Identification of Post Flood Water Level Severity through UAV Images Using Attention Based Deep Learning Techniques

Sanket S Kulkarni[1,*], Ansuman Mahapatra[1]

[1]Department of Computer Science and Engineering, National Institute of Technology Puducherry Karaikal, Puducherry U.T 609609, Puducherry, India

Emails: sanketskulkarni95@gmail.com; helloansuman@gmail.com

Abstract

Floods are among the most devastating natural disasters, causing widespread damage to infrastructure, homes, and human lives. Rapid assessment of flood severity is critical for effective disaster response and resource allocation. This study explores several deep learning approaches for flood water level classification using UAV imagery. A curated dataset of 2,000 UAV images from diverse regions, including India, the United States, and Brazil, was developed and augmented to improve generalization. Multiple architectures were evaluated, including pre-trained CNNs, ResNet50v2, MobileNetv2, Vision Transformers, and Swin Transformers, with and without the Convolutional Block Attention Module (CBAM) and adaptive learning strategies. Experimental results reveal that integrating Vision Transformers with CBAM achieves the highest classification accuracy of 90.6%, while a hybrid CNN–Vision Transformer model further improves performance to 92.3%. These findings highlight the potential of attention-based hybrid models for precise flood severity mapping. The proposed framework can aid rescue teams and disaster management authorities by prioritizing high-risk areas, enabling faster response and optimized allocation of resources during emergency operations.

Keywords: Flood water level; Pre-trained CNN; CBAM; Vision Transformer; Swin transformer; Hybrid vision transformer

1. Introduction

Floods have a severe impact on individuals and communities. It also results in impacts to society. Floods are extremely devastating to individuals and to communities. Society is also affected by it. The level determination of the floodwater based on images of unmanned aerial vehicle (UAV) is a crucial component of disaster response and preparedness. To facilitate in rescue efforts of people, evaluation of property damages and implementation of mitigation measures, post- and peri flood water levels may be of critical importance to give valuable information on the magnitude and scope of flooding. Floods have become a serious issue in most areas of earth, more especially in most parts of Asia, ranging to Ceylon, Afghanistan, Bangladesh, India as well as Pakistan [1].

The emerging approach in this field is floodwater level image classification, which uses deep learning models to analyze UAV (Unmanned Aerial Vehicle) images to estimate post-flood water levels. Nearly all individuals utilize pre-trained CNN models for classifying images of floodwater levels. Further, the recent transformer approaches, including Vision and Swin transformers, have yet to be explored. Top and bottom images in Figure 1 show the high and low flood levels, respectively.

DOI: https://doi.org/10.54216/FPA.210221
Received: April 18, 2025 Revised: July 24, 2025 Accepted: September 02, 2025

Figure 1. Extreme floodwater (on the top) and low floodwater level images (on the bottom)

Flood is one of the most catastrophic natural calamities that damage property, cause deaths as well as cause displacement of communities. The conventional method of measuring water level, e.g. manual inspection or terrestrially based sensors, are time-consuming and the readings obtained may not be accurate during serious floods. One of the traditional methods with which the flood prone areas are managed include visual inspection and marker-based systems, that is, the use of water level markers like poles that have markings defining the level of water. Flood estimates are obtained using remote sensing, stream gauging, and photo comparisons of flood surpassing these markers with water level. Other data collected via remote sensing with the help of a satellite will be unavailable in real time of the catastrophe since revisit time of a satellite is usually too large [35]. This can be one of the disadvantages of some of these techniques but they are possibly affordable. The process of collecting data in flooded areas is quite risky and is highly expensive in cases where it is done manually because a survey of the flooded area has to be done. With an appropriate and prompt measurement of flood and water levels, it would be possible to respond to the disaster and recovery and use the resources efficiently. The goal of this project is to build a classification model, which would be based on the newest techniques in computer vision in order to identify the areas affected by flooding and classify them according to the level of water.

Estimating the state of various places after a flood has occurred is known as post-flood assessment. Flood rescue operations, mapping flood zones, and flood monitoring are some of the major benefits of post-flood assessment [36]. For some viewable ranges, it is feasible to determine if the water level locations are in mild or intense flooding regions.

The key advances of the work are the following:

1. Using unmanned aerial vehicle (UAV) photos create a dataset for estimating flood water levels.
2. Employing CBAM layer to enhance the performance of vision transformers and pre-trained CNN models.
3. Improvement of accuracy through the integration of the most effective pre-trained CNN model and vision transformer.

2. Related Works

2.1 Studies towards estimating flood water levels with images from satellite Imagery

The result of flood zones on satellite images is realized in digital elevation models (DEM) and the most outstanding ones use Synthetic Aperture Radar (SAR) satellite data, as Sentinel-1, estimating water depths based on the amount of elevation information along with flooding extent. The available papers on estimating the water level in case of a flood are more inclined to the techniques of machine learning and deep learning approaches in order to give the estimation. One of the works by Andrea Betterle and Peter Salamon (2024) introduced a methodology that considers enhancing satellite-based inundation maps and approximating depths of floodwaters and sharpened flood boundaries to increase the accuracy of available satellite-based inundation maps. This model will make use of readily available topographical information to increase accuracy in determining the extent of the flood [2].

DOI: https://doi.org/10.54216/FPA.210221
Received: April 18, 2025 Revised: July 24, 2025 Accepted: September 02, 2025

The paper article on the special deep learning model of flood forecasting with the use of satellite imagery is Stateczny et al. (2023), where there was a special deep hybrid model of flood prediction (DHMFP) and combined Harris hawk shuffled shepherd optimization (CHHSSO) based training algorithm of flood prediction [12]. Following this help through application of median filtering process, the input image is taken into the pre-processing. Here the filtering of the image then follows this; later an image segmentation is performed on which the use of a weighted cubic chaotic map is founded on basis of k-means clustering. The model is based on the design of deep ResNet classification and CNN. The hybrid methods are not competitive as when they are combined error rates of the calculations are so small and high accuracy of deep neural network also, increases as the weights finally get the right degree of adjustments and hence a lot of effectiveness has been provided to it.

Cuong Le et al. (2023) offer the FL-Former technology of estimating the extent of flood, the technology that applies the Vision Transformer to estimate the measure of flood based on the camera images in the urban environment. The alternative can be solved by detecting and locating the level of rain and inundation, namely urban towns, i.e., Ho Chi Minh City resorted to by taking image with the help of cameras to ascertain the extent of inundation [13]. They offered to make the online API system and software to supply the citizens with up-to-date data on the flooding and precipitation in various locations of the city that was disclosed as the supplementary data to the models of hydrometeorological forecasts and analysis and was organized in web and mobile apps.

Anusha N and Bharathi B (2020) examined the floods that occurred in the Uttar Pradesh region of India in August 2017 using optical and great quality multi-temporal Synthetic Aperture Radar (SAR) pictures. To figure out which are the flooded districts in the selected Area of Study (AQS), this paper calculates the zonal statistics. [14]. This mapping is performed using district floods using extracted water level as a superimposition. Meteorological measurements are used to guarantee correctness of the outcomes. Consequently, the flood disaster management can be highly improved with the identified flood prone areas, which may in turn be utilized as important input in the flood modeling and analysis.

In three stages, Tanaka et al. (2019) identified flood areas and clumps for each flooded zone to estimate flood water levels in the Mekong floodplain [15]. Buffering of the identified flood locations to extract the edges of the same and spatial transformation of flood water level elevation by executing interpolation of the extracted data. They then compared the estimated levels of water and actual.

Table 1: Existing works on satellite flood water level identification

Study	**Dataset**	**Methodology**	**Key findings**
Chamatadis *et al.* (2024) [3]	Sentinel-1 and Sentinel-2	Transfer learning and Vision transformer	Vision transformer is used in detecting flooding in satellite image with low performance
Wedjao *et al.* (2024) [5]	Sentenel-2	SVM, Random forest, LSTM and Linear Regression	High accuracy in predicting flood-prone areas with effective susceptebility mapping
Zhouyayan Li and Ibhrahim Demeyer (2023) [6]	Sentenel-1	Edge OTSU, Bmax OTSU and Fuzzy OTSU	This work differenciates permanent water and flood pixels trained using pre-trained weights from coarse dataset
Ismail elkrachy (2022) [7]	Sentenel-1, Sentenel -2 and digital surface models (DSM)	Used machine learning (ML), regression algorithms	The RMSE accuracy for all ML algorithms was between 0.18-0.22 m for depth less than 1m
Hossein Hoseiny (2021) [8]	SAR dataset	Used Unet advanced CNN for flood depth estimation	The performance of model is limited to river geometry and flood extent
Bonafilia *et al.* (2020) [9]	Sen1Flood11	Fully Convolutional neural networks (FCNN) to segment permanent flood water	This work focused on surface water detection of floods

DOI: https://doi.org/10.54216/FPA.210221
Received: April 18, 2025 Revised: July 24, 2025 Accepted: September 02, 2025

Cohen *et al.* (2018) [10]	Digital elevation model (DEM)	Developed a floodwater depth estimation tool	This work is limited to specific region where FWDET calculates accuracy in depth for diverse flood scenarios
Cian *et al.* (2018) [11]	LIDAR and DEM	Statisitical analysis for depth estimation	These methods are fast and robust compared to hydrodynamic models

The significant drawbacks of estimating floodwater depth from satellite images face several challenges. Firstly, this image often needs more spatial resolution to accurately capture water depth variations, especially in urban areas with complex structures. Additionally, climatic factors like haze and clouds might obfuscate flood details in the photos. Water depth estimation becomes more complicated due to shadows from buildings, trees, and other objects misinterpreted as water. Additionally, with ground reference data, it is easier to directly measure water depth from images, as satellite sensors primarily capture surface information. Measurements or advanced modelling techniques, which can be complex and resource-intensive.

2.2 The work done on flooding water level calculating with UAV imageries

The methods used by the researchers are estimation of flood water levels through human observation of the water levels as well as use of traditional deep-learning methodology such as CNN to collect data. The image set on social media is the one that is used to forecast the floodwater. There are a few other works that also involve flood water level estimation based on human pose estimation, and some other works on flood water level estimation include street images (traffic signals) for estimating depth using the traffic signal boards' height.

Using the YOLOv4 object recognition model, Zhong et al. (2024) were able to identify submerged items in images, such as the legs of pedestrians and the exhaust pipes of vehicles. The results of their tests showed that the model's accuracy differed for different reference objects. It is worth mentioning that the YOLOv4 model had better accuracy when automobiles were utilized as reference objects rather than pedestrians. [16]. Mosaic technology, which augments images, also significantly improved recognition accuracy. Using pre-existing traffic camera photos or video data, the created technology is able to extract continuous, real-time water depth information.

According to Wan et al. (2024), the primary goal of this study is to present a technique for determining flood levels by identifying submerged vehicles in images. [17]. to elucidate on the Flood status of vehicles in this research paper, the flood status is categorized in five groups whereby each category corresponds to one of five intervals of urban flood levels. The collected and labelled data has two thousand photos representing six thousand three hundred vehicle items.

Deep-learning-based methodology incorporated dense optical-flow field representations calculated using images recorded by a standstill camera by the study of Ranieri et al. (2024). [18]. they had the following criteria to gauge level of water: absolute water level (how low, medium, high or flooding is the water) and relative water level (how high or low is the water). It was stated by the trials that the relative measurement of the water level and pairs of successive grayscale photographs were seriously encouraged by representations on the optical flow and actually gauged the absolute level of the water body.

The methodology that Wienhold et al. (2023) created was derived by relying on the application of the UAV technology, which is the Floodwater Inundation and Depth Mapper (FIDM). The process has three basis and they are; acquire aerial image, pre-process acquired image and seek inundation of flood mapping and inundation-depth mapping. This is the integration of such factors that give a detailed analysis and assessment of floods [19]. Here, a fully integrated (FI) model is proposed in which the UAV sensors construct an elevation dataset but a partially integrated (PI) model, which uses existing data such as LiDAR available publicly, is proposed here. They relate and discuss the advantages and disadvantages of full and partial models. Besides, the model generates flooding and depth that gives a result when compared with validation and ground truth data.

In a study by Popandopulo et al. (2023) analyzed the flood data using a pipeline of deep neural networks. Using this method, researchers may make educated guesses about the magnitude and breadth of floods. [20]. To guarantee the reliability and accuracy of the results of the flood monitoring they made use of the Digital Elevation Model

DOI: https://doi.org/10.54216/FPA.210221
Received: April 18, 2025 Revised: July 24, 2025 Accepted: September 02, 2025

(DEM). The volume of that flood is 0.0087 km^3, so we could estimate its scope. To assist affected regions in becoming responsive and faster in recovery, they proposed a practical school of thought on the way they can enhance the accuracy and speed of flood damage surveys.

Hafiz Salman Munwar et al. (2021) utilized CNN based flood detection and feature selection of landmark based [21]. Flood inundation mapping employed UAV images. This work is limited to flood-affected and flooded regions obtained by UAV images and cannot describe flood depth needed for assessing flood extent. The system's accuracy is improved by using RNN and LSTM deep learning methods for mapping and detection of flood inundation.

On the one hand, related to the flood level estimation in the images of news media, Julia Strebl et al. (2019) proposed MedEval2019. To establish the images of the people that were standing above the knee length in the water, they initially combined the human pose detector and the water detector [22]. The weakness of this technique in the process of estimating the flood depth is that this will not only manifest other objects in the water, but also low extremities too. This work is useful with pixel-wise segmentation of water and a human.

Amir H. Behzadan and Bahareh Alizadeh Kharazi (2021) developed a method in the depth of floodwater and the pole length estimation. They identify drowned stop signs in images of flooded intersections using deep neural network, canny edge detection and probabilistic Hough transform [24]. BluPix 2020.1: 10 regions of the FEMA in the United States: The findings web-mined the stop signs in the water and the dataset contained the linked images of pictures taken under the water. Using the mean squared error error (RMSE), the estimated RMSE values of the pole lengths in post-flood pictures were 17.43 inches and 8.61 inches in the pre-flood pictures to measure, the level of floodwater in Canada was 12.63 inches mean squared error error (MSE).

According to Muhadi et al., Deeplabv3+ and SegNet were compared regarding many performance metrics. [23]. The Deeplabv3+ rated higher in accuracy of 93 percent and Internet of Things (IoT) values of approximately 82 percent on boundary F1 (BF) score when compared to Segnet whose accuracy was 91 percent and 67.20 percent respectively. At the level of performance within specific classes as well, DeepLabv3+ outperformed SegNet due to its greater overall accuracy, BF and IoU score.

Table 2: Existing Works on Flood water depth estimation in UAV images

Study	**Dataset**	**Methodology**	**Key findings**
Lin *et al.* (2020) [28]	559 positional UAV images were used	Random forest, SVM, KNN	VGI quantification for flood water level estimation
Rizk *et al.* (2024) [29]	UAV image from news and website	Used RCNN for detecting house and cars	Developed a system to tackle wild datasets
Liang *et al.* (2023)[30]	UAV annotated videos and images from flooded areas	Novel automatic system for urban flood detection and quantification	Water depth estimation model on reference objects and template matching
Chaudary *et al.* (2019) [31]	UAV social media images	Locating selected class of objects with known sizes of objects	Recognition of objects at same time
J.L. Gan and W Zaliah (2021) [25]	300 images captured from flood scenes	Used CNN for classification	Water level classification to analyze risk
R.J. Pally and S.Samadi (2022) [33]	Flood custom dataset with 9000 images	Developed Flood image classifier to classify and detect objects	Calculates flood water level and innundation areas

DOI: https://doi.org/10.54216/FPA.210221
Received: April 18, 2025 Revised: July 24, 2025 Accepted: September 02, 2025

J. L. Gan and W. Zailah (2021) utilized CNN in their study to categorize the images and integrated a host of data preprocessing and enhancement methods to expand the data by adding more pictures [25]. AlexNet adjusted with a variety of optimizers and learning rates attained a precision of 99.72 percent using the Adam optimizer, a learning rate of 0.0001, and a group size of 16 on a limited dataset.

Feng *et al.* (2020) extracted and mapped a flood severity information by suggesting method of three-step process [26]. As part of the first stage, images related to flooding are accessed with the help of pre-trained CNN used as feature extractors. Again, according to the level of water relative to body parts like ankle, knee, hip and chest, the images can be classified into four severity grades with regard to the extent of partial inundation. Finally yet importantly, the last step will be the creation of the flood extent and severity map based on the latest tweet location data. It was tested with the observation of Hurricane Harvey in 2017 by measuring its accuracy in identifying the locations where flooding or water was experienced with more than 62 percent regions being accurately classified as flooded areas or victimized using water and its frequency of occurrence computed correctly.

Quan et al. (2020) suggested correlating water level, flood extent, and human position to assess flood intensity. This study predicts flood levels using online photo and post human pose data. [27]. The experiments are conducted on a Multimodal flood level estimation dataset, accomplishing 88.31% F1-Score.

In their development on several platforms, Lin et al. (2020), present the volunteered geographic information (VGI), which allows the populace to understand where and when anything happens simultaneously, and image-based VGI will illustrate environmental changes and disasters, including flooding areas and water levels comparisons [28]. The suggested methodology determined the level of water during a flood in a city of Taipei based on photogrammetric algorithm and digital elevation maps.

Chaudhary et al. (2019) used this feature to estimate the water level and included 2000 photos of extreme and low floods that were gathered from different sources. [31]. To train deep-learning model, they first created a set of images depicting floods. Lastly, showed the ability of our trained model to accurately estimate the flood levels and, at the same time recognize objects. This way to measure the significant steps, including locating the classes of objects, whose sizes are known approximately, which is the first step, was presented by the authors of this work. It uses this property in order to forecast the water level. They used the creation of the floodwater image dataset to check the effectiveness of this approach.

Through this literature review, it is necessary to note that object detection and floodwater segmentation methods are relevant to the practice of the floodwater level estimate in UAV imagery. A great number of things are considered when gauging flood water level including human poses, vehicles, among other things. However, the effectiveness of introducing pre-trained CNN models and transformers in ensuring better classification of the flood water-level image has not been delved into by many studies. Moreover, few researchers apply convolutional block attention modules. The appearance of water in images varies based on lighting, vegetation, and land cover, which can confuse algorithms. These factors challenge models to delineate water boundaries across various environments consistently. Image classification is used to estimate the levels of floodwater through establishment of the depths and boundaries of water after analysing aerial or ground level images. Techniques often use deep learning such as CNNs and ordinary DEM.

3. Methodology

This section presents the methodology adopted for classifying floodwater levels from UAV imagery. It includes the dataset description, preprocessing and augmentation strategies, and the design of architectures based on CNNs, Vision Transformers (ViT), Swin Transformers, and their integration with the Convolutional Block Attention Module (CBAM). The proposed hybrid model combining CNN and transformer features is also described.

3.1 Dataset Description

A dataset of **2,000 UAV flood images** was compiled from multiple sources, including reliable websites, YouTube clips, and aerial photographs from news channels. Images were collected from flood-affected regions in **India** (Odisha, Kerala, Tamil Nadu, Delhi, Karnataka) as well as from the **United States, United Arab Emirates, Australia, and Brazil**. The dataset spans the period **January 2023 – August 2024**, covering diverse post-flood scenarios.

DOI: https://doi.org/10.54216/FPA.210221
Received: April 18, 2025 Revised: July 24, 2025 Accepted: September 02, 2025

Table 3: Dataset Description

Total Images / **Ratio**	**Training Images**	**Validation Images**	**Test Images**	**Total Images**
Train (80%)	533	533	534	1600
Validation (10%)	67	67	66	200
Test (10%)	67	67	66	200
Total Images	667	667	666	2000

The UAV images reflect varying **altitudes (20–120 m)**, **pitch and yaw angles**, and scene complexities:

- Houses of varying shapes, sizes, and colors (small, large, occluded, overlapped).
- Both **continuous** and **fragmented** floodwater zones.
- High-detail close-range images (20–30 m) and wide-area lower-resolution images (80–120 m).

Table 4: Data Augmentation techniques

Augmentation technique	**Value of augmentation**
Image Rotation	±30°
Flipping	Horizontal
Brightness	±16%
Exposure	±20%
Gausian blur	σ=1.5 pixels
Noise	up to 3%

3.2 Convolutional Block Attention Module (CBAM)

CBAM improves classification by adaptively emphasizing **channel** and **spatial features** relevant to floodwater detection. For UAV images, this helps the model focus on boundaries, shallow flooding, and structural details, which may otherwise be ambiguous. CBAM applies sequential **channel attention** and **spatial attention** maps, refining feature maps by element-wise multiplication with the attention maps. This lightweight and general-purpose module can be integrated into CNNs without major architectural changes.

In this study, CBAM is applied to enhance pre-trained CNNs by focusing on **flood-related textures, water regions, and housing structures**, thereby improving accuracy and robustness.

For feed-forward CNN, the CBAM serves as an attention module. By successively producing attention maps in two different dimensions, one for the channels and another for the spatial locations, this module processes an intermediate feature map. The process of refining of the adaptive features then requires the multiplication of the attention maps with the input feature map. The lightweight, general-purpose CBAM module can easily be included in CNN designs that seems to involve implicit weights.

A combination of CNN and CBAM uses several convolutional layers, pooling, and activation functions to extract hierarchical features from input images. The standard CNN-models have been employing not only relevant

DOI: https://doi.org/10.54216/FPA.210221
Received: April 18, 2025 Revised: July 24, 2025 Accepted: September 02, 2025

variables but also irrelevant variables in classifying the images. In the present case, CBAM enhances the focus of the model on important aspects.

The use of UAV flood imagery is currently undertaken using this research with an attempt to collect the most pertinent channel and geographical information. The focus of the efforts is to include CBAM into a classification architecture that has already been trained using CNN. This work applies the CBAM module of the CNN model after specific convolutional layers. During the evaluation of channels according to their relevance to flood features, such as water level visibility, the flood water level classification primarily integrates CBAM into a pre-trained CNN architecture. This method allows for the classification of extreme and low floodwater locations by extracting the most relevant spatial and channel information from the flood photos.

CBAM's computational efficiency and ease of integration with current models to increase computational complexity are its primary features. The increased feature representation in CBAM for image classification helps the model better capture significant characteristics by concentrating on spatial regions and the most informative channels. The CBAM offers versatility because it may be efficiently included into current CNN architectures for classification tasks without requiring major modifications.

3.3 Pre-trained CNN model with CBAM

Pre-trained CNN models (ResNet50, ResNet101v2, VGG16, VGG19, Xception, DenseNet201, MobileNetv2, EfficientNet, and Inception-ResNet) were initialized with ImageNet weights. CBAM was integrated after specific convolutional layers to refine extracted features.

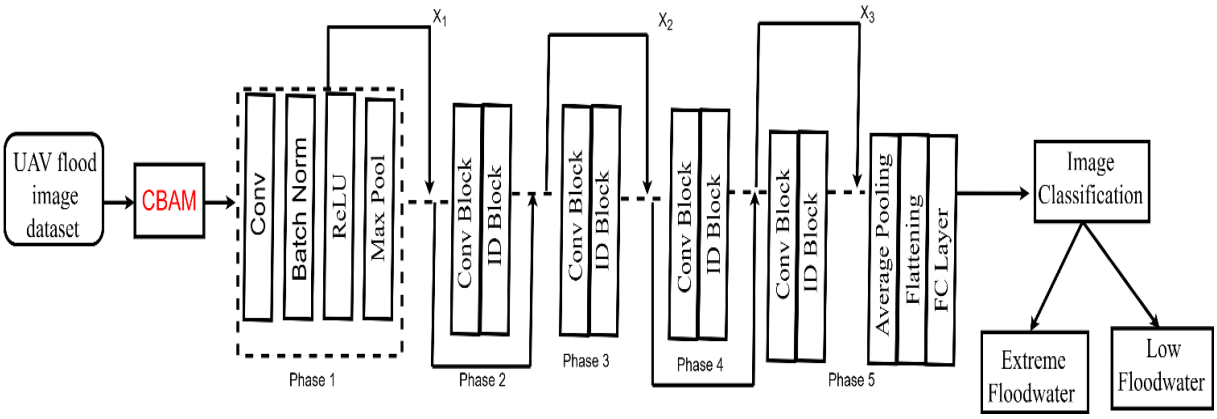

Figure 2. Pre-trained ResNet50 model with CBAM

As illustrated in **Fig. 3**, the CNN extracts hierarchical features, which are further refined by CBAM before being passed to the classification head. The final layer outputs binary classification: **extreme floodwater** vs. **low floodwater**.

3.4 Vision transformer with CBAM

The Vision Transformer (ViT) models global dependencies by dividing images into patches and applying multi-head self-attention. Integrating CBAM further improves global attention, allowing the model to emphasize regions corresponding to flood-prone areas.

Essential parts of the Vision Transformer (ViT) for estimating flood levels include a layer for embedding patches, a mechanism for self-awareness and a classification head. The ViT model for image categorization uses the transformer-like design across patches. ViT's advantage with CBAM is that it increases the accuracy of the classification with complex textures and objects of different sizes. By obtaining particular regions, ViT with CBAM lowers classification mistakes. As shown in Figure 4, the ViT pipeline includes patch embedding, multi-head attention, feed-forward layers, and a classification head. CBAM helps reduce misclassifications caused by complex textures or occlusions by selectively amplifying flood-relevant regions.

Multi-layer Perceptions (MLPs), Layer Norm, and Multi-head Attention Network (MHA) are the three main processing components that make up each of the ViT encoder's several blocks. The model adjusts to changes in training images, and Layer Norm helps maintain the training process's direction.

DOI: https://doi.org/10.54216/FPA.210221
Received: April 18, 2025 Revised: July 24, 2025 Accepted: September 02, 2025

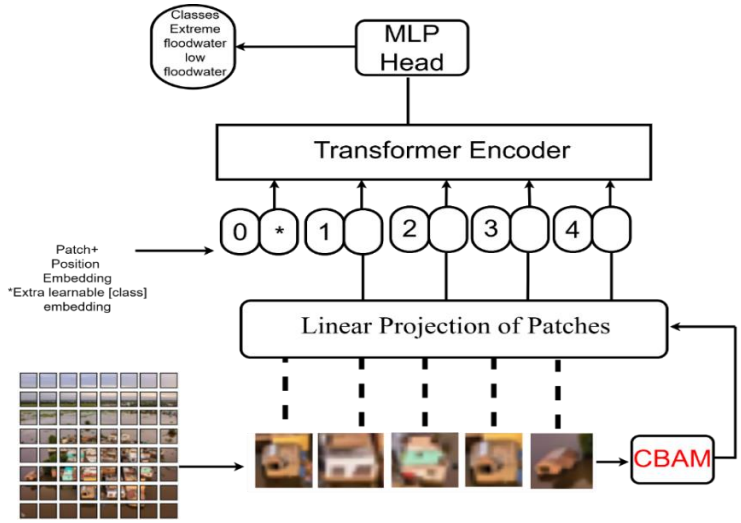

Figure 3. Vision transformer with CBAM

The network is assisted by the attention maps in concentrating on the most important areas of the UAV flood photos. To enhance performance and stabilize training, the transformer encoder employs feed-forward layers and multi-head self-attention. A fully connected layer called the classification head creates the predictions of final class.

3.5 Swin Transformer with CBAM

The Swin Transformer applies **shifted window-based self-attention**, capturing hierarchical spatial relationships at multiple scales. This is particularly effective for UAV flood imagery, where both fine-grained details and large-scale flood extents must be modelled.

By combining picture patches, the Swin transformer creates an organized representation of images. The model can capture high-level features and fine-grained details by merging image patches. With the shifted window technique, each layer of the image is divided into non-overlapping window shifting. By enabling cross-window connections and restricting self-attention computation to non-overlapping local windows, the shifting window technique offers versatility.

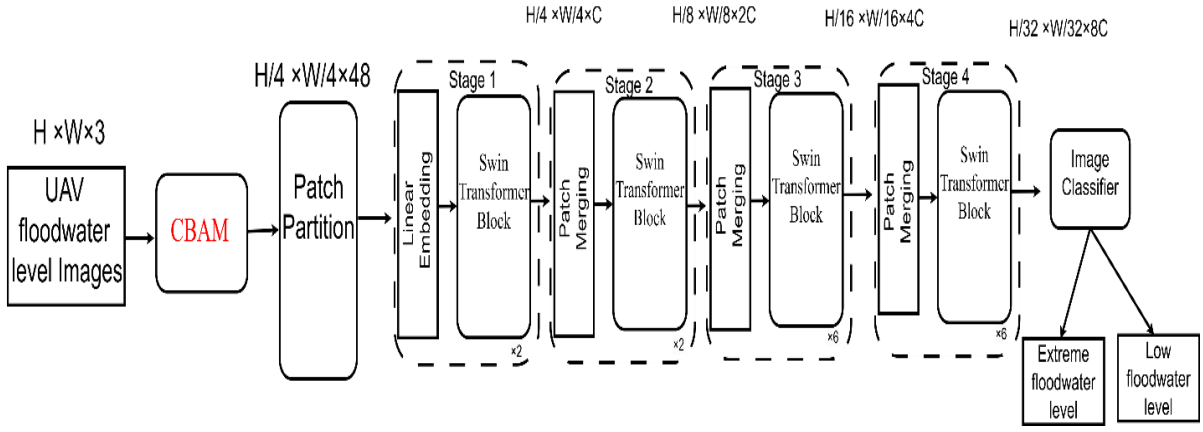

Figure 5. Swin transformer with CBAM

As shown in **Fig. 5**, CBAM is integrated into Swin blocks to refine hierarchical features, enhancing both local (house boundaries, water edges) and global (overall flood coverage) representations.

DOI: https://doi.org/10.54216/FPA.210221
Received: April 18, 2025 Revised: July 24, 2025 Accepted: September 02, 2025

3.6 Hybrid Vision Transformer with CBAM

To combine the strengths of CNNs and transformers, a hybrid architecture was developed. CNNs provide strong local feature extraction, while ViTs capture long-range global context. CBAM refines these features by emphasizing flood-critical regions

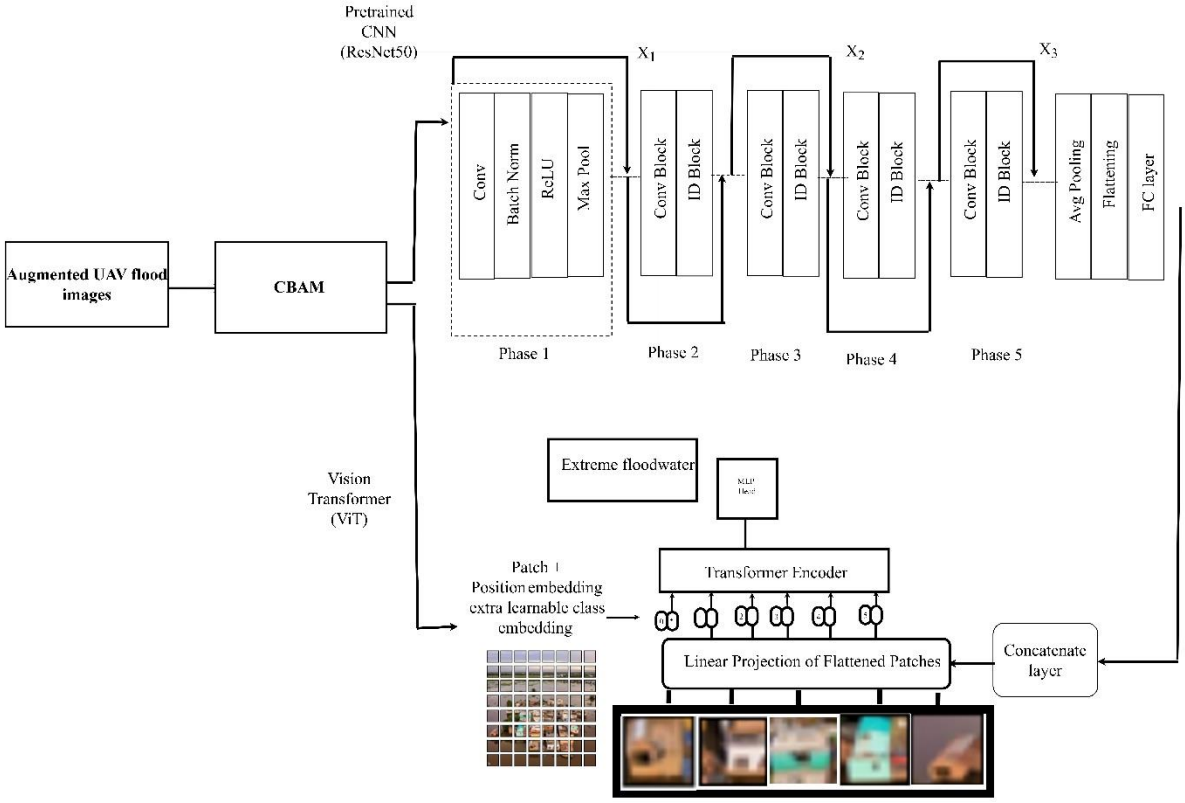

Figure 6. Hybrid vision transformer with CBAM

As illustrated in **Figure 6**, the hybrid pipeline includes:

1. UAV image input and augmentation.
2. Feature extraction using **ResNet50 (CNN backbone)**.
3. Global feature modelling using **Vision Transformer layers**.
4. Attention refinement via **CBAM**.
5. Classification into **extreme** or **low floodwater level**.

This hybrid design leverages complementary strengths, resulting in improved classification accuracy and better generalization across varied UAV flood imagery.

4. Results and Discussion

The tests conducted with pre-trained CNN models and CBAM are the main emphasis of this section. The model of CNN along with CBAM and adaptive learning rate are used in the pre-trained additional tests. Image transformers, such as Swin Transformer [35], and Vision Transformer [32], enhance model performance for flood water level estimate. The CBAM layer, a vision transformer, and pre-trained CNN models were all combined to generate the hybrid vision transformer used in the tests. The hybrid vision transformer performed well with the adaptive learning rate.

DOI: https://doi.org/10.54216/FPA.210221
Received: April 18, 2025 Revised: July 24, 2025 Accepted: September 02, 2025

The tests of the picture classification are twofold. In step one, we categorise pictures of water levels during floods with pre-trained CNN models. The second stage is associated with the CBAM layer and adaptive learning rate as a strategy to adapt the learning rate based on the schedulers. Step decay, RMSProp, exponential decay, cosine annealing, and Adagrad optimizer all belong to this strategy. The goal of this is to make the model not overfit or are stuck in some local minimal by either increasing or decreasing the learning rate depending on the optimization process. In the next step, to evaluate the performance of the models, the latest vision transformer, Swin transformer to estimate the out-of-kind without and with CBAM and rate of adaptive learning are utilized.

The following are some of the experiments conducted is as follows:

1. CNN model that has already been trained without adaptive learning rate and CBAM.
2. CNN model that has been pre-trained using adaptive learning rate and CBAM.
3. Adaptive learning rate and CBAM-free vision transformer.
4. No CBAM vision transformer
5. Adaptive learning rate as well as CBAM in Swin transformer.
6. Adaptable learning rate and CBAM in a hybrid vision transformer.

4.1 Classification of flood water level outcomes using a CNN model that has already been trained

Its trials with pre-trained CNN models occurred in several stages. First, CBAM was left out in the training of these separate models. These experiments used architectures, which include ResNet101v2, VGG19, DenseNet201, ResNet50v2, and MobileNetv2. Such a range of optimizers as Adadelta, Adam, and SGD was used, as well as the respective learning rates at 0.01, 0.1, and 0.001 after 100 training epochs. The ResNet50v2 was among the models that performed the best when it came to the classification of flood water level images where it scored an accurate percentage of 82 without using the CBAM. The table 3 includes only the best models among the pre-trained CNNs without CBAM.. All the experiments were conducted with number of epochs as 100.

Table 5: Flood water level image water level identification using pre-trained CNN models

Model	**Model**	**Optimizer**	**Learning rate**	**Training Accuracy (%)**	**Validation Accuracy (%)**
Pretrained CNN models without CBAM	ResNet101	SGD	0.1	80	74
	DenseNet201	Adadelta	0.01	81	76
	MobileNetv2	Adadelta	0.001	82	77
	ResNet50v2	**SGD**	**0.001**	**91**	**82**
	VGG19	Adam	0.001	89	80
Pretrained CNN models with CBAM	ResNet101	Adam	0.01	93	76
	DenseNet201	SGD	0.001	98	80
	MobileNetv2	**Adam**	**0.001**	**94**	**84**
	XceptionNet	SGD	0.001	81	75
	InceptionV3	Adadelta	0.1	77	74
	VGG16	Adadelta	0.01	71	65

The next experiments use CNN models that have been pre-trained with CBAM from Table 4. Using the Adam optimizer, MobileNetv2 demonstrated good performance, attaining 81% accuracy at a learning rate of 0.001. For experiments in flood water level with CBAM and adaptive learning rate as shown in Table 5 MobileNetv2 achieved 81.60% accuracy.

DOI: https://doi.org/10.54216/FPA.210221
Received: April 18, 2025 Revised: July 24, 2025 Accepted: September 02, 2025

Table 5 shows better overall performance of flood water level classification using pre-trained CNN models. This is because CBAM helps the models focus on the most significant aspects of each image. CBAM uses channel and spatial attention to emphasize key features, such as textures and patterns, while reducing attention to less relevant areas. Additionally, CBAM improves the model's generalization on unseen data, as it can effectively highlight crucial features while suppressing noise. Adding CBAM to pre-trained CNN models enhances ability to classify as extreme floodwater and low flood water levels respectively.

4.2 Results of Hybrid Vision Transformer with CBAM

The image level of flooding captured by UAV was used to run the experiments with batch sizes of 32, 64, 128 and 256. 0.001, 0.01 and 0.1 learning rates were used separately, with various optimizers, Adam, SGD, and Adadelta. Adaptive learning rate adjustments were used in both datasets of these tests. The experiments were undertaken in 100 training epochs.

Table 6: Classifying flood water level images with CBAM and Vision Transformer

Model	Batch Size	Optimizer	Learning Rate	Training Accuracy (%)	Validation Accuracy (%)
Hybrid Vision Transformer	**512**	**Adam**	**0.01**	**97.32**	**92.30**
	64	Adam	0.001	85.25	79.46
	128	Adadelta	0.001	83.20	80.31
	256	SGD	0.1	89.32	81.50
	32	Adadelta	0.01	87.30	83.59

Experiments involving the use of a vision transformer alongside CBAM and adaptive learning rate resulted in a value of accuracy of 81.25 percent as illustrated in Table 6. These analyses were carried out with the batch size of 32, the learning rate of 0.01 and SGD optimizer.

4.3 *Results of Vision Transformer with CBAM and Adaptive learning rate*

The results of the tests, which have been conducted in order to determine the water levels of floods, are presented in Table 7. The examinations use an active-adaptive-learning rate Vision Transformer in which a CBAM layer is also used at the earliest phases of image classification. Some of the optimizers that are tested include Adam, SGD, and Adadelta as well as batch sizes of 32, 64, 128, 256 and 512.

Table 7: Image classification of flood water levels utilizing Vision transformer in conjunction with CBAM.

Model with CBAM	Batch Size	Optimizer	Learning Rate	Training Accuracy (%)	Validation Accuracy (%)
Vision Transformer	**32**	**Adam**	**0.001**	**95.25**	**90.60**
	64	Adadelta	0.1	89.52	82.30
	128	Adam	0.01	85.25	79.46
	256	SGD	0.001	86.12	80.08
	512	Adam	0.001	91.36	87.23

In estimating the level of floodwater, it was observed that the 32-batch, 100-epoch, and Adam optimizer with a gradient descent rate of 0.0001 did 90 percent accuracy. The second model that employed a batch size of 512, a learning rate of 0.0001 and epochs of 100 yielded an accuracy rate of 87.23%.

DOI: https://doi.org/10.54216/FPA.210221
Received: April 18, 2025 Revised: July 24, 2025 Accepted: September 02, 2025

Effective training such as with global contexts and fine details like flood water level classification and with strong feature extraction across the image is incorporated to make the performance of the Vision Transformer better with adaptive learning rates.

4.4 Swin Transformer for Image classification with CBAM and adaptive learning rate

Categories of floodwater were also tested by training the Swin Transformer with different numbers of batches of 32, 64, 128, 256 and 512. It used techniques like RMSProp and squared gradient and batch and adaptive learning rates. This study has animal learning rate strategies, such as squared gradient, batch normalization and RMSProp. The first studies without the CBAM layer were aimed at testing batch sizes and variations of the learning rate with Swin Transformer. Using a batch size of 256 and training epochs of 100, the model got an accuracy of 73.21 percent. The initial learning rate of 0.2 assists the Swin Transformer, which takes advantage of CBAM and has its rate adaptive to achieve an accuracy of 82.18 percent with the help of the Adam optimizer and the 256-size batch (see Table 9).

Table 8: Classifying flood water level images with Swin transformer and CBAM

Model	Batch Size	Optimizer	Learning rate	Training Accuracy (%)	Validation Accuracy (%)
Swin transformer	32	SGD	0.003	70.32	65.20
	32	Adam	0.3	71.27	66.63
	256	**SGD**	**0.1**	**77.08**	**73.21**
	128	Adadelta	0.001	75.38	73.08
	64	Adam	0.2	73.84	71.04

Table 9: Classification of flood water level images using Swin transformer with adaptive learning rate and CBAM

Model	Batch Size	Optimizer	Learning Rate	Training Accuracy (%)	Validation Accuracy (%)
Swin Transformer	32	Adadelta	0.003	72.60	70.52
	32	Adam	0.03	70.32	62.50
	256	**Adam**	**0.2**	**86.47**	**82.18**
	256	SGD	0.002	86.32	74.38
	512	Adadelta	0.001	78.35	71.05

4.5 Results of hybrid vision transformer

The hybrid vision transformer unifies CNN and a vision transformer to enhance the performance of feature extracting and image classification on floodwater-level photos. The combination of the vision improver in picture classification with the good performing CNN model uses the batch size of 32, 64,128,256, and 512. The tests are performed on 100 epochs with the batch size of 512 and adaptive learning rate. Adam optimizer generated an increased accuracy.

DOI: https://doi.org/10.54216/FPA.210221
Received: April 18, 2025 Revised: July 24, 2025 Accepted: September 02, 2025

Table 10: Classifying flood water level images with an adaptive learning rate and a hybrid vision transformer

Model	Batch Size	Optimizer	Learning rate	Training Accuracy (%)	Validation Accuracy (%)
Hybrid Vision Transformer	**512**	**Adam**	**0.01**	**97.32**	**92.30**
	64	Adam	0.001	85.25	79.46
	128	Adadelta	0.1	83.20	80.31
	256	SGD	0.001	89.32	81.50
	32	Adadelta	0.01	88.30	83.50

Based on the results produced in Table 10 the hybrid vision transformer with ResNet50 and vision transformer with 100 epochs, 512 as the batch size, initial learning rate of 0.01 with Adam optimizer, the accuracy value provided by the model on flood water level classification was 92.30\%. Also, the second model, which performed best, revealed that a hybrid vision transformer had an accuracy of 83.50% when the batch size was set at 32, the number of epochs was 100, an Adadelta optimizer, and initial learning rate of 0.01.

4.6 *Comparative Analysis of various deep learning models in flood water level identification*

Table 11 summarizes the key results of our experiments across different models with and without CBAM. Among CNN-based architectures, MobileNetv2+CBAM achieved the highest accuracy (83%). ViT+CBAM outperformed individual CNNs, reaching 90.6% accuracy. Swin+CBAM achieved 86.47% with adaptive learning. The proposed Hybrid ViT+CNN+CBAM model demonstrated the best performance with **92.3% accuracy**, highlighting the benefit of combining local feature extraction and global context modelling.

Table 11: Comparative performance of different models for flood water level classification

Model with CBAM	Accuracy (%)	Precision (%)	Recall (%)	F1-score
ResNet50v2 (baseline)	82.0	81	80	0.80
MobileNetv2+CBAM	83.0	82	83	0.82
Vision Transformer	90.6	0.90	0.91	0.91
Hybrid ViT+CNN	**92.3**	**0.92**	**0.93**	**0.92**

4.7 *Limitations of the Study*

While the proposed framework demonstrates strong classification performance, certain limitations must be acknowledged:

1. Dataset Size and Diversity: The dataset consists of 2,000 UAV images, which, although balanced and augmented, may not fully capture the wide variability of flood scenarios across regions, lighting conditions, and drone altitudes. This limitation may affect the model's generalization to entirely unseen environments.

2. Computational Constraints: The integration of CNN, Vision Transformers, Swin Transformers, and CBAM requires significant GPU memory and computational resources during training. This restricts the scalability of experiments and may limit real-time deployment on resource-constrained UAV platforms.

DOI: https://doi.org/10.54216/FPA.210221
Received: April 18, 2025 Revised: July 24, 2025 Accepted: September 02, 2025

Despite these limitations, the framework offers a strong proof of concept for UAV-assisted flood severity assessment and provides a foundation for future large-scale, real-time disaster management applications

Overall, from the experiments conducted it was found that Hybrid ViT+CNN obtained an accuracy of 92.3%. Initially the experiments are carried out

5. Conclusion

This work presented a systematic investigation into deep learning-based flood water level classification using UAV imagery, ranging from pre-trained CNNs to advanced transformer-based models. Early experiments with CNN architectures such as ResNet50v2 and MobileNetv2 achieved moderate accuracies (82–83%), with improvements observed through the integration of CBAM and adaptive learning. Notably, the Vision Transformer with CBAM achieved **90.6% accuracy**, while the hybrid CNN–Vision Transformer framework reached a peak performance of **92.3% accuracy**, demonstrating the advantages of combining spatial feature extraction with attention-driven global context.

From a **practical perspective**, these results provide a foundation for developing UAV-assisted flood monitoring systems that can **automatically classify water severity levels in real time**, helping disaster management agencies to:

- **Prioritize rescue operations** in severely flooded zones where risks to human and animal lives are greatest.
- **Optimize resource allocation** by distinguishing between high-risk and moderately affected areas.
- **Support rapid situational awareness** for authorities and humanitarian organizations, reducing response delays.

While the study was limited by dataset size and computational constraints, the findings demonstrate the viability of hybrid deep learning architectures in real-world flood response. Future research will expand the dataset, enhance model robustness, and explore lightweight deployments on UAV platforms for **scalable, on-site disaster assessment.**

References

[1] United Nations Office for Disaster Risk Reduction, "Heavy Floods Widespread Across Asia," UNDRR News. https://www.undrr.org/news/heavy-floods-widespread-across-asia.

[2] Betterle and P. Salamon, "Water depth estimate and flood extent enhancement for satellite-based inundation maps," *Nat. Hazards Earth Syst. Sci.*, vol. 24, no. 8, pp. 2817–2836, 2024, doi: 10.5194/nhess-24-2817-2024.

[3] Chamatidis, D. Istrati, and N. D. Lagaros, "Vision Transformer for Flood Detection Using Satellite Images from Sentinel-1 and Sentinel-2," *Water*, vol. 16, no. 12, p. 1670, 2024, doi: 10.3390/w16121670.

[4] G. K. Wedajo, T. D. Lemma, T. Fufa, and P. Gamba, "Integrating Satellite Images and Machine Learning for Flood Prediction and Susceptibility Mapping for the Case of Amibara, Awash Basin, Ethiopia," *Remote Sens.*, vol. 16, no. 12, p. 2163, 2024, doi: 10.3390/rs16122163.

[5] Z. Li and I. Demir, "U-net-based semantic classification for flood extent extraction using SAR imagery and GEE platform: A case study for 2019 central US flooding," *Sci. Total Environ.*, vol. 869, p. 161757, 2023, doi: 10.1016/j.scitotenv.2023.161757.

[6] W. Li, D. Li, and Z. N. Fang, "Intercomparison of automated near-real-time flood mapping algorithms using satellite data and DEM-Based methods: a case study of 2022 Madagascar flood," *Hydrology*, vol. 10, no. 1, p. 17, 2023, doi: 10.3390/hydrology10010017.

[7] Elkhrachy, "Flash flood water depth estimation using SAR images, digital elevation models, and machine learning algorithms," *Remote Sens.*, vol. 14, no. 3, p. 440, 2022, doi: 10.3390/rs14030440.

[8] H. Hosseiny, "A deep learning model for predicting river flood depth and extent," *Environ. Model. Softw.*, vol. 145, p. 105186, 2021, doi: 10.1016/j.envsoft.2021.105186.

[9] D. Bonafilia, B. Tellman, T. Anderson, and E. Issenberg, "Sen1Floods11: A georeferenced dataset to train and test deep learning flood algorithms for sentinel-1," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 210–211, doi: 10.1109/CVPRW50498.2020.00113.

[10] S. Cohen *et al.*, "Estimating floodwater depths from flood inundation maps and topography," *JAWRA J. Amer. Water Resour. Assoc.*, vol. 54, no. 4, pp. 847–858, 2018, doi: 10.1111/1752-1688.12609.

DOI: https://doi.org/10.54216/FPA.210221
Received: April 18, 2025 Revised: July 24, 2025 Accepted: September 02, 2025

[11] F. Cian, M. Marconcini, P. Ceccato, and C. Giupponi, "Flood depth estimation by means of high-resolution SAR images and lidar data," *Nat. Hazards Earth Syst. Sci.*, vol. 18, no. 11, pp. 3063–3084, 2018, doi: 10.5194/nhess-18-3063-2018.

[12] Stateczny, H. D. Praveena, R. H. Krishnappa, K. R. Chythanya, and B. B. Babysarojam, "Optimized Deep Learning Model for Flood Detection Using Satellite Images," *Remote Sens.*, vol. 15, no. 20, p. 5037, 2023, doi: 10.3390/rs15205037.

[13] Q. C. Le, M. Q. Le, M. K. Tran, N. Q. Le, and M. T. Tran, "FL-Former: Flood Level Estimation with Vision Transformer for Images from Cameras in Urban Areas," in *Int. Conf. Multimedia Modeling*, Cham: Springer, 2023, pp. 447–459, doi: 10.1007/978-3-031-27077-2_35.

[14] N. Anusha and B. Bharathi, "Flood detection and flood mapping using multi-temporal synthetic aperture radar and optical data," *Egypt. J. Remote Sens. Space Sci.*, vol. 23, no. 2, pp. 207–219, 2020, doi: 10.1016/j.ejrs.2019.01.001.

[15] K. Tanaka, Y. Fujihara, K. Hoshikawa, and H. Fujii, "Development of a flood water level estimation method using satellite images and a digital elevation model for the Mekong floodplain," *Hydrol. Sci. J.*, vol. 64, no. 2, pp. 241–253, 2019, doi: 10.1080/02626667.2019.1578463.

[16] P. Zhong, Y. Liu, H. Zheng, and J. Zhao, "Detection of urban flood inundation from traffic images using deep learning methods," *Water Resour. Manag.*, vol. 38, no. 1, pp. 287–301, 2024, doi: 10.1007/s11269-023-03669-9.

[17] J. Wan *et al.*, "Automatic detection of urban flood level with YOLOv8 using flooded vehicle dataset," *J. Hydrol.*, vol. 639, p. 131625, 2024, doi: 10.1016/j.jhydrol.2024.131625.

[18] M. Ranieri, T. L. de Souza, M. Nishijima, B. Krishnamachari, and J. Ueyama, "A deep learning workflow enhanced with optical flow fields for flood risk estimation," *Appl. Intell.*, vol. 54, pp. 5536–5557, 2024, doi: 10.1007/s10489-024-05466-2.

[19] K. J. Wienhold, D. Li, W. Li, and Z. N. Fang, "Flood Inundation and Depth Mapping Using Unmanned Aerial Vehicles Combined with High-Resolution Multispectral Imagery," *Hydrology*, vol. 10, no. 8, p. 158, 2023, doi: 10.3390/hydrology10080158.

[20] G. Popandopulo *et al.*, "Flood extent and volume estimation using remote sensing data," *Remote Sens.*, vol. 15, no. 18, p. 4463, 2023, doi: 10.3390/rs15184463.

[21] H. S. Munawar, F. Ullah, S. Qayyum, and A. Heravi, "Application of deep learning on uav-based aerial images for flood detection," *Smart Cities*, vol. 4, no. 3, pp. 1220–1242, 2021, doi: 10.3390/smartcities4030065.

[22] J. Strebl *et al.*, "Flood level estimation from social media images," in *CEUR Workshop Proc.*, vol. 2670, MediaEval, pp. 1–3, 2019.

[23] A. Kharazi and A. H. Behzadan, "Flood depth mapping in street photos with image processing and deep neural networks," *Comput., Environ. Urban Syst.*, vol. 88, p. 101628, 2021, doi: 10.1016/j.compenvurbsys.2021.101628.

[24] J. L. Gan and W. Zailah, "Water level classification for flood monitoring system using convolutional neural network," in *Proc. 11th Natl. Tech. Semin. Unmanned Syst. Technol. 2019: NUSYS'19*, Singapore: Springer, 2021, pp. 299–318, doi: 10.1007/978-981-15-5281-6_21.

[25] Y. Feng, C. Brenner, and M. Sester, "Flood severity mapping from Volunteered Geographic Information by interpreting water level from images containing people: A case study of Hurricane Harvey," *ISPRS J. Photogramm. Remote Sens.*, vol. 169, pp. 301–319, 2020, doi: 10.1016/j.isprsjprs.2020.09.011.

[26] K. A. C. Quan *et al.*, "Flood level prediction via human pose estimation from social media images," in *Proc. 2020 Int. Conf. Multimedia Retrieval*, 2020, pp. 479–485, doi: 10.1145/3372278.3390704.

[27] Y. T. Lin, M. D. Yang, J. Y. Han, Y. F. Su, and J. H. Jang, "Quantifying flood water levels using image-based volunteered geographic information," *Remote Sens.*, vol. 12, no. 4, p. 706, 2020, doi: 10.3390/rs12040706.

[28] H. Rizk, Y. Nishimur, H. Yamaguchi, and T. Higashino, "Drone-based water level detection in flood disasters," *Int. J. Environ. Res. Public Health*, vol. 19, no. 1, p. 237, 2021, doi: 10.3390/ijerph19010237.

DOI: https://doi.org/10.54216/FPA.210221
Received: April 18, 2025 Revised: July 24, 2025 Accepted: September 02, 2025

[29] Y. Liang, X. Li, B. Tsai, Q. Chen, and N. Jafari, "V-FloodNet: A video segmentation system for urban flood detection and quantification," *Environ. Model. Softw.*, vol. 160, p. 105586, 2023, doi: 10.1016/j.envsoft.2022.105586.

[30] P. Chaudhary, S. D'Aronco, M. Moy de Vitry, J. P. Leitão, and J. D. Wegner, "Flood-water level estimation from social media images," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 4, no. 2/W5, pp. 5–12, 2019.

[31] Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv: 2010.11929*, 2020.

[32] Z. Liu *et al.*, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 10012–10022, doi: 10.1109/ICCV48922.2021.00986.

[33] R. J. Pally and S. Samadi, "Application of image processing and convolutional neural networks for flood image classification and semantic segmentation," *Environ. Model. Softw.*, vol. 148, p. 105285, 2022, doi: 10.1016/j.envsoft.2021.105285.

[34] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19, doi: 10.1007/978-3-030-01234-2_1.

[35] S. S. Kulkarni and A. Mahapatra, "Improving Satellite Flood Image Classification Using Attention-Based CNN and Transformer Models," *Int. J. Adv. Comput. Sci. Appl.*, vol. 16, no. 3, 2025, doi: 10.14569/IJACSA.2025.01603101.

[36] S. S. Kulkarni and A. Mahapatra, "Post flood assessment using deep learning techniques," in *AIP Conf. Proc.*, vol. 2917, no. 1, p. 050001, AIP Publishing, 2023, doi: 10.1063/5.0175612.

DOI: https://doi.org/10.54216/FPA.210221
Received: April 18, 2025 Revised: July 24, 2025 Accepted: September 02, 2025