger at the same time. There are four gaps: urban double-bounce scattering, sparse regional labeled data, topographic false positives, and near-real-time inference. We map the way ahead to each.**

***Index Terms*— Synthetic Aperture Radar, Sentinel-1, Flood Mapping, Semantic Segmentation, Vision Transformers, Deep Learning, Indian Subcontinent.**

I. INTRODUCTION

The Indian subcontinent takes a disproportional amount of annual losses and floods are one of them. Incidents such as the 2018 Kerala floods and frequent inundations in Bihar and Assam demonstrate how urgently we should have accurate, timely flood maps to aid emergency response [1, 2]. Here, optical remote sensing is mostly useless = monsoon clouds block the visible and near-infrared bands days or weeks at a time [3].

C-band SAR on the Sentinel-1A/B constellation gets around this. Being an active microwave sensor, it can traverse the cloud cover, and can be used in the absence of sunlight, which can see large areas of flood at 10 m resolution with revisit periods of 6-12 days [4]. The physics are simple, that is, flat open water reflects radar pulses farther away than the sensor itself, resulting in a characteristic low-backscatter signature [5].

The move towards threshold-based classification and to deep learning has transformed the discipline in the last ten years. The thresholding and change-detection ratio between pre- and post-event images, used to develop early SAR flood mapping, originated with Otsu [6, 7]. The latter techniques are applicable to flat homogeneous surfaces, but they fail to differentiate water and shadowed landscapes or smooth agricultural ones-a challenge which becomes increasingly difficult in the varied topography of South Asia [8]. In comparison, semantic segmentation networks learn the spatial context which provides more robust flood boundary delineation even in the presence of heavy speckle and class imbalance [9, 10].

This paper contributes four things: (i) it tracks the architectural evolution of fully convolutional networks to transformer-based segmentation; (ii) it benchmarks three architectures on a regionally stratified dataset consisting of 55% South Asians; (iii) it evaluates domain transfer to Indian flood events; and (iv) it identifies four research gaps that must be bridged before near-real-time flood monitoring can be operationally deployed across the subcontinent.

II. BACKGROUND AND RELATED WORK

A. SAR-Based Flood Mapping: Classical Methods

Initial mapping of the Indian floods used optical indices such as the MNDWI of Sentinel-2. Vishnu et al. (2019) estimated close to 90 percent surface-water expansion during the August 2018 Kerala floods, but nearly 85 percent of the affected pixels were cloud-contaminated at maximum inundation [2].

That gap was filled by SAR thresholding. Tiwari et al. (2020) cross-fertilized Sentinel-1 GRD images with the method of Otsu on Google Earth Engine with an accuracy of more than 94% and demonstrating how cloud-based SAR processing can be scaled to operational use [6]. Vanama et al. (2021) expanded this to multi-temporal acquisitions, and found that normalized change indices provide cleaner flood/non-flood separation than simple backscatter ratios [7]. A Bayesian probabilistic overlay was added to SAR backscatter by Sherpa et al. (2020), and converted the backscatter into pixel-level flood probability and revealed that reservoir releases were more effective drivers of inundation extent than local rainfall [11].

B. Deep Learning to SAR Segmentation

The field was changed by the Sen1Floods11 benchmark (Bonafilia et al., 2020): 4,831 globally distributed 512x512 Sentinel-1 chips with expert labels on 11 floods across 14 countries [12]. This enabled systematic architecture comparisons. Yadav et al. (2022) demonstrated a deep attentive fusion network with spatial and channel-wise attention that achieved 0.67 IoU on Sen1Floods11 by avoiding speckle in feature maps [9]. Paul and Ganju (2021) demonstrated that semi-supervised learning on a small size labelled set, combined with a large size unlabelled archive, can achieve the same performance as fully supervised in high-noise conditions-promising to label-scarce regional datasets [10].

End-to-end pixel-wise prediction using fully convolutional networks was proposed by Long et al. (2015) [13]. U-Net was proposed by Ronneberger et al. (2015), and the encoder-decoder of this architecture is symmetric with skip connections [14]. Zhou et al. (2018) enhanced it with UNet++, which replaces simple skip connections with nested dense pathways that bridging the semantic gap between shallow and deep features, and improving the boundary accuracy under limited training data [15]. Li and Demir (2023) validated that UNet++ works well on continental scale flooding [16].

C. Global Benchmarks and Foundation Models

Two new datasets took the field a step further. Kuro Siwo (Bountos et al., 2023) covers 33 billion square meters of the inundation in several continents with expert-labeled multi-temporal Sentinel-1 images, which provides the depth of time required to train and evaluate multi-temporal architectures [17]. The first Sentinel-1 SLC benchmark of urban floods is UrbanSARFloods (Zhao et al., 2024), which reveals an actual vulnerability: the intensity-only nature of the SARFloods approach prevented the handling of double-bouncing returns related to the facades of buildings [18].

Geo-foundational models are now also coming into the scene. As shown by Scheibenreif et al. (2024), fine-tuning a fraction of a pre-trained vision encoder with minimal labeled data closely resembles full supervised training-which is directly relevant in label-scarce regional applications [19]. Portales-Julia et al. (2025) cross-sensored the magnitude of performance degradation due to cross-sensory transfer of flood detection models across sensor modalities and geographic areas [20].

III. ARCHITECTURAL PROGRESSION: CNNS TO TRANSFORMERS

A. Convolutional Encoder–Decoders

UNet++ is the most popular CNN baseline used to segment SAR floods. Its nested skip pathways enable deep supervision at various depths of decoders, which aids in the delineation of the boundaries when there is a dearth of labeled data-critical in the light of limited South Asian flood images. EfficientNet-B4 encoder is an encoder that extracts strong multi-scale features, and is capable of being trained on consumer-grade GPUs [21]. Melgar-Garcia et al. (2023) compared UNet, WU-Net and UNet++ on predicting tropical floods, indicating that UNet++ is consistently superior in terms of boundary or boundary accuracy, but it still has false positives on smooth agricultural fields with backscatter of water [22].

B. Hierarchical Vision Transformers

Dosovitskiy et al. (2021) showed self-attention over image patches can match convolutional architectures on natural image benchmarks [23]. However, the standard ViT computes all patches in the world, which is computationally infeasible on high-resolution SAR tiles. This was resolved by Liu et al. (2021) using the Swin Transformer: non-overlapping local windows with shifted-window self-attention to capture cross-window dependencies [24]. Swin-UNet, which is applied using a U-Net-like decoder, is capable of capturing long-range correlations between flood bodies separated by elevated terrain, something that local convolutions cannot achieve. In a Siamese arrangement, Tang et al. (2024) scaled it down and recorded high performance in terms of multi-temporal flood boundary extraction [25].

C. SegFormer: Efficiency Frontier

Xie et al. (2021) introduced SegFormer, which pairs a hierarchical Mix Vision Transformer encoder with a lightweight all-MLP decoder, specifically without positional encodings to enhance generalization at input resolutions [26]. The MiT-B0 version does this with only 3.7M parameters, and can run on 6 GB VRAM with latency that is compatible with emergency mapping. Fine-tuning on smaller regional datasets avoids the resolution mismatch inherent to standard ViT variants, a valuable practical consideration to Indian subcontinent workflows.

IV. BENCHMARKING AND REGIONAL PERFORMANCE

A. Data set and Evaluation Protocol.

We sample a regionally stratified subset of Sen1Floods11: 1,932 non-overlapping 512×512 chips at 10 m resolution, with India, Pakistan and Sri Lanka contributing 55 percent of the chips. To avoid leakage of spatial autocorrelation in our models, like other known sources of inflated measures in geospatial deep learning, we use event-wise spatial splits: entire flood events are held out as test units. Ablation experiments verified that random tile splitting boosted IoU by 512 percentage points compared to event-wise splitting [27].

There is a great imbalance between classes: median flood coverage per chip is 1.23%, with 86 percent of chips covering less than 10 percent of flooded pixels. We use focal loss (γ = 2.0) [28], which down-weights simple non-flood pixels and concentrates gradient changes on ambiguous flood boundaries. All models are trained with mixed-precision and gradient accumulation in 6 GB VRAM budget on consumer hardware.

Table I: Segmentation Performance on Indian Subcontinent Test Set

Model	Paradigm	Params (M)	IoU	Dice (F1)	Precision	AUC-ROC
UNet++ (Eff-B4)	CNN	8.6	0.4708	0.6402	0.7569	0.9482
Swin-Tiny UNet	ViT	11.4	0.6231	0.7678	0.7506	0.9737
SegFormer-B0	ViT	3.7	0.6404	0.7808	0.8282	0.9767
DAM-Net* [9]	Attn. CNN	17.5	~0.67	~0.80	—	—
Attentive U-Net* [9]	Attn. CNN	—	0.6700	—	—	—

**Results from published literature, not re-implemented here.*

B. Quantitative Results

Table I presents a performance of segmentation in all three architectures together with published results of more complex fusion models. The transition of CNN to transformer architectures results in a relative increase in the mean IoU of approximately 36% or 0.17. SegFormer-B0 has the best precision (0.8282), and this implies that the SAR speckle and low-backscatter terrain will have fewer false positives. The best recall (0.7858) is reached by Swin-Tiny UNet, which retrieves fragmented urban flood areas that CNNs miss. Both transformer models have larger than 0.97 AUC-ROC ratio, which indicates reliable discrimination throughout the full range of operating threshold-important during the calibration of agencies of detection thresholds to acceptable false-alarm rates.

C. Domain Transfer and South Asia Gap.

The models that are trained using the data that is distributed globally will always perform poorly when used in a particular region. Portales-Julia et al. (2025) measured this: SAR-trained models lose 5 -18 IoU points when transferred to new sensor-domain combinations [20]. The Indian subcontinent has a gap driven by three factors: (1) during the monsoon season, high-moisture rice paddies cause SAR backscatter to be close to shallow inundation; (2) densely built flood-prone corridors generate flood patterns that do not have a close analogue within European or North American training data [6, 8, 11].

V. RESEARCH GAPS AND FUTURE DIRECTIONS

A. Urban Double-Bounce Scattering

Sentinel-1 SLC data is phase-based data, which is largely ignored by intensity-based flood segmentation. Radar pulses in urban areas reflect between surfaces of water and building facades, creating high-backscatter returns, which resemble dry land in intensity-only pipelines. This problem has been formalised by UrbanSARFloods [18]) which however, has no widely used segmentation architecture integrated yet to combine interfermetric coherence with intensity. Coherence change detection-measuring the decorrelation between pre- and post-flood SLC pairs-is physically sensitive to surface flooding irrespective of which way the building is oriented. It is computationally complex and technically mature, and it is the best prospect to gain accurate mapping of urban inundation.

B. Regional Data Sparseness and Labeling

Sen1Floods11 [12] is both broad geographically and thin on South Asian monsoon processes, agricultural mosaics densities, and backscatter on tropical vegetation. Kuro Siwo [17] enhances the quality of labels worldwide, though professional marking of the events of the Indian subcontinent is so far a rare occurrence. Semi-supervised learning [10] and self-supervised pre-training [19] are helpful, but the field still requires a large, professionally labelled SAR flood corpus covering Bihar, Assam, Odisha and the Brahmaputra Delta.

C. Topographic False Positives

In mountainous areas radar layover and shadow produce a long low-backscatter region that appears like open water. The existing segmentation models lack the terrain ancillaries. DEM derivatives-topographic wetness index, slope gradient, and flow accumulation-carry strong physical priors regarding flood susceptibility and could potentially lead to a significant decrease in false detections in steep terrain. The addition of SRTM or Copernicus DEM layers as an additional input channel is a low cost augmentation [7, 11] that has shown promise in Kerala flood studies but has not been systematically tested across architectures.

D. Operational Timeliness and Edge Deployment

Flood maps provided in excess of 12 hours beyond the peak inundation have little value in terms of emergency evacuation guidance. State-of-the-art fusion architectures such as DAM-Net demand 4.6 GB VRAM and inference times that are unfeasible with 1-hour deployment requirements on field hardware. Geo-models Foundational geo-models, large pre-trained encoders, that are fine-tuned with limited labeled data [19] might offer a route to not only accuracy but also speed, but their real-time characteristics in embedded systems have not been studied. As demonstrated by Ryd and Nearing (2025), the use of local gauge and rainfall data by fine-tuning global forecasting models can significantly enhance spatial precision [29], indicating that hybrid physics-ML methods can be superior to pure deep learning in operational environments.

VI. CONCLUSION

We have demonstrated the shift between threshold-based SAR flood detection to hierarchical transformer segmentation, and have shown that current vision transformers-especially SegFormer-B0-offer the best accuracytoparameter trade-off to flood monitoring across the Indian subcontinent. Transformers achieve about 36 percent relative IoU relative to CNN baselines by learning long-range spatial dependence, which cannot be learned using the CNN baselines. However, four gaps must be bridged-urban double-bounce scattering, paucity of regional data, topographic false positives and operational timeliness-before these advances can work to become disaster management tools. The future direction is a combination of multi-source data fusion (SAR coherence, DEM ancillaries, IoT sensor streams), large expert-labeled South Asian flood datasets, and lightweight foundational geo-models created to quickly regional fine-tune on consumer hardware.

REFERENCES

[1] V. Tiwari et al., "Flood inundation mapping—Kerala 2018: Harnessing the power of SAR, automatic threshold detection methods, and Google Earth Engine," PLOS ONE, vol. 15, no. 8, e0237324, 2020.
[2] C. L. Vishnu et al., "Satellite-based assessment of the August 2018 flood in parts of Kerala, India," Geomatics, Natural Hazards and Risk, vol. 10, no. 1, pp. 758–767, 2019.
[3] P. Tripathy and T. Malladi, "Global Flood Mapper: A novel Google Earth Engine application for rapid flood mapping using Sentinel-1 SAR," Natural Hazards, vol. 114, no. 2, pp. 1341–1363, 2022.
[4] S. Dhanabalan et al., "Flood monitoring using Sentinel-1 SAR data: A case study of the 2018 and 2019 floods in southern Kerala," Int. Archives Photogrammetry, Remote Sensing & Spatial Info. Sciences, vol. 44, pp. 37–41, 2021.
[5] N. Anusha and B. Bharathi, "Flood detection and flood mapping using multi-temporal synthetic aperture radar and optical data," Egyptian Journal of Remote Sensing and Space Science, vol. 23, no. 2, pp. 207–219, 2020.
[6] V. Tiwari et al., "Flood inundation mapping—Kerala 2018," PLOS ONE, vol. 15, no. 8, e0237324, 2020.
[7] V. Vanama, Y. Rao, and C. Bhatt, "Change detection-based flood mapping using multi-temporal Earth observation satellite images: The 2018 flood event in Kerala, India," European Journal of Remote Sensing, vol. 54, no. 1, pp. 42–58, 2021.
[8] R. Yadav, A. Nascetti, and Y. Ban, "Deep attentive fusion network for flood detection on uni-temporal Sentinel-1 data," Frontiers in Remote Sensing, vol. 3, p. 1060144, 2022.
[9] R. Yadav, A. Nascetti, and Y. Ban, "Deep attentive fusion network for flood detection," Frontiers in Remote Sensing, vol. 3, p. 1060144, 2022.
[10] S. Paul and S. Ganju, "Flood segmentation on Sentinel-1 SAR imagery with semi-supervised learning," arXiv:2107.08369, 2021.
[11] S. F. Sherpa et al., "Probabilistic mapping of the August 2018 flood in Kerala, India, using space-borne SAR," IEEE J. Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 896–913, 2020.
[12] D. Bonafilia, B. Tellman, T. Anderson, and E. Issenberg, "Sen1Floods11: A georeferenced dataset to train and test deep learning flood algorithms for Sentinel-1," Proc. IEEE/CVF CVPR Workshops, pp. 835–845, 2020.
[13] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," Proc. IEEE CVPR, pp. 3431–3440, 2015.
[14] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," Proc. MICCAI, pp. 234–241, 2015.
[15] Z. Zhou et al., "UNet++: A nested U-Net architecture for medical image segmentation," DLMIA/ML-CDS, LNCS, vol. 11045, pp. 3–11, 2018.
[16] Z. Li and I. Demir, "U-Net-based semantic classification for flood extent extraction using SAR imagery and the GEE platform," Science of the Total Environment, vol. 869, p. 161757, 2023.
[17] N. I. Bountos et al., "KURO SIWO: 33 billion m² under the water—A global multi-temporal satellite dataset for rapid flood mapping," arXiv:2311.12056, 2023.
[18] J. Zhao, Z. Xiong, and X. X. Zhu, "UrbanSARFloods: Sentinel-1 SLC-based benchmark dataset for urban and open-area flood mapping," Proc. IEEE/CVF CVPR, pp. 419–429, 2024.
[19] L. Scheibenreif, M. Mommert, and D. Borth, "Parameter-efficient self-supervised geospatial domain adaptation," Proc. IEEE/CVF CVPR, pp. 27841–27851, 2024.

ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Volume 13, Issue 05, May 2026

[20] E. Portales-Julia, G. Mateo-García, and L. Gómez-Chova, "Understanding flood detection models across Sentinel-1 and Sentinel-2 modalities and benchmark datasets," Remote Sensing of Environment, vol. 328, p. 114882, 2025.
[21] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," Proc. ICML, pp. 6105–6114, 2019.
[22] L. Melgar-García et al., "A novel semantic segmentation approach based on U-Net, WU-Net, and U-Net++ for predicting pluvial flood-sensitive areas in tropical regions," Int. Journal of Digital Earth, vol. 16, no. 1, pp. 3661–3679, 2023.
[23] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," Proc. ICLR, 2021.
[24] Z. Liu et al., "Swin Transformer: Hierarchical vision transformer using shifted windows," Proc. IEEE/CVF ICCV, pp. 10012–10022, 2021.
[25] Y. Tang, Z. Cao, N. Guo, and M. Jiang, "A Siamese Swin-UNet for image change detection," Scientific Reports, vol. 14, no. 1, p. 4577, 2024.
[26] E. Xie et al., "SegFormer: Simple and efficient design for semantic segmentation with transformers," Advances in Neural Information Processing Systems, vol. 34, pp. 12077–12090, 2021.
[27] G. Kattenborn et al., "Review on convolutional neural networks (CNN) in vegetation remote sensing," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 173, pp. 24–49, 2021.
[28] T.-Y. Lin et al., "Focal loss for dense object detection," Proc. IEEE/CVF ICCV, pp. 2980–2988, 2017.
[29] E. Ryd and G. Nearing, "Fine flood forecasts: Incorporating local data into global models through fine-tuning," arXiv:2504.12559, 2025.
[30] V. Mateo-García et al., "Towards global flood mapping onboard low cost satellites with machine learning," Scientific Reports, vol. 11, p. 7249, 2021.