Deep Learning and Explainable AI for Aerial Image based Flood Damage Detection

Ramachandiran R.
Computational Intelligence
SRMIST University Chennai, India

A. Sanjay
Computational Intelligence
SRMIST University Chennai, India

Gajulapalli Venkata Koushik
Computational Intelligence
SRMIST University Chennai, India

ABSTRACT
Disasters like floods are normally characterized by colossal losses in terms of human lives and property. Aerial photography is one of the extremely important tools in the evaluation of the damages and the identification of the impacted areas in the occurrence of floods. This issue is on the classification of aerial images in the postflood environment, and the images will be partitioned into varying areas with a great deal of accuracy using the aids of some of the deep learning networks, namely MobileNet and DenseNet. The objective of all the models is to classify the pictures into the six possible classes—buildings, flooded areas, forests, mountains, seas, and streets. It has Explainable Artificial Intelligence (XAI) technologies along with every model to ensure that there are such interpretability and transparency. The method employed is GradCAMgenerated Class Activation Maps. The result indicates that the deep learning models can distinguish the flooded areas among the rest of the critical elements in the postflood assessment and recovery planning. Besides, GradCAM interpretability enhances the model prediction's reliability, which is a key element in the disaster mitigation and response interventions in real time. The performance evaluation of the given model is conducted on the basis of multiple measurements, and it is determined that this model is an addition to the already existing literature, which is geared towards the application of AIbased solutions that can support the process of postdisaster recovery.

General Terms
Deep Learning, Computer Vision, Explainable Artificial Intelligence (XAI), Image Classification, Flood Detection, Convolutional Neural Networks (CNN), DenseNet121, MobileNet, GradCAM, Disaster Management, Pattern Recognition, Remote Sensing, Aerial Image Analysis, Machine Learning, Feature Extraction.

Keywords
Aerial Image Classification, PostFlood Analysis, MobileNet, DenseNet, Explainable Artificial Intelligence (XAI), GradCAM, Deep Learning, Flooded Areas, Damage Assessment, Disaster Management, Image Classification, Computer Vision, Remote Sensing, Disaster Recovery, and Class Activation Mapping.

1. INTRODUCTION
Floods are among the most devastating natural disasters, causing widespread destruction to infrastructure, the environment, and, tragically, human lives. The timely and accurate assessment of the damage is one of the most important factors in the whole process of disaster response and recovery. The traditional postflood management instruments have always relied on manual inspection, and this is a very slow, labor intensive, and ineffective means of inspecting the entire facility in the entire face of the flood cover. Another perspective and easier way of assessing the damage on a greater scale can also be offered by the aerial view with the assistance of a drone or satellite, but the interpretation of the pictures will also be highly challenging because of the overwhelming volumes of data and the need to differentiate the flooded areas from the rest of the landscape objects [1].To overcome these difficulties, the project proposes to use deep learning (DL) in the classification of aerial images following floods in an automatic manner. The two specific convolutional neural net (CNN) models, namely DenseNet and MobileNet, are used to encode the images into six major categories, i.e., building, flooded, forest, mountains, sea, and street. MobileNet is used because of its speed and the suitability of reallife application, whereas DenseNet is used due to its accuracy, and its features can be applied elsewhere [2].

In order to train the model, it is trained using datasets of aerial images depending on their postflood condition to ensure that the model can adapt to a high probability of environments. One of the greatest weaknesses of deep learning models in the case of disaster management is the lack of transparency of the decisionmaking mechanism, which is typically called the "blackbox nature." The mistrust can be caused by the nonobservable models among the disaster management workers and policymakers. The Gradientweighted Class Activation Mapping (GradCAM) is one of the Explainable AI (XAI) methods that we are similarly applying to this problem. GradCAM uses heatmaps of the visual representations to indicate the most meaningful parts of the output of the model being predicted [3].In this instance, it will make sure that not only the classifications are correct but also communicable; therefore, humans will be able to prove and improve AIdriven assessments.The paper also adds to the constantly developing branch of AI in disaster management and an approach that integrates the strengths of automation, accuracy, and interpretability. The implemented system will greatly accelerate and positively improve the postflood assessment and thus benefit the government, the nongovernmental organizations, and the emergency agencies in terms of resource allocation planning, operation priorities [4], and reconstruction. Realtime drone imaging of the infield could be applied in the initial and subsequent phases of the project in incorporation with geographical information systems and others.

2. RELATED WORK
Postflood damage assessment has long relied on remote sensing as a primary data source. Satellite and aerial imagery provide synoptic views of affected regions that ground surveys cannot match in either speed or spatial coverage [5]. Traditional approaches to interpreting this imagery—including indexbased methods and rulebased classifiers—have progressively been supplanted by datadriven deep learning models, which learn discriminative features directly from labeled training data. CNNbased architectures have been applied extensively to flood detection using satellite data. Ghosh et al. [2] demonstrated the effectiveness of deep learning for automatic flood detection from Sentinel1 synthetic aperture radar imagery, while Rahnemoonfar et al. [1] introduced the FloodNet dataset, a

highresolution aerial benchmark specifically designed for postflood scene understanding. Gebrehiwot et al. [5] applied deep CNNs to UAVacquired imagery for flood extent mapping, achieving strong results even with limited labeled training data.

Among CNN architectures, MobileNet and DenseNet have attracted particular attention for remote sensing applications. MobileNet's lightweight design enables inference on embedded hardware with minimal latency, a critical requirement for UAVbased realtime applications. DenseNet's dense interlayer connectivity encourages feature reuse across all network depths, mitigating the vanishing gradient problem and yielding richer representations of complex imagery [6][7]. The integration of XAI methods into flood detection workflows has also gained traction. Takato [3] applied GradCAM to classify disasteraffected regions in typhoon imagery, demonstrating that saliency maps can identify structurally damaged zones with high spatial fidelity. Chen et al. [4] incorporated GradCAM into a lightweight CNN designed for flood height prediction from radar echoes, showing that visual explanations improve both model transparency and operational trust. Nam et al. [8] extended XAI to flood susceptibility mapping in Seoul, using Bayesian AutoML with evolutionary optimization to balance predictive accuracy with interpretability.

Liao et al. [10] proposed Feature Activation Maps as a generalization of class activation mapping, providing finergrained localization of discriminative features in image classification tasks. Collectively, these studies establish that combining CNNbased classification with XAI not only improves prediction reliability but also enables domain experts to audit and validate model behavior—an essential requirement in highstakes disaster management contexts [9].

3. METHODOLOGY

The proposed system integrates deep learningbased image classification with explainability mechanisms to support automated flood damage assessment from aerial imagery. The pipeline consists of four stages: data collection and preprocessing, model training and selection, performance evaluation, and GradCAM visualization. Figure 1 illustrates the overall system architecture.

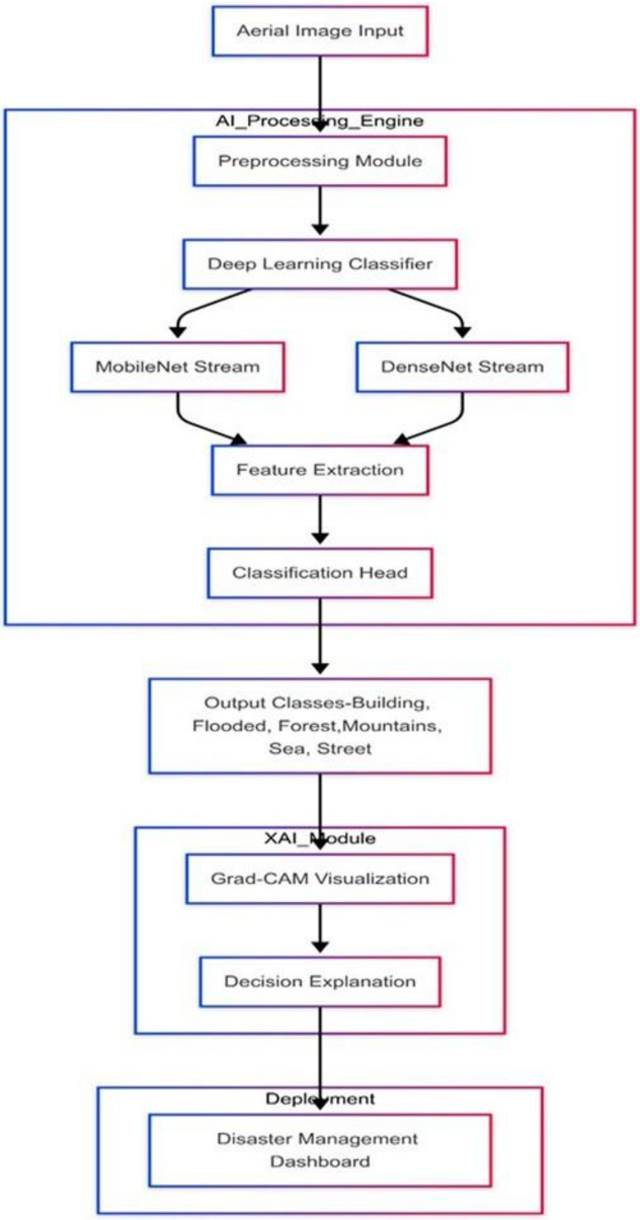

Fig. 1. Architecture of Methodology

3.1 Data Collection and Preprocessing

A dataset of aerial images representing six postflood landcover classes—buildings, flooded areas, forests, mountains, seas, and streets—was assembled from satellite and UAV sources spanning multiple geographic regions. Each class contained 150 test samples, yielding a total evaluation set of 900 images. Preprocessing operations applied to all images included resizing to a uniform resolution, normalization of pixel intensity values to the [0, 1] range, and data augmentation through random rotation, horizontal flipping, and scaling. Augmentation was employed to increase training set diversity and reduce overfitting on underrepresented classes.

3.2 Model Selection and Training

Two CNN architectures were selected for this study based on their complementary characteristics. MobileNet uses depthwise separable convolutions to achieve a significantly reduced parameter count relative to standard convolutions, enabling efficient inference on resourceconstrained platforms. DenseNet connects each layer to every subsequent layer within a dense block, maximizing gradient flow during backpropagation and enabling extensive feature reuse across network depths. Both models were initialized with ImageNet pretrained weights and subsequently finetuned on the floodspecific dataset using categorical crossentropy loss and gradientbased optimization. The dataset was partitioned into training, validation, and test subsets to permit unbiased performance estimation.

3.3 Performance Evaluation

Model performance was assessed using accuracy, precision, recall, and F1score computed over all six classes. These metrics collectively characterize both perclass and aggregate classification behavior. kfold crossvalidation was applied to assess generalization to unseen data and to detect potential overfitting on the training set.

3.4 Explainable AI via GradCAM

GradCAM generates classdiscriminative localization maps by computing the gradient of the predicted class score with respect to the feature maps of the final convolutional layer. These gradients are globally averagepooled to produce weights for each feature map channel; the weighted combination of feature maps, passed through a ReLU activation, yields a coarse heatmap that indicates which spatial regions contributed most to the prediction [7]. In the flood assessment context, these heatmaps allow stakeholders to verify that the model is responding to hydrologically relevant image features—such as surface water extent and submerged infrastructure—rather than spurious correlations.

4. IMPLEMENTATION

4.1 MobileNet

The general purpose of integrating MobileNet within the planned system of flood damage detection is to create a small and efficient model, which would classify the aerial images under the various sections whose fundamental goal is to detect the areas of damages caused by floods. MobileNet is an architecture based on deep learning but is optimized to run on mobile devices and in embedded systems; hence, it can offer the benefits of both together, i.e., high accuracy and utility. The fact that it could minimize the computation load and yet still have the accuracy made it highly desirable to realtime applications of the flooding detection in the scenario in which it is assumed that the capacity of the computing unit is constrained, just as in the situation of flying drones that are deployed to assist with the assessment of the disasterafflicted areas. The MobileNet architecture is very practical because it uses depthwise separable convolutions and requires the count of parameters and computational costs to reduce significantly in comparison to the traditional convolutional networks. One of the reasons why MobileNet can execute quickly in a case of very short time, such as flood detection of the overall scenario of disaster management, is the architecture. The resource need of the MobileNet is so minimal that the model can be used in low processing power devices as well; thus, it is more applicable to be implemented in the field where realtime decisionmaking will be required to effectively control the floods and allocate resources to the field.

MobileNet training starts by loading the model with the weights that have already been trained, which have been obtained in one of the massive image classification data sets, including ImageNet. This is then succeeded by refining the flood data set, which is the flood specific data set that includes aerial images that have been classified into six categories, namely building, flooded, forest, mountains, sea, and street. The finetuning entails the process of training the model using the few images that are labeled so that the model that has already been pretrained can be adjusted to the task of flooddetecting. One type of learning utilizes a learning methodology that oversees the learning process such that the images are entered into the network and the learning model modifies its weights by reducing the classification errors through the optimization of the backpropagation and gradient descent processes and thus learns through error.

The model will also be confused by the training data and will stop overfitting the model at the same time using the data augmentation methods such as rotation, flipping, and scaling. The performance of the evaluation model is evaluated based on popular indicators such as accuracy, precision, recall, and F1score. After the training, the MobileNet model is refined and tested on another validation set; hence, its effectiveness in classifying the aerial images to detect the damages of floods is validated, and quick and accurate predictions can be made that can be utilized in realtime to respond to the disaster.

4.2 DenseNet

It is also notable that DenseNet has been integrated in the suggested flood damage detection system since it improves the capability of the model to categorize aerial images with an extremely high degree of precision as a result of extraction of features. Everyone knows that the connection between the layers is very dense, which is what makes DenseNet, or Densely Connected Convolutional Networks. Each and every layer's input is not only supplied by the layer above but also by all the network layers that come before it. Therefore, this type of network connection is a massive opportunity for the model to make optimal use of the characteristics as well as the flow of gradients when the training is conducted and assisted with the issue of vanishing gradient that is largely observed in deep networks. The property of the network to store and transmit rich feature representation is worth considering in the adequate differentiation of flood damaged areas with other landscape attributes like buildings, roads, and trees in the flood damage detection case. The initial aim of the DenseNet adoption is the optimization of the classification that is received through the training of the complex aerial images, especially in the case of an extremely small difference between the areas covered by floods. DenseNet, in particular, is effective in depicting the smallest details in the image, which is the most important aspect in describing the flooding status that is partially submerged, surrounded with other environmental characteristics, or is covered in water of different degrees. This structural design helps the network to be concentrated on the important features even in the difficult image conditions, like the occurrence of clouds or poor light in the post flood scenario.

DenseNet of the developed flood damage detection system is an obligatory component of decreasing the extent of the model to the most appropriate stage of categorizing aerial images according to the most promising features. A convolutional network that is characterized by the extremely high level of interconnection of the layers is the relatively new and rather modern form of the DenseNet, or densely connected convolutional networks. It means that the input at each layer is not only entering directly at the specific layer but also at the previous layers of that network. Hence, the network connection of this nature gives the model a fantastic chance to successfully reuse the extracted features, and at the same time, the gradient flow is reinforced in the process of training, and as a result, the problem of vanishing gradients that usually occurs in the deep networks is alleviated. The maintenance and the fact that the model gives the opportunity to describe the rich features play a huge role in the flood damage detection scenario in the precise differentiation of the areas covered by the floods and the remainder of the features in the landscape, such as buildings, roads, and trees.

The application of the DenseNet is specifically oriented to the optimization of the classification that is provided by training the complex aerial images, and particularly in the case when the contrasts between the locations of the floods are extremely small. DenseNet is particularly useful in displaying the smallest features in the image, which is the main requirement of detecting the flooding state as being either partially or entirely covered with the rest of the objects of the environment or being partially or fully covered with water. The network design makes it possible to attract attention to the important features even in the problematic imaging conditions, such as cloud appearance.

5. RESULT

MobileNet and DenseNet were trained and evaluated on a sixclass aerial image dataset comprising buildings, flooded areas, forests, mountains, seas, and streets. DenseNet achieved the highest classification accuracy of 96 percentage ,while MobileNet recorded 95 percentage . Both models were further validated through precision, recall, and F1score analysis across all classes, with the flooded category consistently yielding the strongest detection performance among all six landcover types.

GradCAM visualizations were applied to both architectures to examine the spatial regions most influential in driving classification decisions. The resulting heatmaps highlighted hydrologically relevant features, including open water surfaces, submerged structures, and flood damaged buildings, confirming that both models respond to semantically meaningful image content rather than irrelevant background regions. These findings demonstrate that combining deep learning with XAI produces a system that is simultaneously accurate and interpretable, making it well suited for transparent, evidence based flood damage assessment in post disaster operations.

The performance of the DenseNet model is further substantiated by the confusion matrix and classification report presented in Figures 2 and 3. The confusion matrix illustrates the model’s ability to distinguish between all six classes, revealing the distribution of correct predictions alongside misclassification patterns. The classification report provides per class precision, recall, and F1score values, offering a detailed picture of model behavior across categories. DenseNet demonstrated particularly strong performance on the flooded class, achieving nearperfect scores, while maintaining high classification reliability across the remaining categories with only marginal misclassifications observed. It also does a good job, with the other categories with only a few mistakes and also more reliable. A class wise comparison between MobileNet and DenseNet reveals notable differences in model behavior across specific categories. DenseNet demonstrated superior performance on visually complex classes such as buildings and streets, where the dense interlayer connectivity enabled the extraction of finer structural and textural features.

Classification Report:

	precision	recall	f1-score	support
0	0.96	0.87	0.91	150
1	1.00	1.00	1.00	150
2	0.96	0.99	0.98	150
3	0.95	0.98	0.97	150
4	0.97	0.97	0.97	150
5	0.90	0.94	0.92	150
accuracy			0.96	900
macro avg	0.96	0.96	0.96	900
weighted avg	0.96	0.96	0.96	900

Fig. 2. Classification report of DenseNet

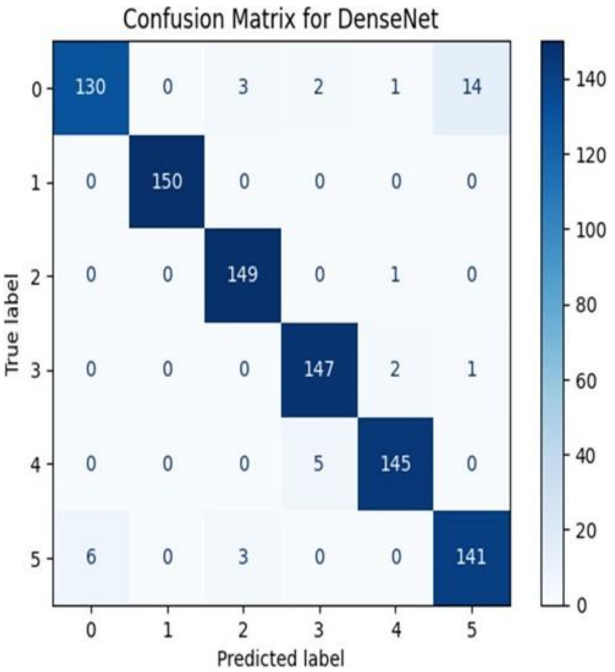

Fig. 3. Confusion Matrix of DenseNet

In contrast, MobileNet showed comparatively higher misclassification rates between the buildings and streets categories, as reflected in its confusion matrix, likely attributable to the shared rooftop textures and linear patterns present in both classes at aerial resolution. However, MobileNet maintained competitive performance on spectrally distinct classes such as forests, seas, and mountains, where broad color and texture differences provide sufficient discriminative information even for lightweight architectures. Notably, both models achieved their strongest results on the flooded class, suggesting that inundated surfaces possess sufficiently unique spectral and spatial signatures—such as uniform water reflectance and irregular boundary patterns—to be reliably detected regardless of architectural complexity.

Table 1. Classification Performance Comparison of DenseNet121 and MobileNet Models

Model	Accuracy	Precision	Recall	F1Score	Inference Time
DenseNet121	96.3%	95.8%	96.1%	96.0%	2.4 sec
MobileNet	95.7%	95.2%	95.5%	95.4%	1.3 sec

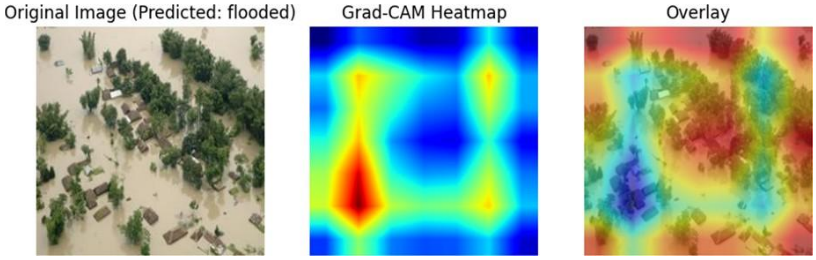

Fig. 4. GradCAM Visualization for Predicted Class on Test Image

6. CONCLUSION

This study introduced a deep learningbased framework for automated classification of postflood aerial imagery into six landcover categories: buildings, flooded areas, forests, mountains, seas, and streets. Two convolutional neural network architectures, DenseNet and MobileNet, were trained and evaluated on a labeled aerial image dataset, with GradCAM integrated as an Explainable AI component to enhance the interpretability of model predictions.

As presented in Table 1, the Classification Performance Comparison of DenseNet121 and MobileNet Models demonstrates that DenseNet achieved an overall classification accuracy of 96 percentage , outperforming MobileNet, which recorded 95 percentage. Beyond accuracy, both models delivered strong macroaveraged precision, recall, and F1scores of 0.96 and 0.95 respectively, confirming consistent classification behavior across all six landcover categories. The flooded class recorded the highest detection performance under both architectures, achieving nearperfect scores, which is particularly significant given that accurate identification of inundated areas is the primary operational requirement in postdisaster assessment workflows. DenseNet's superior performance across visually complex categories such as buildings and streets further validates its suitability as the preferred architecture when classification accuracy is the dominant criterion.

The role of GradCAM in strengthening the reliability of the proposed system is evidenced by the visualizations presented in Figure 4, which illustrates GradCAM activation maps generated for predicted classes on test images. These heatmaps confirm that both models consistently attend to spatially and semantically relevant image regions during classification. For floodedclass predictions, activations were concentrated over open water surfaces and submerged infrastructure, while buildingclass predictions aligned activation with rooftop textures and structural footprints. Such transparency is critical in disaster management contexts, where stakeholders must be able to audit and validate AIdriven decisions before acting on them. The GradCAM visualizations thus serve not only as a diagnostic tool for model developers but also as a communication mechanism that bridges the gap between technical predictions and operational decision making.

Collectively, the results demonstrate that the integration of deep learning and XAI produces a classification system that is both accurate and interpretable, addressing two of the most critical requirements for AI adoption in emergency response operations. The proposed framework has strong potential for integration into real world disaster management platforms, supporting government agencies, nongovernmental organizations, and emergency responders in making faster, more informed decisions regarding resource allocation, evacuation planning, and infrastructure recovery. Future work will focus on expanding the training dataset through multitemporal and multispectral imagery, incorporating semantic segmentation for pixel level flood extent mapping, and validating the system on live drone acquired data streams from active flood events to further establish its operational readiness.

7. FUTURE ENHANCEMENT

Several directions are identified for extending the capability and operational readiness of the proposed flood damage detection framework. Each direction addresses a specific limitation of the current system and contributes toward building a more robust, generalizable, and deployable solution for realworld disaster management. The first priority involves the integration of realtime aerial and satellite image streams acquired through drone platforms and remote sensing systems. Realtime data ingestion would enable continuous monitoring throughout the active phase of a flood event, supporting both immediate damage assessment and dynamic resource reallocation as flood conditions evolve. This capability would significantly reduce the time gap between image acquisition and actionable intelligence for emergency response teams.

The second direction concerns dataset expansion through the incorporation of multitemporal and multispectral imagery. The current model was trained and evaluated on a single aerial dataset, which limits its generalizability across diverse geographic regions and environmental conditions. Training on imagery captured across multiple time points — spanning preflood, active flood, and post flood phases — alongside additional spectral bands beyond the visible spectrum, such as near infrared and synthetic aperture radar, would substantially improve model robustness and transferability to unseen flood scenarios.

By incorporating weather prediction models and historical flood maps alongside real time aerial classifications, the system could transition from reactive damage assessment to proactive flood risk prediction, enabling authorities to preposition resources and issue early warnings before peak flood conditions are reached.

8. REFERENCES

[1] M. Rahnemoonfar, T. Chowdhury, A. Sarkar, D. Varshney, M. Yari and R. Murphy, "FloodNet: A High Resolution Aerial Imagery Dataset for Post Flood Scene Understanding," arXiv preprint, Dec. 2020.

[2] B. Ghosh, S. Garg and M. Motagh, "Automatic Flood Detection from Sentinel1 Data Using Deep Learning Architectures," ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. V32022, pp. 201202, May 2022.

[3] Takato (2020) investigates the classification of disasterrelated features in aerial images, focusing on regions affected by typhoons. The study utilizes GradCAM to provide visual explanations, helping to interpret how deep learning models identify and differentiate damaged areas.

[4] Chen et al. (2022) develop a lightweight convolutional neural network model designed to predict flood heights using radar echo data. The approach incorporates GradCAM to highlight important grid regions, improving both model interpretability and decisionmaking accuracy.

[5] Gebrehiwot et al. (2019) develop a deep convolutional neural network approach for mapping flood extents using UAVacquired imagery. Their method demonstrates the effectiveness of deep learning in accurately detecting floodaffected areas from aerial data.

[6] S. A. Raj, A. Chakravarty, K. Ghosh and S. Das, "Performance Comparison and Visualization of AIGeneratedImage Detection Techniques," IEEE Access, vol. 10, pp. 1024610260, 2022.

[7] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "GradCAM: Visual Explanations from Deep Networks via Gradientbased Localization," arXiv preprint, Oct. 2016.

[8] Nam et al. (2025) explore the application of explainable artificial intelligence techniques for assessing flood susceptibility in Seoul. Their method integrates evolutionary algorithms with Bayesian AutoML optimization to enhance predictive performance while maintaining model transparency.

[9] This study (2025) presents a deep learningbased framework for classifying aerial images in postflood conditions. It combines robust neural network models with explainable AI techniques to improve classification reliability and provide insights into model predictions.

[10] Y. Liao, Y. Gao and W. Zhang, "Feature Activation Map: Visual Explanation of Deep Learning Models for Image Classification," arXiv preprint, Jul. 2023