Improving Satellite Flood Image Classification Using Attention-Based CNN and Transformer Models

Sanket S Kulkarni, Ansuman Mahapatra
Department of Computer Science and Engineering,
National Institute of Technology Puducherry, Karaikal, Puducherry 609609, India

***Abstract*—Floods are among the most frequent and devastating natural disasters, significantly impacting infrastructure, ecosystems, and human communities. Accurate satellite-based flood image classification is crucial for assessing flood-affected regions and supporting emergency response efforts. This study uses Convolutional Neural Networks (CNNs) and transformer-based architectures to enhance flood classification, integrating the Convolutional Block Attention Module (CBAM) to improve feature extraction. Using the xView2 xBD dataset, we classify houses as completely or partially surrounded by floodwater. Experimental evaluations demonstrate that ResNet101v2 achieved an accuracy of 86.87%, while a hybrid CNN model (MobileNetV2- DenseNet201) attained 85.83%, further improving to 89.54CBAM. The Vision Transformer (ViT) with CBAM achieved the highest accuracy of 90.75%, showcasing the effectiveness of attention-based hybrid models for flood image classification. These results highlight the potential of integrating CBAM with deep learning architectures to enhance classification accuracy and improve flood impact assessment.**

Keywords*—CNN; DenseNet; ResNet101v2; VGG16; hybrid CNN model; CBAM; vision transformer; xView2 Building Damage (xBD)

I. Introduction

Floods significantly impact society every year, i.e. causing considerable losses to humans and livestock due to urbanization and global climate changes. Many Asian countries such as India, China and Bangladesh have been prone to the significant effects of flooding recently, as per the reports from the United Nations Office of Disaster Risk Reduction (UNDRR) [1]. Reports from the National Disaster Management Authority (NDMA) indicate that a significant portion of India's geographical area is prone to flooding, highlighting the need for effective flood management strategies [2]. Floods increase in frequency and intensity due to climate change and unplanned urbanization. There are two flood assessment methods, namely, pre-flood and post-flood assessment techniques. The pre-flood assessment techniques refer to determining flood mitigation strategies and evaluating the risk of flooding from all potential sources. The pre-flood evaluation has several issues, such as building roads, reservoirs, and dams, which would be expensive and time-consuming. Conventional methods for managing floods include allowing the flood peak to pass without overflowing and reducing intensity by holding or diverting a portion of inflows or increasing the capacity of the stream. Therefore, post-flood assessment is given more emphasis due to the drawbacks of pre-flood assessment techniques.

Fig. 1(a) shows the completely surrounded house by flood water, and Fig. 1(b) shows the partially surrounded house by floodwater. The regions completely covered by flood water indicate the possibility of trapped humans and livestock; hence, it helps the rescue teams focus on areas surrounded by flood water.

(a)

(b)

Fig. 1. (a) Completely covered house by flood water, and (b) Partially surrounded house by flood water.

Deep learning-based satellite flood image classification has a wide range of critical applications, particularly in disaster response and relief operations. By leveraging advanced deep learning models, this work enables rescue teams to accurately identify and prioritize areas where resources are most needed, such as houses completely surrounded by floodwater, ensuring efficient and timely interventions. Post-flood assessment techniques refer to estimating the conditions of different areas after the flood has occurred. The significant advantages of post-flood assessment include flood monitoring ([3],[7],[6]), flood zone mapping ([8], [9], [10]), flood forecasting ([11], [12]), and flood rescue operations ([13], [14]). For specific visible ranges, it is possible to identify whether regions are completely surrounded or partially surrounded by flood water.

The major contribution of this work includes:

- Creating an image dataset by segregating the satellite images into two classes: houses completely and partially surrounded by floodwater.
- Experimentally fine-tuning hyper-parameters for pre-trained CNN, hybridizing top-performing architectures and various transformer models.
- Integrating Convolutional Block attention Module(CBAM) on various CNN models and transformers for performance improvement.

Section II discusses related works on satellite flood image classification; Section III discusses dataset description, augmentation techniques, and various architectures used for image classification. Section IV discusses the results of the experiments carried out. Section V discusses inferences from the results of experiments carried out. Section VI discusses the conclusion and future scope.

II. Related Works

This section focuses on the most recent satellite-based post-flood assessment research. Most of the researchers use satellite images to map flood areas. Only limited work is related to classifying houses as damaged or not damaged. The existing works on satellite flood images are discussed here in this section. Chamatidis *et al.* (2024) utilized a Vision Transformer combined with transfer learning to detect flooding in satellite imagery [16]. This work uses two distinct datasets to train two separate datasets. Sentinel-1 comprises Synthetic Aperture Radar (SAR) images capturing flood and non-flood events across various regions. The second dataset, Sentinel-2, consists of multispectral imagery acquired from multiple flood and non-flood scenarios in different locations.In their work, Saleh *et al.* (2024) proposed a semantic token as SemT-Former, which operates by prioritizing changes of interest rather than fully comprehending the entire image scene [15].

Kaur *et al.* (2023) used a novel transformer-based network for assessing building damage [31]. The transformer-based network used hierarchical spatial features of multiple resolutions and captured temporal differences in the feature domain by applying a transformer encoder to the spatial features.

Gupta *et al.* (2019) has created a vast satellite image dataset on many natural disasters in different regions of the world. They have classified the houses as damaged or not damaged post-disaster scenarios [30]. xBD dataset is a large dataset developed for building damage assessment to provide humanitarian aid and help in rescue operations. Jiang *et al.* (2021) proposed a segmentation algorithm for automatic flood mapping in near real-time, spanning vast areas and in all weather conditions by integrating Sentinal-1 SAR imagery with an unsupervised machine learning approach named Felz-CNN [25]. Munoz *et al.* developed a deep learning and fusion framework for large-scale compound flood mapping [33]. Pham *et al.* (2021) proposed a novel approach for flood risk assessment, which is a combination of a deep learning algorithm and Multi-Criteria Decision Analysis (MCDA) and also a flood risk assessment framework for integration of hazard, exposure, and vulnerability mask [34]. Hafizi Mohd Ali *et al.* proposed a time series model with layer normalization and leaky ReLU activation function [41]. Rahnemoonfar *et al.* proposed the FloodNet dataset to demonstrate the post-flood damages of the affected areas [32]. They compared and contrasted the performance of baseline methods for image classification, semantic segmentation, and visualization of flood data. Wu *et al.* dual-polarization SAR data and multi-scale features of SAR images, an effective flood detection method for SAR images [35]. Table I lists some satellite image classification works related to flood areas. The literature review shows a minimal number of works on satellite image classification for floods due to the low resolution of the images. There is no work on classifications of buildings completely or partially surrounded by floodwater.

III. Methodology

A. Dataset Description

The challenges, such as the scarcity of high-resolution images and the limited availability of datasets, often constrain the classification of satellite flood images, reducing classification accuracy. There are various other problems, such as imbalanced class distribution. The Satellite flood image classification datasets encounter limitations such as class imbalances, geographic biases, and challenges posed by occlusions from clouds or vegetation. The xBD satellite flood image dataset is sourced from Maxar/DigitalGlobe open data, featuring high-resolution imagery [30]. The geographical area covered is approximately 18000 km^2, with high-resolution images providing a detailed analysis of regions affected by flooding. The xBD dataset includes images from various areas, including those capturing the Midwest US Floods between January 3 and May 31, 2019. These floods primarily impacted the midwestern United States, particularly along the Missouri River.

The xBD dataset used in this work is categorized into two classes: completely surrounded houses by floodwater and partially surrounded houses by floodwater. In the completely surrounded house category, the house is fully submerged, with no visible escape routes such as roads or pathways, indicating a critical need for immediate rescue. Conversely, partially surrounded houses may have accessible pathways or roads that could serve as potential escape routes for trapped individuals, requiring less urgent attention but still necessitating intervention.

In the xBD dataset for our model training, 5382 images are segregated into two classes, namely houses completely or partially surrounded by flood water. Each class contains 2691 images, which is equally balanced. Table II shows the number of images used for classification. Images are split into two folders with train (70%) and Validation (30%), respectively.

B. Dataset Augmentation

Data augmentation techniques were applied to the xBD satellite flood image dataset to address the limited availability of images and enhance the training dataset's diversity. The augmentation process includes image rotation, flipping, and saturation adjustment. These transformations, as summarized in Table III, simulate variations in lighting conditions, color intensities, and the time of image capture, thereby improving the robustness and generalization of the classification techniques.

TABLE I. RELATED WORKS ON POST-FLOOD ASSESSMENT FROM SATELLITE IMAGES

Method	Dataset Used	Features	Application
Wu Z*et al.* (2024) [5]	GID dataset and GIH-Water dataset	Multi-scale transformer-based algorithm for floodwater contour extraction	Flood water body delination Roboust solution on disaster stuck areas
Wu L*et al.* (2024) [4]	The dataset comprising of 2945 flood house images with four damage level	Proposed dual-view CNN for post-flood damage levels in houses	Identify damage flood house level
Montello*et al.* (2022) [21]	Dataset contains 1,748 Sentinel-1 acquisitions comprising 95 flood events	flood delineation task using deep learning models to evaluate the performance gains of entropy-based sampling and multi encoder architecture.	Assessment of flood areas accurately
Jackson *et al.* (2023) [19]	FloodNet Dataset	ResNet18, VGG16, MobileNetv2 for building damage assessment	Identification of flood risk areas
Pech *et al.* (2023) [20]	SAR images from Campeche,Chiapas and Tabasco,Mexico	U-Net for flood mapping	Detection of flooded areas
Islam *et al.* (2022) [18]	The dataset comprises three classes	Inceptionv3, DenseNet CNN approach for flood severity assessment	Identify flood areas and help in rescue operations
J. Ha and J.E Kang (2022) [22]	Flood data from Busan city	Flood risk level using random forest model	Identify flood risk areas
Bouchard *et al.* (2022) [23]	xBD dataset	CNN in building damage assessment from post-disaster	Flood building damage assessment
Franceschini *et al.* (2021) [17]	Spatial aerial flood image	Detect and localize flood buildings	Building damage assessment
Shen *et al.* (2021) [24]	xBD dataset	Two stage CNN for building damage assessment	Building damage assessment
Xin *et al.* (2021) [25]	Sentinel-1a and Sentinel-1b for mapping flood inundation area	Unsupervised machine learning approach Felz-CNN for flood mapping	Effective monitoring of flood conditions to aid disaster governance
Opella *et al.* (2019) [26]	Used data from GIS	Fused ConvNet, along with SVM	Effective and robust flood map for image classification
Moya *et al.* (2019) [28]	TerraSAR-X intensity images	3DGLCM for building damage classification	Flood building damage assessment
Chandrama Sarker *et al.* (2019) [27]	Landsat and WOfS images	Fully convolutional neural networks (F-CNNs)	Flood extent mapping from Landsat satellite images

TABLE II. DATASET DESCRIPTION OF IMAGES USED FOR CLASSIFICATION

Dataset	Completely Surrounded	Partially Surrounded	Total Images
Train (70%)	1883	1883	3766
Validation (30%)	808	808	1616

The model is better equipped to handle real-world scenarios with diverse environmental conditions and perspectives by augmenting the dataset.

TABLE III. DATA AUGMENTATION FOR FLOOD IMAGE CLASSIFICATION

Transformation Applied	Value of Transformation
Image Rotation	0^0, 90^0, 180^0,270^0
Image Flipping	50%
Saturation	±30%
Exposure	±15%

C. Convolutional Block Attention Module (CBAM)

The Convolutional Block Attention Module (CBAM) is an attention module for feed-forward convolutional neural networks. Given an intermediate feature map, this module would sequentially infer attention maps along two separate dimensions, channel and spatial. Then, the attention maps are multiplied by the input feature map for adaptive feature refinement [42] as shown in Fig. 2. CBAM is a lightweight and general module that can easily integrate into CNN architectures, seemingly with integrated weights.

CBAM, added with CNN, extracts hierarchical features from input images through multiple convolutional layers followed by pooling and activation functions. During image classification, traditional CNN models consist of relevant and irrelevant features. Here, adding CBAM enhances the model's attention to essential features.

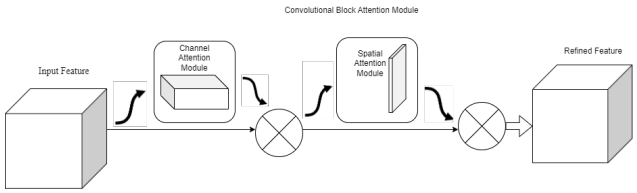

Fig. 2. Convolutional Block Attention Module (CBAM).

The key characteristics of using CBAM are that it is computationally efficient and easily integrates with existing models to improve computational complexity. CBAM for image classification includes enhanced feature representation, which improves the model's ability to capture essential features by focusing on the most informative channels and spatial regions. The CBAM provides flexibility since it can be easily integrated into existing CNN architectures without significant changes for classification tasks.

D. Individual Pre-trained CNN Models with CBAM

There are ten pretrained individual architectures such as VGG16, VGG19, ResNet50, XceptionNet, MobileNetv2, ResNet101v2, DesnseNet201, Inceptionv3, XceptionNet, and Inception-ResNet ResNet are fine-tuned with our dataset to classify the houses in satellite images as partially or completely surrounded by flood water. Fig. 3 shows the various stages of image classification using individual pre-trained architecture. These pre-trained CNN models are selected since they are top-performing models in terms of image classification.

Individual pre-trained CNN models for flood image classification are vital because they can extract robust and hierarchical features from images. These models, pre-trained on large datasets like ImageNet, can be fine-tuned for flood-specific tasks, such as distinguishing between partially and fully flooded areas. Their convolutional layers effectively capture spatial patterns, such as water boundaries and submerged structures, which are critical for accurate flood assessment. Moreover, these models' adaptability to various datasets and computational efficiency make them suitable for real-time applications in disaster response, flood monitoring, and resource allocation.

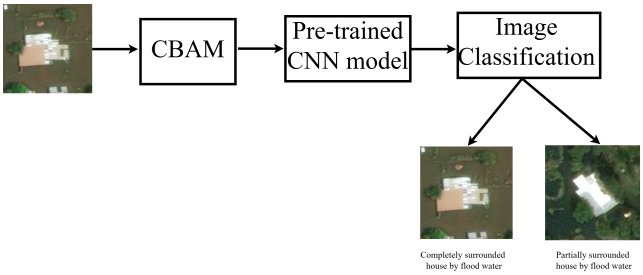

Fig. 3. Stages involved in the individual pre-trained CNN architecture.

In order to include the CBAM in the design of individual pre-trained models, attention modules that apply spatial and channel-wise attention mechanisms successively are incorporated. These attention modules enable the model to focus on the most relevant regions of the flood images, such as water-logged areas around houses, while suppressing less informative background details. The combined model is fine-tuned on the flood image dataset to adapt the pre-trained features and CBAM-enhanced attention to the specific classification task.

E. Hybrid CNN Architecture Convolutional Block Attention Module

Based on the performance of individual pre-trained CNN models, a hybrid architecture was designed by combining two best individual pre-trained CNN models for feature extraction. This architecture capitalizes on the complementary strengths of both models, leveraging their distinct feature extraction capabilities, as depicted in Fig. 4. The inclusion of CBAM is further refined the attention mechanism, improving model accuracy. This architecture was selected through iterative experimentation, ensuring an optimal balance between computational efficiency and classification performance.

The feature maps of both pre-trained network layers are concatenated. The concatenation layer merges the features extracted by both pre-trained networks, allowing the hybrid model to utilize features from both architectures for enhanced classification. In the initial stage, the flood image dataset is provided as input to the two pre-trained CNN models, namely, pre-trained model 1 and pre-trained model 2. In pre-trained model 1, the model processes the input images through its layers and generates feature maps. An averaging layer computes the average value across each feature map to reduce dimensionality. Similarly, the pre-trained model 2 extracts feature representations from the input images. Further, it is given as input to the averaging layer, ensuring that the feature maps are reduced to a manageable size. Then, further, each pre-trained model is followed by a dense Prediction layer, which generates a set of output predictions based on the features extracted by the respective models. These dense prediction layers classify the flood images using the information obtained by each pre-trained model.

The outputs from feature maps of the dense prediction layers from the two models are merged through a concatenation layer, subsequently serving as input to a dense prediction layer. This layer is responsible for classifying the images into two classes: completely surrounded houses by floodwater or partially surrounded houses by floodwater. The Convolutional Block Attention Module (CBAM) is a lightweight and effective attention mechanism that can enhance the performance of deep learning models in satellite image classification tasks, such as flood detection and assessment. By sequentially applying channel and spatial attention, CBAM enables the model to focus on the most relevant features in satellite imagery, such as water bodies, flood extents, and damaged areas, while suppressing irrelevant background information.

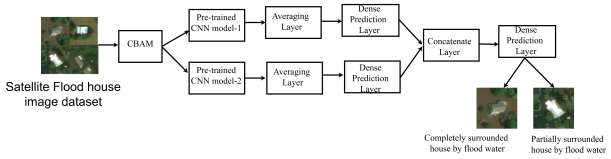

Fig. 4. Stages involved in the design of hybrid CNN architecture.

Fig. 4 shows the architecture of modification made to the pre-trained CNN architectures after applying CBAM. The CBAM is added before the final descent prediction layer, which classifies the houses as completely or partially surrounded houses by flood water.CBAM processes the extracted feature maps to refine them by emphasizing relevant spatial and channel-specific features. After applying CBAM processes, the extracted feature maps are refined by emphasizing relevant spatial and channel-specific features.

F. Architecture for Data Efficient Image Transformer (DeiT) with CBAM

The Data-Efficient Image Transformer (DeiT) is employed for satellite flood image classification, leveraging its efficiency in learning from datasets with high accuracy [36].

The DeiT incorporates a teacher-student learning distillation mechanism that enhances transferring from the convolutional neural network (CNN) teacher model to the transformer. For satellite flood classification, the input images are pre-processed into fixed-size patches, embedded, and processed

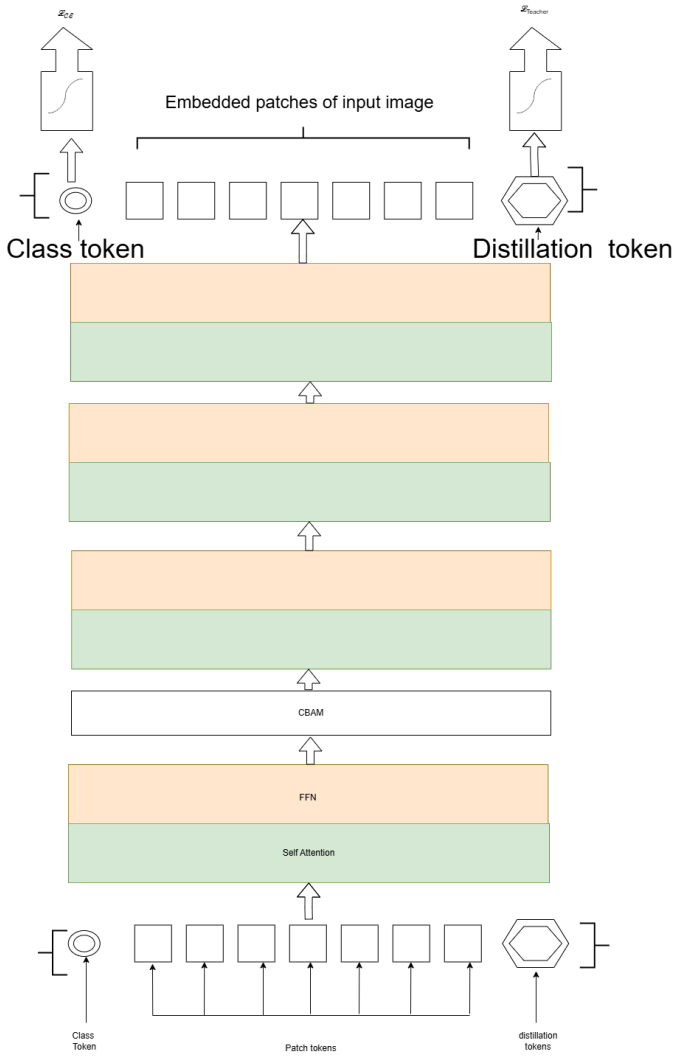

Fig. 5. Architecture of data efficient image transformer.

through multiple transformer layers, allowing both flood-specific patterns and global spatial dependencies to be captured during this process.

The advantages of DeiT are that it effectively trains on all datasets, it is compact with variants, and The distillation process enhances DeiT accuracy, making it competitive with state-of-the-art convolutional neural networks (CNN). The DeiT model achieves higher accuracy when compared to pre-trained CNN models. The hierarchical self-attention mechanism makes it salable for small and large-scale image classification. The DeiT leverages global self-attention, no inductive bias, and parallelization to obtain global and local features. Fig. 5 shows the architecture diagram for flood image classification. The DeiT for satellite flood image classification ensures that the model focuses on critical flood-related features, such as identifying houses completely or partially surrounded by flood water.

DeiT uses a self-attention mechanism to capture global dependencies and identify subtle patterns and features indicative of the flood effect. The feature extraction process is enhanced by integrating the CBAM to focus on critical regions of images. DeiT with CBAM enhances the accurate classification of houses completely or partially surrounded by flood water.

To further enhance performance, specific challenges in flood house image classification, such as variations in lighting, viewing angles, and physical obstructions, are addressed by fine-tuning the DeiT model's architecture. The DeiT-integrated CBAM emphasizes flood-relevant features while suppressing irrelevant or noisy information in the images. This combination allows the model to capture critical spatial and contextual patterns effectively, improving its robustness and accuracy in classifying flood-affected houses in diverse scenarios.

G. Architecture for Multiscale Vision Transformer (MviT) for Satellite Flood Image Classification with CBAM

The Multiscale Vision Transformer (MViT) model efficiently captures global and local spatial features across multiple scales. By incorporating multiscale attention mechanisms, the MViT adaptively focuses on fine-grained features, such as flooded areas, to enhance classification accuracy. The model is configured with a patch-based tokenization strategy, ensuring the preservation of critical spatial features throughout the processing pipeline.

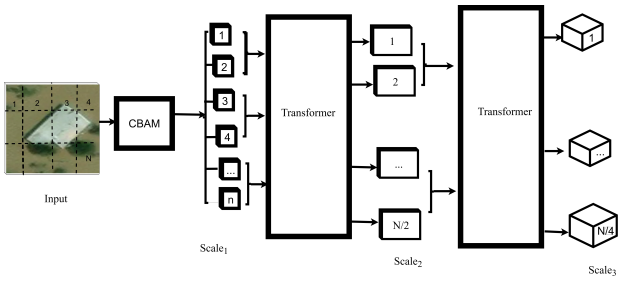

Fig. 6. Architecture of Multiscale Vision Transformer (MViT).

Fig. 6 shows the architecture for a Multiscale vision transformer (MviT). The CBAM is integrated into the architecture to enhance feature refinement by selectively emphasizing flood-relevant spatial and channel-wise information.

H. Architecture for Swin Transformer for Satellite Flood Image Classification with CBAM

The Swin Transformer is considered well-suited for flood image classification due to its hierarchical architecture and shifted window mechanism. It effectively captures global and local features, accurately identifying flood-affected regions in satellite images [37]. By leveraging its multiscale representation, the Swin Transformer can differentiate between partially and fully inundated areas, contributing to accurate flood zone mapping and rescue prioritization. Its efficiency and scalability make it ideal for processing high-resolution flood imagery in real-world disaster scenarios.

The CBAM, which includes channel and spatial attention modules, is integrated into the Swin Transformer to enhance its feature extraction capabilities. The integration of CBAM with the Swin Transformer occurs at key stages of the model architecture.CBAM is integrated into the model by inserting it

after the attention layers of the transformer blocks, allowing the network to refine its attention maps and focus on more relevant features.

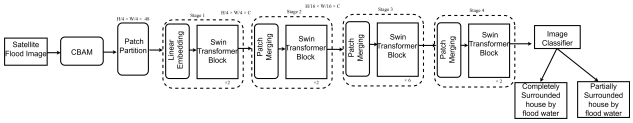

Fig. 7. Swin transformer with CBAM.

Fig. 7 presents the Swin Transformer with CBAM, where the input undergoes sequential processing through multiple transformer blocks. After each transformer block, CBAM is incorporated before the output is passed to the next block or the final classification layer. This setup allows for the refinement of spatial and channel features after each block's multi-head self-attention and MLP operations, ensuring the extraction of distinct features at each stage.

I. Architecture for Sparse Swin Transformer for Flood Image Classification with CBAM

Sparse Swin Transformer is a variation of the Swin Transformer architecture where the attention mechanism is designed to focus on only the most critical parts of an image, effectively sparsifying the attention leading to faster computation leading to faster computation and potentially improved accuracy with few parameters compared to standard Swin transformer [38].

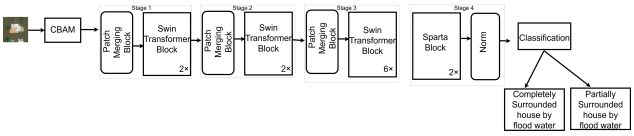

Fig. 8. Sparse Swin transformer with CBAM.

Fig. 8 shows the architecture diagram for Sparse Swin Transformer with CBAM selectively focusing on critical flood-relevant regions, such as water boundaries and inundated areas, reducing computational complexity without compromising feature extraction. The hierarchical architecture of the Sparse Swin Transformer facilitates multi-scale feature learning, enabling the model to capture both local details and global context from the images. For this study, the satellite datasets were pre-processed into patches and fed into the transformer, preserving spatial information. The model integrates CBAM (Convolutional Block Attention Module) to enhance attention to flood-relevant features in spatial and channel dimensions.

J. Architecture for Hierarchical Vision Transformer(HVT) for Flood Image Classification with CBAM

The Hierarchical Vision Transformer (HVT) is utilized for flood image classification to effectively analyze satellite imagery by leveraging its hierarchical structure and multiscale feature extraction capabilities [39].

The Hierarchical Vision Transformer (HVT) model, integrated with CBAM, is utilized for flood image classification to capture local and global contextual information as shown in Fig. 9. The hierarchical structure of HVT enables efficient

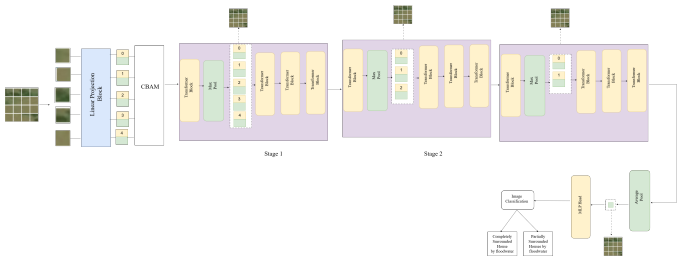

Fig. 9. Hierarchical vision transformer with CBAM.

processing of high-resolution flood images by focusing on multiscale features. CBAM further enhances this by selectively emphasizing flood-relevant features and suppressing irrelevant ones, improving the model's ability to accurately classify flood-related patterns. This combination leads to a more precise and robust flood image classification.

The hierarchical vision transformer divides input satellite images into progressively finer patches, allowing the model to capture global contextual information. The hierarchical vision transformer architecture was augmented with a Convolutional Block Attention Module (CBAM) to enhance spatial and channel-level attention, ensuring a more targeted focus on flood-relevant features. Initially, the experiments are carried out without adding the CBAM layer, where the focus is distributed across all parts of the image rather than directed toward specific, critical regions. This approach provides a baseline performance, allowing for a comparison to evaluate the impact of CBAM in enhancing feature selection and improving classification accuracy.

K. Architecture for Vision Transformer for Satellite Flood Image Classification with CBAM

The Vision Transformers with CBAM enhance the features to identify the flooded houses [40]. ViT effectively captures long-range dependencies. ViT processes the image as a patch sequence, allowing it to learn from a global context for satellite image classification. Using a pre-trained ViT model, typically fine-tuned on large image datasets, allows leveraging learned representations to improve satellite image performance.

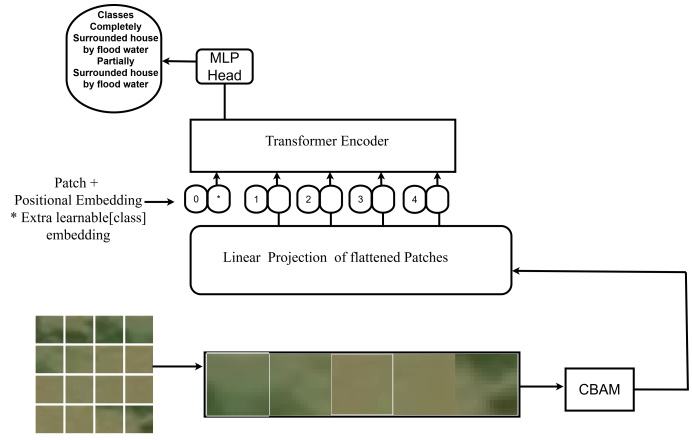

Fig. 10. Vision transformer with CBAM.

The satellite images are passed through the transformer layers to learn the spatial and semantic features. A classification head processes the output tokens, typically a fully connected layer, to predict the class of the satellite image. Fig. 10 shows the architecture for flood image classification with CBAM layer, which is added after satellite flood image patches to focus on relevant features such as identification of flooded houses. The Vision Transformer (ViT) architecture combines the strengths of attention-based mechanisms in both spatial and channel dimensions, enhancing its performance for image classification tasks.

IV. Results

This section mainly focuses on the experiments conducted with varying learning rates. The various experiments carried out include:

- Flood image classification using individual pre-trained CNN models.
- Flood image classification using hybrid CNN model.
- Flood image classification using Sparse Swin Transformer .
- Flood image classification using Data efficient Image Transformer (DeIT)
- Flood image classification using Multiscale Vision transformer (MviT).
- Flood image classification using Swin transformer.
- Flood image classification using Hierarchical Vision transformer(HVT).
- Flood image classification using Vision transformer (ViT).

Various experiments were performed using optimizers such as Adam, SGD, and Adadelta, with learning rates of 0.1, 0.01, 0.001, and 0.0001. They perform the experiments for both 50 and 100 epochs, consistently observing that the models achieve peak accuracy within 50 epochs. Additionally, the impact of adding attention mechanisms like CBAM is analyzed to determine their variation in accuracy improvements by fine-tuning the hyperparameters. Experiments comprises of low learning rates such as 0.0001,0.001,0.01 since the pre-trained models have already been trained on numerous images, hence the flood image classification is performed with lower learning rates to classify houses as completely or partially surrounded by flood water.

A. Results of Individual Pre-trained CNN Models

The experiments performed for flood image classification on satellite images to classify houses completely or partially surrounded by floodwater for these individual pre-trained CNN models. Table IV lists only the best hyperparameters that perform well for individual pre-trained CNN models for flood image classification.

However, we have experimented with all the possible combinations of the hyperparameters. The ResNet101V2 model yields the best accuracy, with a learning rate of 0.0001 and an Adam optimizer of 86.87%. ResNet101v2 benefits from residual connections, which helps in effective training. Also, the ability to learn from fine-grained details helped improved accuracy when compared to other models. Initially, the experiments are carried out without CBAM for individual pre-trained CNN model ResNet101v2 with Adam optimizer and learning rate of 0.00001 obtained an accuracy of 86.87%. After applying the CBAM layer there was an improvement in performance wherein ResNet101v2 with Adam optimizer, learning rate of 0.01 obtained an accuracy of 88.60%.

TABLE IV. Results of Experiments Conducted on Individual Pre-trained CNN Models

	Model	Optimizer	Learning Rate	Training Accuracy (%)	Validation Accuracy (%)
Without CBAM	VGG16	Adadelta	0.01	84.88	83.28
	VGG19	Adam	0.01	84.41	82.66
	ResNet50	Adam	0.01	67.90	71.72
	XceptionNet	SGD	0.01	84.49	83.44
	MobileNetv2	Adadelta	0.01	89.45	85.83
	ResNet101v2	**Adam**	**0.0001**	**87.20**	**86.87**
	DenseNet201	Adam	0.01	87.48	85.00
	Inceptionv3	SGD	0.01	85.84	80.16
With CBAM	VGG16	Adadelta	0.01	86.35	85.10
	VGG19	Adam	0.001	85.00	84.35
	ResNet50	Adam	0.1	69.75	68.35
	XceptionNet	SGD	0.001	85.30	84.05
	MobileNetv2	Adadelta	0.1	89.50	86.15
	ResNet101v2	**Adam**	**0.01**	**89.35**	**88.60**
	DenseNet201	Adam	0.001	88.50	86.35
	Inceptionv3	SGD	0.1	86.24	81.75
	Inception-ResNet	Adam	0.001	87.10	81.35

Here the low learning rates such as 0.0001, were used to ensure stable convergence and avoiding for optimization. The lower learning rates require more iterations but they contribute to improved generalization. However, experiments were conducted with other learning rates too such as 0.01,0.1, etc. for classification without CBAM pre-trained model ResNet101v2 with Adam optimizer, learning rate of 0.00001 obtained slightly better accuracy of 86.87%.

B. Results of Hybrid CNN Models for Flood Image Classification

Out of the ten pre-trained models, the top five were selected based on their superior performance in previous experiments. Various combinations of these pre-trained and hybridized networks are followed by experiments utilizing different hyperparameter configurations. The top five results are shown in Table IV, with the hybrid model of MobileNetv2 and DenseNet201 achieving the highest accuracy of 85.83% with SGD optimizer and learning rate of 0.1. Followed by a hybrid model comprising VGG19 and ResNet101v2, it obtained an accuracy of 85.78% for 50 epochs with SGD optimizer and a learning rate of 0.1.

Table V shows the results of experiments performed for hybrid CNN models with CBAM. The best-performing individual models are hybridized. CBAM is added to pre-trained CNN models, allowing fine-tuning to benefit from the attention mechanism of spatial attention, which helps to identify flooded critical regions. The channel attention highlights features like water texture or patterns aiding better classification accuracy.

Among these hybrid models, the performance of MobileNetv2 and DenseNet201 with SGD optimizer learning rate of 0.01 obtained an accuracy of 90.54% after applying CBAM.

TABLE V. RESULTS OF HYBRID CNN MODEL FOR SATELLITE FLOOD IMAGE CLASSIFICATION

	Model	Optimizer	Learning Rate	Training Accuracy (%)	Validation Accuracy (%)
Without CBAM	**Mobilenetv2 and DenseNet201**	**SGD**	**0.1**	**87.76**	**85.83**
	ResNet50 and DenseNet201	Adam	0.1	89.19	83.91
	VGG19 and DenseNet201	Adam	0.1	89.13	83.44
	VGG19 and ResNet101v2	SGD	0.1	88.62	85.78
	ResNet101v2 and DenseNet201	SGD	0.001	86.56	84.84
With CBAM	**Mobilenetv2 and DenseNet201**	**SGD**	**0.01**	**95.36**	**90.54**
	ResNet50 and DenseNet201	Adam	0.001	89.85	86.53
	VGG19 and DenseNet201	Adam	0.1	89.31	85.50
	VGG19 and ResNet101v2	Adadelta	0.01	89.45	86.30
	ResNet101v2 and DenseNet201	Adam	0.1	89.31	85.50

C. Results of Sparse Swin Transformer

Table VI shows the experiments that are carried out with varying learning rates of 0.001, 0.01, 0.1 and different optimizers such as Adam. SGD and Adadelta optimizer. Only the best-performing results for image classification are listed. Initially, the experiments were performed without CBAM for the Adadelta optimizer with a learning rate of 0.001, batch size of 32, and number of epochs as 100, obtaining an accuracy of 71.35%.

TABLE VI. RESULTS OF SPARSE SWIN TRANSFORMER FOR IMAGE CLASSIFICATION

	Model	Optimizer	Learning Rate	Training Accuracy (%)	Validation Accuracy (%)
Without CBAM	Sparse Swin Transformer	Adam	0.001	53.28	50.48
	Sparse Swin Transformer	Adam	0.01	60.15	59.45
	Sparse Swin Transformer	SGD	0.01	68.22	64.44
	Sparse Swin Transformer	SGD	0.1	73.55	69.75
	Sparse Swin Transformer	**Adadelta**	**0.001**	**72.15**	**71.35**
	Sparse Swin Transformer	Adadelta	0.01	66.57	64.39
With CBAM	Sparse Swin Transformer	Adam	0.01	70.26	68.89
	Sparse Swin Transformer	**SGD**	**0.001**	**93.40**	**89.10**
	Sparse Swin Transformer	SGD	0.01	91.30	86.70
	Sparse Swin Transformer	SGD	0.1	90.20	82.35
	Sparse Swin Transformer	Adadelta	0.001	85.35	81.65
	Sparse Swin Transformer	Adadelta	0.01	84.94	80.00

After applying CBAM to the Sparse Swin transformer the improved results were obtained for the Adam optimizer with a learning rate of 0.001, obtaining an overall accuracy of 89.10%. The improved satellite flood image classification performance by leveraging its sparse attention mechanism significantly reduces computational overhead while maintaining accuracy. The hierarchical architecture of the Sparse Swin Transformer allowed for multiscale feature extraction, enhancing its ability to analyze local and global satellite imagery patterns.

D. Results of Data Efficient Image Transformer (DeiT) for Flood Image Classification

Table VII shows the experiments that are carried out with varying learning rates of 0.001,0.01,0.1 and different optimizers such as Adam. SGD and Adadelta optimizer. Only the best-performing results for image classification are listed. Among the experiments performed, improved results were obtained for the SGD optimizer with a learning rate of 0.1, obtaining an overall accuracy of 84.63%.

TABLE VII. RESULTS OF DEIT TRANSFORMER FOR IMAGE CLASSIFICATION

	Model	Optimizer	Learning Rate	Training Accuracy (%)	Validation Accuracy (%)
Without CBAM	DeiT	Adam	0.001	72.04	67.14
	DeiT	Adam	0.01	68.22	64.44
	DeiT	Adam	0.1	58.43	56.48
	DeiT	SGD	0.001	60.50	58.30
	DeiT	SGD	0.01	62.05	60.38
	DeiT	SGD	0.1	65.75	63.58
	DeiT	Adadelta	0.001	67.05	65.35
	DeiT	Adadelta	0.01	69.05	70.43
	DeiT	**Adadelta**	**0.1**	**75.05**	**72.35**
With CBAM	DeiT	Adam	0.001	97.11	80.19
	DeiT	Adam	0.01	66.55	63.70
	DeiT	Adam	0.1	53.52	54.81
	DeiT	SGD	0.001	91.30	88.70
	DeiT	**SGD**	**0.01**	**93.40**	**89.10**
	DeiT	SGD	0.1	85.35	80.65
	DeiT	Adadelta	0.001	93.25	78.75
	DeiT	Adadelta	0.01	84.94	80.00
	DeiT	Adadelta	0.1	72.44	71.65

DeiT provides better image classification results by effectively leveraging its data-efficient training strategy and attention mechanism. DeiT's incorporation of distillation tokens further enhanced learning by providing additional supervision, leading to a better generalization of houses completely or partially surrounded by flood water.

E. Results of Multiscale Vision Transformer (MViT) using CBAM for Flood Image Classification

The Multiscale Vision Transformer (MViT) demonstrated its effectiveness in flood image classification by efficiently capturing both global and local features across varying scales. With its multiscale attention mechanisms and patch-based tokenization, the model achieved high accuracy, particularly in scenarios involving complex spatial patterns such as flooded regions. Initially the experiments are carried out without CBAM where the performance was slightly less and then further the CBAM is added to MviT to improve the performance of model.

Table VIIIshows the experiments performed for image classification on satellite images with varying learning rates of 0.001, 0.01, 0.1 etc., with different optimizers such as Adam, SGD, Adadelta optimizer with 100 epoch. Among the experiments performed without CBAM the accuracy obtained was better with Adadelta optimizer learning rate of 0.01, with accuracy of 71.10% whereas on applying the CBAM the performance was improved with a learning rate of 0.001, Adam optimizer, accuracy obtained was 85.65%.

TABLE VIII. RESULTS OF MULTISCALE VISION TRANSFORMER (MVIT) FOR FLOOD IMAGE CLASSIFICATION

	Model	Optimizer	Learning Rate	Training Accuracy (%)	Validation Accuracy (%)
Without CBAM	Multiscale Vision Transformer (MViT)	Adam	0.001	66.25	61.25
	Multiscale Vision Transformer (MViT)	Adam	0.01	71.06	69.35
	Multiscale Vision Transformer (MViT)	Adam	0.1	79.25	65.60
	Multiscale Vision Transformer (MViT)	SGD	0.001	69.35	54.05
	Multiscale Vision Transformer (MViT)	SGD	0.01	70.50	68.52
	Multiscale Vision Transformer (MViT)	SGD	0.1	59.30	57.35
	Multiscale Vision Transformer (MViT)	Adadelta	0.001	62.60	55.20
	Multiscale Vision transformer (MviT)	**Adadelta**	**0.01**	**75.54**	**71.10**
	Multiscale Vision Transformer (MViT)	Adadelta	0.1	73.51	67.35
With CBAM	**Multiscale Vision Transformer (MViT)**	**Adam**	**0.001**	**88.32**	**85.65**
	Multiscale Vision Transformer (MViT)	Adam	0.01	87.65	83.50
	Multiscale Vision Transformer (MViT)	Adam	0.1	85.45	81.15
	Multiscale Vision Transformer (MViT)	SGD	0.001	77.31	75.89
	Multiscale Vision Transformer (MViT)	SGD	0.01	85.42	79.32
	Multiscale Vision Transformer (MViT)	SGD	0.1	83.19	80.15
	Multiscale Vision Transformer (MViT)	Adadelta	0.001	89.22	76.24
	Multiscale Vision Transformer (MViT)	Adadelta	0.01	92.77	75.25
	Multiscale Vision Transformer (MViT)	Adadelta	0.1	97.51	88.35

F. Results of Vision Transformer for Flood Image Classification

Table IX shows the experiments performed for image classification on satellite images with varying learning rates of 0.001, 0.01, 0.1 etc., with different optimizers such as Adam, SGD, Adadelta optimizer. For the initial experiments for Vision transformer without CBAM it was found using SGD optimizer, learning rate of 0.01, obtained an accuracy of 73.08%. The improved performance for Vision transformer, which performed well for a learning rate of 0.01 with the Adadelta optimizer, obtained an accuracy of 90.75% for 100 epochs with CBAM.

TABLE IX. RESULTS OF VISION TRANSFORMER FOR IMAGE CLASSIFICATION

	Model	Optimizer	Learning Rate	Training Accuracy (%)	Validation Accuracy (%)
Without CBAM	Vision Transformer	Adam	0.001	65.07	62.41
	Vision Transformer	Adam	0.01	67.29	66.67
	Vision Transformer	Adam	0.1	63.42	61.30
	Vision Transformer	SGD	0.001	67.40	58.97
	Vision Transformer	**SGD**	**0.01**	**75.38**	**73.08**
	Vision Transformer	SGD	0.1	77.08	71.21
	Vision Transformer	Adadelta	0.001	79.87	70.43
	Vision Transformer	Adadelta	0.01	68.75	65.69
	Vision Transformer	Adadelta	0.1	73.21	66.63
With CBAM	Vision Transformer	Adam	0.001	92.61	84.58
	Vision Transformer	Adam	0.01	91.06	79.87
	Vision Transformer	Adam	0.1	89.73	81.21
	Vision Transformer	SGD	0.001	92.89	82.56
	Vision Transformer	SGD	0.01	93.97	80.94
	Vision Transformer	SGD	0.1	94.28	79.83
	Vision Transformer	Adadelta	0.001	93.97	80.94
	Vision Transformer	**Adadelta**	**0.01**	**93.94**	**91.75**
	Vision Transformer	Adadelta	0.1	87.62	83.00

The CBAM added to the Vision transformer (ViT) significantly improved performance enabling the model to focus on the most relevant features in terms of spatial and channel-wise attention.

G. Results of Hierarchical Vision Transformer

There are a number of experiments used for flood house image classification with Hierarchical Vision Transformer (HviT) wherein Table X shows the experiments that are performed with varying learning rates of 0.001, 0.001, 0.1 with different optimizers such as Adam, SGD, and Adadelta optimizers;

TABLE X. RESULTS OF HIERARCHICAL VISION TRANSFORMER FOR FLOOD IMAGE CLASSIFICATION

	Model	Optimizer	Learning Rate	Training Accuracy (%)	Validation Accuracy (%)
Without CBAM	Hierarchical Vision Transfomer	Adam	0.001	68.89	67.75
	Hierarchical Vision Transformer	Adam	0.01	50.00	49.62
	Hierarchical Vision Transformer	Adam	0.1	54.16	51.26
	Hierarchical Vision Transformer	SGD	0.001	68.35	64.15
	Hierarchical Vision Transformer	SGD	0.01	61.73	59.45
	Hierarchical Vision Transformer	SGD	0.1	63.25	60.75
	Hierarchical Vision Transformer	Adadelta	0.001	78.30	75.00
	Hierarchical Vision Transformer	**Adadelta**	**0.01**	**80.10**	**76.80**
	Hierarchical Vision Transformer	Adadelta	0.1	74.20	71.00
With CBAM	Hierarchical Vision Transformer	Adam	0.001	87.25	70.85
	Hierarchical Vision Transformer	Adam	0.01	89.56	78.50
	Hierarchical Vision Transformer	Adam	0.1	95.62	83.00
	Hierarchical Vision Transformer	**SGD**	**0.001**	**89.31**	**87.50**
	Hierarchical Vision Transformer	SGD	0.01	92.56	81.50
	Hierarchical Vision Transformer	SGD	0.1	95.62	83.00
	Hierarchical Vision Transformer	Adadelta	0.001	84.94	80.00
	Hierarchical Vision Transformer	Adadelta	0.01	87.62	85.15
	Hierarchical Vision Transformer	Adadelta	0.1	94.12	82.25

Only the best-performing results are listed. The best performance for classification using HviT. Among these the performance of image classification was found to be better with SGD optimizer, learning rate of 0.001 obtained an accuracy of 87.50%. Adding CBAM to the Hierarchical Vision Transformer enhances the model's ability to focus on relevant features by refining both spatial and channel-level attention. This results in improved feature representation, allowing the model to better capture important flood-related patterns and improve classification accuracy.

H. Results of SWIN Transformer using CBAM for Flood Image Classification

Table XI shows the experiments which are carried out with varying learning rates of 0.001, 0.01, 0.1 and different optimizers such as Adam. SGD and Adadelta optimizer. Only the best-performing results for image classification are listed. Among the experiments performed, the improved results were obtained for the Adam optimizer with a learning rate of 0.001, obtained an overall accuracy of 85.35%.

TABLE XI. RESULTS OF SWIN TRANSFORMER FOR FLOOD IMAGE CLASSIFICATION

	Model	Optimizer	Learning Rate	Training Accuracy (%)	Validation Accuracy (%)
Without CBAM	Swin Transformer	Adam	0.003	70.32	65.20
	Swin Transformer	**Adam**	**0.02**	**73.84**	**72.04**
	Swin Transformer	Adam	0.01	64.36	60.56
	Swin Transformer	SGD	0.3	70.32	65.20
	Swin Transformer	SGD	0.2	64.36	60.56
	Swin Transformer	SGD	0.001	55.36	53.39
	Swin Transformer	Adadelta	0.03	65.35	62.10
	Swin Transformer	Adadelta	0.002	67.40	58.97
	Swin Transformer	Adadelta	0.001	65.30	60.75
With CBAM	Swin Transformer	Adam	0.3	78.60	68.35
	Swin Transformer	Adam	0.002	76.53	72.52
	Swin Transformer	Adam	0.01	74.30	70.35
	Swin Transformer	SGD	0.3	77.50	71.00
	Swin Transformer	SGD	0.2	86.12	79.30
	Swin Transformer	**SGD**	**0.01**	**85.46**	**81.69**
	Swin Transformer	Adadelta	0.003	74.08	72.05
	Swin Transformer	Adadelta	0.02	75.42	61.35
	Swin Transformer	Adadelta	0.1	78.35	71.05

The improved performance of the Swin transformer with the CBAM layer is due to the ability of the SWIN transformer to capture long-range dependencies with spatial regions of image such as flooded houses i.e. completely surrounded houses or partially surrounded houses by flood water.

I. Performance Comparison of Models for Flood Image Classification

Table XII Shows the overall comparison of various experiments performed with varying learning rates of 0.001,0.01, and 0.1, with different optimizers such as Adam, SGD, and Adadelta optimizers, respectively it was found that the performance of Vision transformer with a learning rate of 0.01, Adadelta optimizer obtained a better accuracy of 90.75%. This improved performance of Vision transformer with CBAM is as a result of the ability of the Vision transformer to capture intricate regions.

TABLE XII. SUMMARY OF PERFORMANCE COMPARISON FOR VARIOUS MODELS

Model	Optimizer	Learning rate	Training Accuracy (%)	Validation Accuracy (%)
ResNet101v2	Adam	0.0001	87.20	86.87
MobileNetv2 [29]	Adam	0.1	94.23	75.00
MobileNetv2 and DenseNet201	SGD	0.01	95.36	89.54
Sparse Swin Transformer	SGD	0.001	93.40	89.10
DeiT	SGD	0.1	86.35	84.63
MViT	Adam	0.0001	88.32	85.65
Hierarchical Vision Transformer	SGD	0.001	89.31	87.50
Vision transformer	**Adadelta**	**0.01**	**93.94**	**90.75**
Swin transformer	Adam	0.01	75.60	72.52

Sparse Swin Transformer is highly efficient for flood image classification due to its sparse attention mechanism and hierarchical design, enabling effective analysis of high-resolution images with localized and global patterns. Hierarchical Vision Transformer (HVT) captures multi-scale features, making it suitable for identifying fine details, such as partially submerged areas, and broader flood zones. DeiT is ideal for scenarios with limited labeled flood image datasets, leveraging data-efficient training and compact architecture to achieve high accuracy. Multiscale Vision Transformer (MViT) balances computational cost and performance with its multi-scale attention mechanism, effectively classifying diverse flood scenarios. Hybrid CNN models combine the strengths of multiple architectures and integrate CBAM for refined spatial and channel-wise feature extraction, offering robust generalization on complex flood datasets. In contrast, individual pre-trained models, such as ResNet and MobileNet, provide strong baseline performance and quick adaptability, making them suitable for resource-constrained environments or binary flood/non-flood classification tasks. Each model brings unique strengths, enabling tailored solutions for diverse flood image classification challenges.

V. DISCUSSION

ResNet101v2 outperformed other models due to its skip connections, which effectively help deep networks learn residual functions, enabling better training and generalization. Hybrid CNN models like MobileNetV2-DenseNet201 also performed well, leveraging MobileNetV2's efficient architecture and DenseNet201's feature reuse capability. Transformer-based models such as DeiT, MViT, Swin Transformer, and Hierarchical Vision Transformer (HVT) excelled in flood image classification by capturing long-range dependencies and multi-scale features, making them particularly effective for satellite imagery. Incorporating CBAM in ViT and Swin Transformer further improved accuracy by enhancing important spatial and channel-wise features, helping distinguish flood-specific patterns like water levels and house surroundings. Overall, transformer-based models, especially ViT with CBAM, outperformed CNNs by focusing on global features and improving flood classification accuracy.

The effectiveness of CBAM lies in its ability to adaptively refine feature maps, emphasizing critical flood-specific details while suppressing irrelevant information. Traditional CNNs process all features uniformly, which may lead to misclassification in complex flood scenarios. In contrast, models with CBAM enhance feature discrimination by focusing on water texture, surrounding structures, and flood extent, resulting in better classification of houses as completely or partially submerged. Advanced transformer-based models like Sparse Swin Transformer and Hierarchical Vision Transformer with CBAM further refine this process, making them superior to conventional CNNs and hybrid models in flood image classification.

VI. CONCLUSION AND FUTURE SCOPE

This article systematically evaluates various pre-trained CNN architectures and transformer models for satellite flood image classification, specifically identifying houses as completely or partially surrounded by floodwater. The fine-tuning of hyperparameters and hybridizing top-performing architectures with vision transformer modules, we achieved significant improvements in classification accuracy. Among CNN models, ResNet101V2 demonstrated the highest accuracy of 86.87%, while a hybrid CNN combining MobileNetV2 and DenseNet201 reached 85.83%, further improving to 90.54% with CBAM integration. Transformer-based models also performed well, with Vision Transformer achieving 91.75% accuracy, Sparse Swin Transformer reaching 89.10%, and DeiT obtaining 84.63%. The key takeaway from this work is the

integrating CBAM with hybrid CNN architectures and leveraging transformer-based models significantly enhances flood classification accuracy in satellite imagery. These findings can aid disaster response teams in prioritizing affected areas and improving flood impact assessment through flood image classification. Future work can focus on expanding the dataset to improve model generalization and adapting these models for different types of satellite flood imagery to enhance their applicability across diverse disaster scenarios.

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] National Disaster Management Authority (NDMA), Floods: Natural Hazards, *ndma floods*, https://ndma.gov.in/Natural-Hazards/Floods.

[3] R. Colacicco, A. Refice, R. Nutricato, F. Bovenga, G. Caporusso, A. D'Addabbo, M. La Salandra, F. P. Lovergine, D. O. Nitti, and D. Capolongo, "High-Resolution Flood Monitoring Based on Advanced Statistical Modeling of Sentinel-1 Multi-Temporal Stacks," *Remote Sensing*, vol. 16, no. 2, p. 294, 2024. doi:https://doi.org/10.3390/rs16020294.

[4] Wu, Luyuan, Jingbo Tong, Zifa Wang, Jianhui Li, Meng Li, Hui Li, and Yi Feng. "Post-flood disaster damaged houses classification based on dual-view image fusion and Concentration-Based Attention Module." Sustainable Cities and Society 103 (2024): 105234. doi: https://doi.org/10.1016/j.scs.2024.105234

[5] Z. Wu, Z. Dong, K. Yang, Q. Liu, and W. Wang, "Floodwater Extraction from UAV Orthoimagery Based on a Transformer Model," *Remote Sens.*, vol. 16, no. 21, p. 4052, 2024, doi: https://doi.org/10.3390/rs16214052

[6] H. Farhadi, A. Esmaeily, and M. Najafzadeh, "Flood monitoring by integration of remote sensing technique and multi-criteria decision making method," *Computers & Geosciences*, vol. 160, p. 105045, 2022.

[7] D. Amitrano, G. Di Martino, A. Di Simone, and P. Imperatore, "Flood detection with SAR: A review of techniques and datasets," *Remote Sensing*, vol. 16, no. 4, p. 656, 2024. doi: https://doi.org/10.3390/rs16040656.

[8] K. Vashist and K. K. Singh, "Flood hazard mapping using GIS-based AHP approach for Krishna River basin," *Hydrological Processes*, vol. 38, no. 6, p. e15212, 2024. doi:https://doi.org/10.1002/hyp.15212

[9] D. Tadesse, K. V. Suryabhagavan, D. Nedaw, and B. T. Hailu, "A model-based flood hazard mapping in Itang district of the Gambella region, Ethiopia," *Geology, Ecology, and Landscapes*, vol. 8, no. 1, pp. 8–25, 2024.doi:https://doi.org/10.1080/24749508.2021.2022833.

[10] S. S. Rana, S. A. Habib, M. N. H. Sharifee, N. Sultana, and S. H. Rahman, "Flood risk mapping of the flood-prone Rangpur Division of Bangladesh using remote sensing and multi-criteria analysis," *Natural Hazards Research*, vol. 4, no. 1, pp. 20–31, 2024, doi:https://doi.org/10.1016/j.nhres.2023.09.012.

[11] F. Y. Dtissibe, A. A. A. Ari, H. Abboubakar, A. N. Njoya, A. Mohamadou, and O. Thiare, "A comparative study of machine learning and deep learning methods for flood forecasting in the Far North Region, Cameroon," *Scientific African*, vol. 23, p. e02053, 2024, doi: https://doi.org/10.1016/j.sciaf.2023.e02053.

[12] Y. D. Jhong, C. S. Chen, B. C. Jhong, C. H. Tsai, and S. Y. Yang, "Optimization of LSTM parameters for flash flood forecasting using genetic algorithm," *Water Resources Management*, vol. 38, no. 3, pp. 1141–1164, 2024, doi:https://doi.org/10.1007/s11269-023-03713-8.

[13] A. Matsuki and M. Hatayama, "Risk analysis of mutual influence relationships among residents under rescue operations in long-term flooded areas," *International Journal of Disaster Risk Reduction*, vol. 100, p. 104216, 2024, doi: https://doi.org/10.1016/j.ijdrr.2023.104216.

[14] P. U. Nehete, D. S. Dharrao, P. Pise, and A. Bongale, "Object detection and classification in human rescue operations: Deep learning strategies for flooded environments," *International Journal of Safety & Security Engineering*, vol. 14, no. 2, 2024, doi: https://doi.org/10.18280/ijsse.140226.

[15] T. Saleh, S. Holail, X. Xiao, and G. S. Xia, "High-precision flood detection and mapping via multi-temporal SAR change analysis with semantic token-based transformer," *International Journal of Applied Earth Observation and Geoinformation*, vol. 131, p. 103991, 2024, doi: https://doi.org/10.1016/j.jag.2024.103991.

[16] I. 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: https://doi.org/10.3390/w16121670.

[17] R. G. Franceschini, J. Liu, and S. Amin, "Damage estimation and localization from sparse aerial imagery," in *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 128–134, IEEE, 2021, doi:https://doi.org/10.1109/ICMLA52953.2021.00028.

[18] M. A. Islam, S. I. Rashid, N. U. I. Hossain, R. Fleming, and A. Sokolov, "An integrated convolutional neural network and sorting algorithm for image classification for efficient flood disaster management," *Decision Analytics Journal*, vol. 7, p. 100225, 2023, doi: https://doi.org/10.1016/j.dajour.2023.100225.

[19] J. Jackson, S. B. Yussif, R. A. Patamia, K. Sarpong, and Z. Qin, "Flood or non-flooded: A comparative study of state-of-the-art models for flood image classification using the FloodNet dataset with uncertainty offset analysis," *Water*, vol. 15, no. 5, p. 875, 2023, doi: https://doi.org/10.3390/w15050875

[20] F. Pech-May, J. V. Sanchez-Hernández, L. A. López-Gómez, J. Magaña-Govea, and E. M. Mil-Chontal, "Flooded areas detection through SAR images and U-Net deep learning model," *Computación y Sistemas*, vol. 27, no. 2, pp. 449–458, 2023, doi: https://doi.org/10.13053/cys-27-2-4624.

[21] F. Montello, E. Arnaudo, and C. Rossi, "MMFlood: A multimodal dataset for flood delineation from satellite imagery," *IEEE Access*, vol. 10, pp. 96774–96787, 2022, doi: https://doi.org/10.1109/ACCESS.2022.3205419

[22] J. Ha and J. E. Kang, "Assessment of flood-risk areas using random forest techniques: Busan metropolitan city," *Natural Hazards*, pp. 1–23, 2022, doi: https://doi.org/10.1007/s11069-021-05142-5.

[23] I. Bouchard, M. E. Rancourt, D. Aloise, and F. Kalaitzis, "On transfer learning for building damage assessment from satellite imagery in emergency contexts," *Remote Sensing*, vol. 14, no. 11, p. 2532, 2022, doi:https://doi.org/10.3390/rs14112532.

[24] Y. Shen, S. Zhu, T. Yang, C. Chen, D. Pan, J. Chen, L. Xiao, and Q. Du, "BDANet: Multiscale convolutional neural network with cross-directional attention for building damage assessment from satellite images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2021, doi: https://doi.org/10.1109/TGRS.2021.3080580

[25] X. Jiang, S. Liang, X. He, A. D. Ziegler, P. Lin, M. Pan, D. Wang, J. Zou, D. Hao, G. Mao, et al., "Rapid and large-scale mapping of flood inundation via integrating spaceborne synthetic aperture radar imagery with unsupervised deep learning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 178, pp. 36–50, 2021, doi: https://doi.org/10.1016/j.isprsjprs.2021.05.019

[26] J. M. A. Opella and A. A. Hernandez, "Developing a flood risk assessment using support vector machine and convolutional neural network: A conceptual framework," in *2019 IEEE 15th International Colloquium on Signal Processing & Its Applications (CSPA)*, pp. 260–265, IEEE, 2019, doi: https://doi.org/10.1109/CSPA.2019.8695980.

[27] C. Sarker, L. Mejias, F. Maire, and A. Woodley, "Flood mapping with convolutional neural networks using spatio-contextual pixel information," *Remote Sensing*, vol. 11, no. 19, p. 2331, 2019, doi:https://doi.org/10.3390/rs11192331.

[28] L. Moya, H. Zakeri, F. Yamazaki, W. Liu, E. Mas, and S. Koshimura, "3D gray level co-occurrence matrix and its application to identifying collapsed buildings," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 149, pp. 14–28, 2019, doi:https://doi.org/10.1016/j.isprsjprs.2019.01.008.

[29] S. S. Kulkarni and A. Mahapatra, "Post flood assessment using deep learning techniques," in *AIP Conference Proceedings*, vol. 2917, AIP Publishing, 2023, doi: https://doi.org/10.1063/5.0175612

[30] R. Gupta, R. Hosfelt, S. Sajeev, N. Patel, B. Goodman, J. Doshi, E. Heim, H. Choset, and M. Gaston, "XBD: A dataset for assessing building damage from satellite imagery," *arXiv preprint arXiv:1911.09296*, pp. 1–9, 2019.

[31] N. Kaur, C. C. Lee, A. Mostafavi, and A. Mahdavi-Amiri, "Large-scale building damage assessment using a novel hierarchical transformer architecture on satellite images," *Computer-Aided Civil and Infrastructure Engineering*, vol. 38, no. 15, pp. 2072–2091, 2023, doi: https://doi.org/10.1111/mice.12981.

[32] M. Rahnemoonfar, T. Chowdhury, A. Sarkar, D. Varshney, M. Yari, and R. R. Murphy, "FloodNet: A high resolution aerial imagery dataset for post flood scene understanding," *IEEE Access*, vol. 9, pp. 89644–89654, 2021, doi: https://doi.org/10.1109/ACCESS.2021.3090981.

[33] D. F. Muñoz, P. Muñoz, H. Moftakhari, and H. Moradkhani, "From local to regional compound flood mapping with deep learning and data fusion techniques," *Science of the Total Environment*, vol. 782, p. 146927, 2021, doi: https://doi.org/10.1016/j.scitotenv.2021.146927.

[34] B. T. Pham, C. Luu, D. V. Dao, T. V. Phong, H. D. Nguyen, H. V. Le, J. von Meding, and I. Prakash, "Flood risk assessment using deep learning integrated with multi-criteria decision analysis," *Knowledge-Based Systems*, vol. 219, p. 106899, 2021, doi: https://doi.org/10.1016/j.knosys.2021.106899.

[35] H. Wu, H. Song, J. Huang, H. Zhong, R. Zhan, X. Teng, Z. Qiu, M. He, and J. Cao, "Flood detection in dual-polarization SAR images based on multi-scale DeepLab model," *Remote Sensing*, vol. 14, no. 20, p. 5181, 2022, doi: https://doi.org/10.3390/rs14205181.

[36] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, "Training Data-Efficient Image Transformers & Distillation Through Attention," in *Proc. Int. Conf. Mach. Learn. (ICML)*, PMLR, 2021, pp. 10347–10357.

[37] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2021, pp. 10012–10022. doi: https://doi.ieeecomputersociety.org/10.1109/ICCV48922.2021.00986.

[38] K. Pinasthika, B. S. P. Laksono, R. B. P. Irsal, N. Yudistira, *et al.*, "SparseSwin: Swin Transformer with Sparse Transformer Block," *Neurocomputing*, vol. 580, p. 127433, 2024, doi:https://doi.org/10.1016/j.neucom.2023.127433.

[39] X. Zhang, Y. Tian, L. Xie, W. Huang, Q. Dai, Q. Ye, and Q. Tian, "HiViT: A Simpler and More Efficient Design of Hierarchical Vision Transformer," in *Proc. 11th Int. Conf. Learn. Represent. (ICLR)*, 2023.

[40] A. Dosovitskiy, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *arXiv preprint arXiv:2010.11929*, 2020.

[41] M. H. M. Ali, S. A. Asmai, Z. Z. Abidin, Z. A. Abas, and N. A. Emran, "Flood Prediction using Deep Learning Models," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 9, 2022, doi: http://dx.doi.org/10.14569/IJACSA.2022.01309112.

[42] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3–19, 2018.