U-Net++ and ConvLSTM based Semantic Event Stream Processing for Real-Time Flood Monitoring

A. Suresh[1], *Smitha* Kurian[2*], *Subhra* Chakraborty[3], *P.* Rajeswari[4], and *Leeladhar* Gudala[5]

[1]Department of Computer Science and Engineering, GRT Institute of Engineering and Technology, Tiruttani, India
[2]Department of Computer Science, HKBK College of Engineering, Bengaluru, India
[3]Department of Electronics and Communication Engineering, Nitte Meenakshi Institute of Technology, Nitte (Deemed to be University), Bengaluru, India
[4]Department of Electronics and Communication Engineering, Karpagam Institute of Technology, Coimbatore, India
[5]Independent Researcher

Abstract. Currently, Floods are the most popular and devastating natural hazards, and their effects on economic stability and social well-being are continuously increasing. Recent methods have focused on fast and accurate detection of submerged regions to improve emergency planning and damage estimation in spatial and temporal dimensions. Satellite Multispectral Images have limited spectral bands and low spatial resolution, which restrict the depth of analysis. Image processing methods use intensity-hue-saturation and statistical features to enhance the spatial and spectral data before segmentation and classification. The proposed UConvFloodNet starts by removing noise for sharper inputs using Wiener filtering, and increases the spatial contrast by converting pixel intensities. With these enhanced images, U-Net++ segments key areas such as flood zones, water bodies, and land covers and tracks how flooding evolves over time, and ConvLSTM adds a temporal layer, capturing changes frame by frame. Together, these steps form a streamlined outline for real-time flood monitoring, which combines image enhancement, spatial enhancement, precise segmentation, and dynamic tracking. The UConvFloodNet performs an Accuracy of 99.31 and 99.06 for Sen1Floods11 and S1GFloods datasets which is better than existing Compact Convolutional Tokenizer integrating U-Net and Vision Transformer (CCT-U-ViT).

1 Introduction

Recently, natural disasters have become more common; more dangerous, heavy rains and strong storms have caused serious damage to dams and other structures in many places [1]. Specialized analytical tools are designed in accordance with suggested guidelines for producing flood hazard maps [2]. Remote sensing is useful for tracking floods over large areas; however, satellite and drone sensors can struggle with low image quality and cloudy skies during storms [3]. Floods damage people and their homes, harm important infrastructure, pollute water and upset natural ecosystems, and cause a chain of related disasters [4]. Large storms such as typhoons are currently regular, but are sudden, forceful local rainstorms, which is unusual and is growing rapidly in number [5]. As global temperatures rise, climate change is caused by abnormal weather patterns, including intense storms and flooding, and the prediction depends heavily on remote monitoring systems [6]. Floods occur for many reasons, such as melting snow, large ocean waves, broken dams or earthworks, and cities growing too fast without proper drainage systems [7]. Quick and accurate flood detection helps representatives respond more quickly and reduce damage. It is also important to alert emergency teams about risky areas to prepare and take immediate action [8]. Previous research shown that combining optical and Synthetic Aperture Radar (SAR) data improves flood mapping, but many studies still depend on only one method, most regularly microwave remote sensing [9]. Flood mapping is typically carried out by comparing the current water coverage to a reference image that reflects normal hydrological conditions, a satellite scene captured before the flood, preferably during the same season [10]. Fernando Pech-May et al. [11] introduced the U-Net architecture to segment and visualize flooded areas using Sentinel-1 SAR satellite imagery. The introduced U-Net includes preprocessing steps, such as radiometric calibration, terrain correction, and normalization of polarization bands, to enhance image clarity. To identify flooded zones, an encoder-decoder structure and skip connections that allow the capture of both low-level and high-level spatial features are designed. The segmentation quality and modified accuracy for multiple epochs measured by training using binary cross-entropy loss and the Adam optimizer. However, the U-Net-based segmentation method demonstrates strong performance in identifying flooded areas that still face challenges, as segmentation is binary, which restricts its ability to capture advanced flood

* Corresponding author: smitha.nikku@gmail.com

characteristics, such as depth or partial flooding. Max Bereczky et al. [12] used U-Net alternates, particularly AlbuNet-34, which was integrated with the ResNet-34 encoder and used summation-based skip connections to reduce the model complexity and inference time. AlbuNet-34 architecture improvements include atrous spatial pyramid pooling in DeepLabV3+ and multiscale feature fusion in U-Net++ to improve segmentation in challenging environments. Radiometric data augmentation techniques consisting of intensity shifts and speckle noise simulations were used to enhance the simplification. A weighted combination of cross-entropy and Lovasz loss functions is used to optimize both the pixel-level classification and region-level overlap. However, the method used did not incorporate multitemporal sequences, which limits its ability to capture flood evolution or to distinguish between temporary water bodies and stable features. Abbas Mohammed Noori et al. [13] introduced a hybrid DL architecture, the Compact Convolutional Tokenizer-based model (CCT-U-ViT) is designed for bi-temporal flood detection using Synthetic Aperture Radar (SAR) images. It combines the precise spatial segmentation capabilities of a U-Net encoder–decoder framework with the broad contextual understanding offered by a Vision Transformer (ViT). The integration is further enhanced by a compact convolutional token that refines the input representation before processing. The introduced architecture operates on both pre- and post-flood SAR images through parallel branches, one dedicated to extracting layered spatial features through convolutional operations and the other focused on modeling long-range dependencies using multiheaded self-attention mechanisms and positional encoding. This dual-path approach enables the model to effectively identify and classify flooded regions by leveraging both local details and global context. The featured images were fused and passed through a classification head to describe flooded regions with high accuracy. However, the introduced CCT-U-ViT struggled in heterogeneous environments, especially with spectral confusion in areas such as wetlands, urban zones, and regions with high soil moisture, where radar backscatter signatures closely resemble floodwater. Jagannath Jadhav et al. [14] used a Long Short-Term Memory (LSTM) neural network, a multistep process for flood detection using remote sensing images. Wiener filtering was applied to reduce noise and enhance image clarity while preserving structural edges, followed by segmentation using the Versatile Linear Edge Detection (VLED) algorithm to isolate key regions such as water bodies, land, and built structures based on pixel intensity and edge discontinuities. After segmentation, the structural regions were clustered and passed to the classification stage using LSTM integrated with its memory units to capture spatial and contextual patterns across image layers, enabling it to distinguish flooded areas from non-flooded zones. However, LSTM does not train on specific structural features extracted from Remote Sensing Images (RSI), which is not generalizable to diverse geographic regions with variable land, urban layouts, or vegetation cover. Xuzhong Yan et al. [15] introduced a computer vision-based Multi-Object Tracking (MOT) framework for intelligent flood section understanding in dynamic video streams. The MOT method uses real-time video data to monitor, detect, and track multiple objects such as buildings, vehicles, boats, and people during flood events. Three major methods were integrated: a segmentation model for object detection, an optical-flow-based method to estimate trajectories of undetected objects, and a lightweight Convolutional Neural Network (CNN) to recover objects lost due to long-term occlusions. By combining spatial and temporal features, MOT confirms robust and continuous tracking, even in complex flood environments with nonlinear motion and visual obstructions. However, classification based on binary results, such as flooded versus non-flooded without considering flood severity, depth, or partial flood, restricts its utility for detailed disaster assessment.

The contribution of the proposed UConvFloodNet is as follows:

- The proposed UConvFloodNet handles spatial and temporal complexity better, making it suitable for various areas, such as urban, rural, vegetated, and mountainous regions.
- The Wiener filtering method, which effectively removes salt-and-pepper noise while preserving important edge details, is used to enhance remote sensing images.
- UNet++ was applied for precise image segmentation, allowing accurate identification of flood zones, water bodies, and land areas.
- The proposed method, UConvFloodNet, is designed to work effortlessly from preprocessing to classification, allowing real-time flood monitoring during emergencies.

The following Section 2 provides information regarding the problem statement, which is solved in Section 3 of the proposed method. Section 4 and 5 discuss the experimental results and conclusions of this paper.

2 Problem Statement

Floods are becoming increasingly frequent and destructive, and current remote sensing methods struggle to deliver appropriate and accurate flood detection due to limitations in image quality, spectral resolution, and temporal tracking. Traditional segmentation models fail to capture the developing nature of floods, particularly in complex lands and under noisy conditions. Additionally, existing methods depend on single-frame analysis, which lacks the ability to monitor the flood progression over time. To address these challenges, a robust real-time flood monitoring model that integrates noise reduction, precise segmentation, and temporal classification is required to accurately detect and track flood events in diverse global regions.

3 Proposed Method

The proposed UConvFloodNet model is developed for real-time flood monitoring by combining image-

enhancement techniques with an advanced neural network architecture. The core of the proposed method is a two-phase process that begins with the preparation of satellite images and then feeds them into a specialized DL model for segmentation and temporal tracking. The architecture integrates U-Net++ for spatial segmentation and a Convolutional Long Short-Term Memory (ConvLSTM) network to analyze changes in flood areas over time, as shown in Figure 1. This dual-component design ensures both the accurate identification of flood zones and dynamic monitoring of their evolution, addressing the restrictions of previous methods that lacked depth in spatiotemporal analysis.

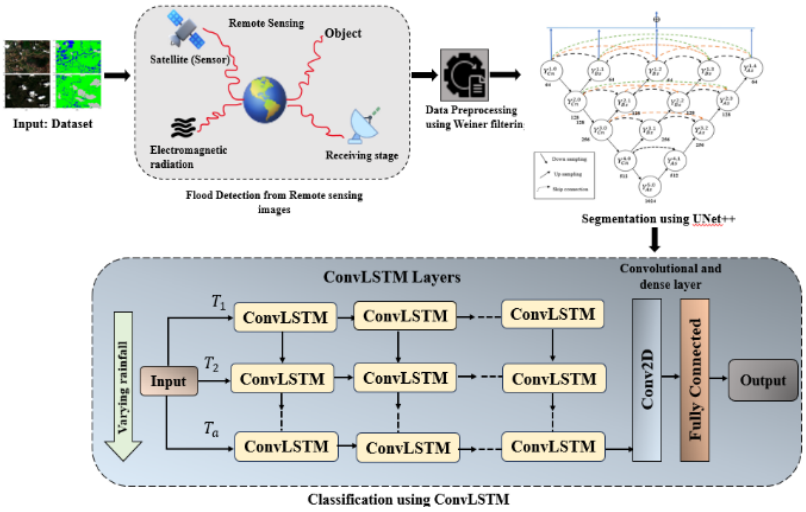

Fig. 1. Proposed architecture for flood monitoring.

3.1 Dataset Description

The performance of the UConvFloodNet model is evaluated using two publicly available datasets specifically for flood detection: Sen1Floods11[16] and S1GFloods [17]. These datasets are essential for training and testing models that utilize remote-sensing data to identify flood events and their features. The Sen1Floods11 and S1GFloods datasets serve as benchmark sources for evaluating flood detection models, each offering distinct characteristics. Sen1Floods11 combine data from Sentinel-1 SAR and Sentinel-2 optical satellites, whereas S1GFloods depend only on Sentinel-1 SAR images. Both datasets are provided by the European Space Agency (ESA) and tailored for remote sensing applications. Sen1Floods11 include rebuilt Ground Range Detected data enriched with metainformation from SAR and optical sources, whereas S1GFloods consist of SAR radar images formatted in RGB channels. Although the image size for Sen1Floods11 is unspecified, S1GFloods offer standardized inputs of 256×256 pixels. Sen1Floods11 capture a wide range of flood events from multiple studies with global geographic coverage, whereas S1GFloods focus on specific scenarios, including intense rainfall events, overflow of rivers, structural breaches in dams or levees, and extreme weather systems such as cyclones and tropical storms, spanning rural, mountainous, urban, and vegetated zones. Temporally, Sen1Floods11 provide pre- and post-event imagery, while S1GFloods include previous floods, floods, and variations in labeled images. Both datasets follow a 90:10 training-to-testing split, supporting a robust model evaluation.

3.2 Image Preprocessing

Image processing plays an important role in enhancing both conventional and satellite image by reducing noise, correcting blur, and applying various filtering techniques. The preprocessing phase typically involves two key components such as inverse filtering, which helps restore degraded image details, and noise smoothing which targets irregular brightness or color fluctuations especially common in Remote Sensing Images (RSI). Image scaling is also applied to adjust the pixel dimensions for consistent analysis. To address salt-and-pepper noise characterized by random black and white pixel disruptions specialized filtering methods are used to clean up the image and improve clarity.

3.2.1 Weiner Filter Method for noise reduction

The noise filter removes unwanted signals, but keeps the edges sharp. After basic processing, the image is divided into color channels and encoded. The middle channel handles bright areas using how the pixels are related to each other. The pixels are sorted in a small window to clean the images. Because important parts of the image have a steady brightness compared to the background, the filter can effectively remove noise.

3.2.2 Algorithm step for Weiner Filter

Step 1: Input Remote Sensing Images (RSI)

$$S_{min} = lowest\ pixel\ intensity\ value\ in\ A_{u,v}$$

Step 2: Initially, the kernel median value is calculated, and the pixel values are examined to identify and match the characteristics of salt-and-pepper noise.

$$S_{max} = highest\ pixel\ intensity\ value\ in\ A_{u,v}$$

$$Z_{a,b} = [\frac{M \times (a,b)}{|\ M \times (a,b)|^2 + C}]\ L(a,b)$$

Evaluate $Z_{(a,b)}$, Calculate the degraded image $L_{(a,b)}$, Estimate the degraded function $M_{(a,b)}$ and $M \times (a,b)\ is\ complex\ conjugate\ of\ M_{(a,b)}\ and\ constant\ is\ C$.

Step 3: Noise suppression is performed previous to feature extraction and image analysis.

Step 4: Output of preprocessing image.

3.3 Image Segmentation in RSI

Image segmentation refers to the process of dividing a remote sensing image (RSI) into separate regions to isolate meaningful features. In the context of flood analysis, this technique helps to differentiate between water bodies, land surfaces, and built structures. Segmentation is typically caused by variations in color or changes in grey-level intensity, allowing the detection of edges and boundaries within the image. The discontinuities serve as signs for separating regions, enabling precise identification of flood-affected areas. Edge and line detection is important in identifying structural features and different landmarks within RSI. Once these features are extracted, the image is partitioned into regions based on pixel similarity. Clustering methods are then applied to group pixels that share common characteristics, such as intensity or texture. To enhance segmentation accuracy, techniques like thresholding, region growing, and region splitting and merging are employed. These methods help to isolate areas such as flood zones or land surfaces by analyzing spatial continuity and contrast variations.

3.3.1 UNet++

UNet++ builds on the original UNet; an encoder–decoder architecture is first supported for medical image segmentation and has been adopted in areas such as remote sensing. In the classic UNet design, each decoder stage only merges feature maps of the same spatial size through direct skip connections; therefore, the information at encoder level i feeds only decoder level i. This keeps the segmentation map uniform, which leads to the blurring of fine details and weakening of boundary localization. UNet++ addresses this by inserting nested, densely interconnected skip paths: outputs from multiple encoder depths are progressively merged and refined before reaching each decoder stage, allowing richer, multi-scale feature fusion and more precise delineation of target regions. Equations (1) to (4) are used in the merging process. The UNet++ model consists of multiple intermediate convolutional blocks between the encoder and decoder stages, enabling multi-scale feature fusion. The developed model is designed to capture fine-grained flood boundaries, especially in heterogeneous terrains where water features blend with urban or vegetative backgrounds. By leveraging deep supervision, U-Net++ facilitated faster convergence and improved generalization across various flood scenarios.

$$Y_{As}^{1.4} = Y_{Bs}^{1.3} + Y_{Bs}^{1.2} + Y_{Bs}^{1.1} + Y_{Cn}^{1.0} + Y_{As}^{2.3} \quad (1)$$

$$Y_{Bs}^{1.3} = Y_{Bs}^{1.2} + Y_{Bs}^{1.1} + Y_{Cn}^{1.0} + Y_{Bs}^{2.2} \quad (2)$$

$$Y_{Bs}^{1.2} = Y_{Bs}^{1.1} + Y_{Cn}^{1.0} + Y_{Bs}^{2.1} \quad (3)$$

$$Y_{Bs}^{1.1} = Y_{Cn}^{1.0} + Y_{Cn}^{2.0} \quad (4)$$

3.4 Image Classification using ConvLSTM

Image classification includes organizing visual data into expressive categories based on consistent patterns across different layers and spectral channels. In remote sensing, grouping pixels with similar intensity values to identify specific features within an image. Supervised classification methods depend on training datasets known as training sites to teach the model how to differentiate between classes, such as flooded and non-flooded areas, with high accuracy. Using Python-based techniques, flood detection is performed on RSI by applying classification algorithms that learn from labeled data. These methods analyze clustered image segments to identify affected regions and collect them into a structured dataset. One efficient method is the ConvLSTM model, which processes image sequences across multiple layers. It scans the entire image once, enabling rapid and precise identification of flood zones through a dedicated classification layer. Each image is labelled with geographic coordinates, including latitude and longitude, to support in identification. The data processing workflow involves two key stages: testing and optimization. A portion of the dataset is reserved for testing typically around half to evaluate the system's performance. The model used is a mobile-integrated composite model that first preprocesses the input data before feeding it into the classification pipeline. During preprocessing, the data is divided into segments of size ten, and each segment is further split into three distinct components. This structured method allows the function to generate machine-readable image data, enabling efficient analysis and classification.

3.4.1 ConvLSTM

The ConvLSTM model is a DL algorithm designed for predicting spatiotemporal sequences that enhances the traditional LSTM architecture by incorporating convolutional operations, allowing it to effectively capture both spatial and temporal patterns. ConvLSTM achieves superior performance in multiple fields because of its strong ability to capture and learn spatiotemporal features. Traditional LSTM uses feed-forward neural network computations between gates, making it suitable for handling time-series data. However, it is difficult to represent 3D time series data. To address this, ConvLSTM introduces convolutional operations into the LSTM structure, allowing it to capture spatial features and learn spatial patterns from

time series datasets. When the input is a 3D image, ConvLSTM can effectively extract the relevant features from it. The ConvLSTM model replaces the feed-forward computations between gates in traditional LSTM with convolution operations. Unlike standard LSTM, which processes one-dimensional inputs, ConvLSTM applies convolution instead of matrix multiplication at each gate, allowing it to handle 3D input data effectively. The states of the gates (input, forget, and output) in the ConvLSTM model at each moment are computed as in Equations (5) to (9) where a_t is the input gate, e_t is the forget gate, D_t is the state at present moment, o_t is output gate.D_{t-1} represents the state of prior moment, g_t represents the final output, W is the weight coefficients, c represents the corresponding bias coefficient,σ is the sigmoid function,$\circ$ is the Hadamard product and, $\times$ is the convolution operation. In UConvFloodNet, ConvLSTM plays an important role in modeling the temporal evolution of flood patterns, enabling the network to learn how water bodies expand, shift, or recede across consecutive frames. By integrating memory cells with spatial filters, ConvLSTM allows the model to retain the context from previous observations while refining current predictions. temporal awareness enhances segmentation accuracy in dynamic flood scenarios, particularly when dealing with bi-temporal or multi-temporal SAR data.

$$a_t = \sigma(K_{ua} \times u_t + K_{ga} \times G_{t-1} + K_{da} \circ D_{t-1} + c_a) \quad (5)$$

$$\mathrm{e_t} = \sigma(\mathrm{K_{ue}} \times \mathrm{u_t} + \mathrm{K_{ge}} \times \mathrm{G_{t-1}} + \mathrm{K_{de}} \circ \mathrm{D_{t-1}} + \mathrm{c_e}) \quad (6)$$

$$\mathrm{D_t} = \mathrm{e_t} \circ \mathrm{G_{t-1}} + \mathrm{a_t} \circ \tan h(\mathrm{K_{ud}} \times \mathrm{u_t} + \mathrm{K_{gd}} \times \mathrm{G_{t-1}} + \mathrm{c_d}) \quad (7)$$

$$p_t = \sigma(K_{up} \times u_t + K_{gp} \times G_{t-1} + K_{dp} \circ D_t + c_p) \quad (8)$$

$$g_t = p_t \circ \ tan\ h(D_t) \quad (9)$$

4 Experimental Results

The proposed UConvFloodNet method is implemented using Python in the Spyder environment, where RSI data consistently provide richer and more detailed information than standard flood images, which helps to enhance the overall quality and importance of the results. The evaluation depends on the specific application and simulation parameters, using tools such as Spyder, Anaconda 3, and Python as the programming language, all of which run on an Intel Core i5 processor.

4.1 Evaluation Metrics

Accuracy is the ratio of correctly identified pixels (both flood and non-flood) to the total number of pixels in an image. The F1 Score is the harmonic mean of the precision and recall, making it useful for evaluating models on imbalanced datasets. The intersection over union (IoU) measures the overlap between the predicted and actual ground-truth flood areas. A higher IoU value indicates better alignment between the predicted and actual flood regions. The Kappa Coefficient (Cohen's kappa) measures the intercoder reliability of a model. Equations (10) to (13) show the formulae of each metric calculation, where TP is true positive, TN is true negative, FP is false positive, and FN is false negative. P_o is the observed agreement (overall accuracy) and P_e is the expected agreement by chance.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

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

$$IoU = \frac{TP}{TP+FP+FN} \quad (12)$$

$$Kappa = \frac{P_o - P_e}{1-P_e} \quad (13)$$

4.2 Performance Evaluation

To evaluate the effectiveness of the proposed UConvFloodNet architecture, present a detailed performance comparison with the existing state-of-the-art models. The evaluation s conducted using two benchmark datasets, Sen1Floods11 and S1GFloods, which were selected for their diversity in flood scenarios and geographic coverage. Tables 1 and 2 summarize key metrics, such as Accuracy, F1-score, Kappa coefficient, and IoU, presenting a complete view of the model segmentation and temporal tracking capabilities.

Table 1. Performance measurement for state of art methods using Sen1Floods11 dataset.

Dataset	Models	Accuracy	F1 score	Kappa	IoU
Sen1Floods11	InceptionV3	95.14	92.34	80.71	81.64
	ResNet-50	96.47	95.71	81.78	76.78
	Fully Convolutional Network	93.87	92.47	71.34	80.74
	UConvFloodNet (proposed)	99.31	97.60	85.07	84.98

Table 2. Performance measurement for state of art methods using S1GFloods dataset.

Dataset	Models	Accuracy	F1 score	Kappa	IoU
S1GFloods	InceptionV3	90.78	89.58	81.13	78.64
	ResNet-50	96.45	88.56	76.87	73.45
	Fully Convolutional Network	95.21	90.47	73.48	82.71
	UConvFloodNet (proposed)	99.06	98.27	86.17	86.62

4.3 Complexity and Statistical Analysis

The proposed UConvFloodNet model demonstrates efficient computational performance with a time complexity of approximately $O(n.d)^2$, where n is the number of input frames and d is the spatial resolution. Memory usage remained within practical bounds because of the modular integration of ConvLSTM and U-Net++ components, averaging 2.3 GB during inference on 512×512 inputs. To validate the statistical significance, a Wilcoxon signed-rank test was conducted to compare IoU scores across baseline models and UConvFloodNet. The results showed $p < 0.01$, confirming that the improvements are not due to random variation. The findings highlight the suitability of the models for real-time flood monitoring in resource-constrained systems, balancing accuracy with computational efficiency.

4.4 Ablation Study

To understand the individual contributions of each component within the proposed UConvFloodNet architecture, an ablation study is conducted, as shown in Figures 2 and 3. By selectively removing or modifying each module, such as the ConvLSTM layer, U-Net++ backbone, and preprocessing enhancements (Wiener filtering and spatial contrast adjustment), the impact on overall segmentation accuracy and temporal consistency. The overall performance of the combined models is better than that of each individual model, which represents the contribution of the proposed method.

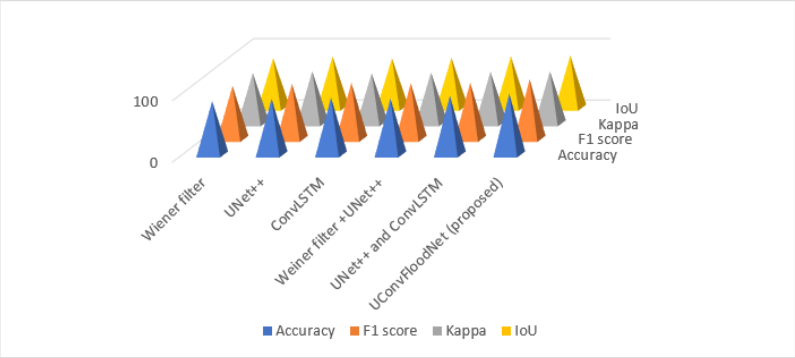

Fig. 2. Graph showing component wise performance for proposed using Sen1Floods11 dataset.

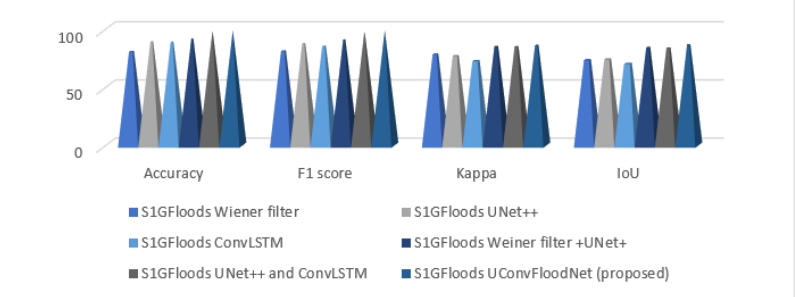

Fig. 3. Graph showing component wise performance for proposed using S1GFloods dataset.

4.5 Comparison Study

To analyze the performance of the proposed UConvFloodNet, a comparative evaluation of established flood-detection architectures, including U-Net, DeepLabV3+, AlbuNet-34, and CCT-U-ViT performed. The comparison is based on the key segmentation metrics Accuracy, F1-score, Kappa coefficient, and IoU for two benchmark datasets: Sen1Floods11 and S1GFloods, as shown in Table 3. The proposed method, UConvFloodNet, provides a superior value compared to other existing models, which shows the robustness of the model.

Table 3. Comparing the proposed method with existing model.

Dataset	Models	Accuracy	F1 score	Kappa	IoU
Sen1Floods11	U-Net [11]	94.31	84.41	NA	73.02
	AlbuNet-34+FCN+DeepLabV3+ U-Net+ U-Net++ [12]	98.9	82.3	81.6	74.0
S1GFloods	CCT-U-ViT [13]	91.24	91.21	82.48	83.83
Sen1Floods11	UConvFloodNet (proposed)	99.31	97.60	85.07	84.98
S1GFloods	UConvFloodNet (proposed)	99.06	98.27	86.17	86.62

4.6 Discussion

The proposed UConvFloodNet model shows advantages over traditional flood detection methods, both in terms of how it segments flooded areas and tracks changes over time. By combining U-Net++ with ConvLSTM, the proposed architecture captures detailed flood boundaries while adapting to evolving conditions in the satellite images. When compared to the established CNN and Transformer-based models, UConvFloodNet consistently produces better, more interpretable results across various flood scenarios. The ablation study highlights the importance of each component in removing temporal or preprocessing modules that noticeably weaken performance. Visualization outputs further support the reliability of the model, showing that it consistently focuses on flood-relevant regions. Despite its complexity, the proposed model remains sufficiently efficient for real-time use, retaining a practical balance between accuracy and resource demand. Overall, UConvFloodNet provides a strong and scalable foundation for flood monitoring and can be extended to other geospatial or disaster-related applications.

5 Conclusion

Th presents UConvFloodNet, a hybrid DL model that integrates U-Net++ and ConvLSTM for effective flood detection using satellite images. The model addresses both spatial segmentation and temporal tracking challenges by applying conventional CNN and Transformer-based architectures across diverse flood situations. A complete comparative analysis and ablation studies showed that each component of the architecture contributed to the overall performance. Visual interpretability confirms that the model consistently focuses on flood-relevant regions, thereby enhancing its reliability for real-world deployment. Importantly, UConvFloodNet maintains its computational efficiency, making it suitable for real-time applications in resource-constrained environments. Adaptability to variable geographies and flood pattern positions is a scalable solution for operational flood monitoring. Beyond flood detection, the architecture has the capacity for broader geospatial tasks, such as disaster response, land cover change analysis, and environmental monitoring. Future work may explore the integration of IoT systems and multimodal data sources to further enhance responsiveness and decision support in emergency scenarios.

References

1. B.B. Nair, P. Vallimeena, U. Gopalakrishnan, S.N. Rao, S. Krishnamoorthy, Enhanced urban flood monitoring: Integrating advanced semantic segmentation and human facial feature and posture analysis. IEEE Access. **12**, 185807-185825 (2024). https://doi.org/10.1109/ACCESS.2024.3513239
2. M. Kögel, L. Feile, F. Möldner, D. Carstensen, SSegRef2Surf—Near real-time photogrammetric flood monitoring and refinement of classified water surfaces. Remote Sens. **17**, 1351 (2025). https://doi.org/10.3390/rs17081351
3. Y. Wang, Y. Shen, B. Salahshour, M. Cetin, K. Iftekharuddin, N. Tahvildari, G. Huang, D.K. Harris, K. Ampofo, J.L. Goodall, Urban flood extent segmentation and evaluation from real-world surveillance camera images using deep convolutional neural network. Environ. Model. Softw. **173**, 105939 (2024). https://doi.org/10.1016/j.envsoft.2023.105939
4. P. Dang, J. Zhu, C. Dang, H. Zhang, A method for assessing urban flood impacts by combining hydrodynamic models with entity semantic evolution. Int. J. Digit. Earth **18**, 2515268 (2025). https://doi.org/10.1080/17538947.2025.2515268
5. B.J. Jang, I. Jung, Development of high-precision urban flood-monitoring technology for sustainable smart cities. Sensors **23**, 9167 (2023). https://doi.org/10.3390/s23229167
6. L. Hashemi-Beni, M. Puthenparampil, A. Jamali, A low-cost IoT-based deep learning method of water gauge measurement for flood monitoring. Geomatics Nat. Hazards Risk **15**, 2364777 (2024). https://doi.org/10.1080/19475705.2024.2364777
7. G. Savitha, S. Girisha, P. Sughosh, D.K. Shetty, J.M. Balakrishnan, R. Paul, N. Naik, Consistency regularization for semi-supervised semantic segmentation of flood regions from SAR images. IEEE Access. **13**, 9642-9653 (2025). https://doi.org/10.1109/ACCESS.2025.3526244
8. N. Humaira, V.S. Samadi, N.C. Hubig, DX-FloodLine: End-to-end deep explainable pipeline for real-time flood scene object detection from multimedia images. IEEE Access **11**, 110644-110655 (2023). https://doi.org/10.1109/ACCESS.2023.3321312
9. C. Krullikowski, C. Chow, M. Wieland, S. Martinis, B. Bauer-Marschallinger, F. Roth, P. Matgen, M. Chini, R. Hostache, Y. Li, P. Salamon, Estimating ensemble likelihoods for the Sentinel-1 based Global Flood Monitoring product of the Copernicus Emergency Management Service. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **16**, 6917-6930 (2023). https://doi.org/10.1109/JSTARS.2023.3292350
10. F. Fichtner, N. Mandery, M. Wieland, S. Groth, S. Martinis, T. Riedlinger, Time-series analysis of Sentinel-1/2 data for flood detection using a discrete global grid system and seasonal decomposition. Int. J. Appl. Earth Obs. Geoinf. **119**, 103329 (2023). https://doi.org/10.1016/j.jag.2023.103329
11. F. Pech-May, R. Aquino-Santos, O. Álvarez-Cárdenas, J.L. Arandia, G. Rios-Toledo, Segmentation and visualization of flooded areas through Sentinel-1 images and U-net. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **17**, 8996-9008

(2024).
https://doi.org/10.1109/JSTARS.2024.3387452

12. M. Bereczky, M. Wieland, C. Krullikowski, S. Martinis, S. Plank, Sentinel-1-based water and flood mapping: Benchmarking convolutional neural networks against an operational rule-based processing chain. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **15**, 2023-2036 (2022). https://doi.org/10.1109/JSTARS.2022.3152127
13. A.M. Noori, A.R.T. Ziboon, A.N. Al-Hameedawi, Deep-learning integration of CNN–Transformer and U-net for bi-temporal SAR flash-flood detection. Appl. Sci. **15**, 7770 (2025). https://doi.org/10.3390/app15147770
14. A. Sonavale, M. Chakkaravarthy, S. Srinivasa Rao, H.B.M. Salleh, J. Jadhav, Automatic semantic segmentation and classification of remote sensing image data for flood detection using novel LSTM neural network. SN Comput. Sci. **5**, 992 (2024). https://doi.org/10.1007/s42979-024-03336-9
15. X. Yan, Y. Zhu, Z. Wang, B. Xu, L. He, R. Xia, Intelligent flood scene understanding using computer vision-based multi-object tracking. Water **17**, 2111 (2025). https://doi.org/10.3390/w17142111
16. Dataset 1:SEN1FLOODS11-essentials
17. Dataset 2: S1GFloods_ALL