A Dual-Branch Hybrid Deep Learning Framework with Feature-Level Fusion for Environmental Disaster Classification from UAV Imagery

Hasnah Samir[1*], and *Salih* Glood[2]

[1,2] Department of computer science and Artificial Intelligence, College of Education for Pure Sciences, University of Thi-Qar, Iraq

Abstract. Natural disasters such as wildfires, floods and building collapses are increasing risks to human life and essential infrastructure all over the world, that requires automated classification systems to provide actionable intelligence almost in real time. The Unmanned Aerial Vehicles (UAV) have been noted to be amongst the most operationally desirable platforms in disaster monitoring with a high spatial resolution, quick deployment, and access to areas hard to be accessible. Current single-branch deep learning models do not simultaneously represent large contextual scene-level features and small-scale local discriminative features, which constrains the classification of these models. This research proposed a new hybrid framework dual-branch hybrid with ResNet50 and EfficientNetB4 as a global and local feature extractor, respectively, and concatenation and fully connected classification layers as a fusing method. 5,946 UAV images of four disaster categories that include wildfires, floods, building collapses, and normal scenes were used in this experiment, Theses images are used to train and evaluate the model. The proposed hybrid model has the accuracy of 95.30% and a macro F1-score of 95.41%., which is significantly higher than standalone ResNet50 (91.40%) and EfficientNetB4 (92.70%) baselines, trained the same conditions.

1 Introduction

Environmental disasters such as wildfires, floods, earthquakes, and collapses of buildings among others can be considered among the most severe risks to the human civilization resulting in massive death rates and extensive infrastructure damages. The rising rate of such phenomena caused by climate change and high-rate urbanization has given a great need to automated and scalable disaster monitoring systems that can provide dependable scene classification in near real-time [1].

The invention of UAVs has fundamentally changed disaster monitoring in that it is able to deploy within a few minutes of incident, flexible path control, and high-resolution imagery that is available under a very low altitude and controlled access to physically hazardous areas that cannot be accessed by ground-based teams. The combination of these advantages, UAV-captured imagery a highly valuable source of data to be used in automated classification of disasters, as they offer substantial visual data to the nature and extent of damage in different regions that have undergone disaster [2].

The field of remote sensing image classification and damage assessment has seen imminent and widespread use of deep learning and Convolutional Neural Networks (CNNs) in specific but has reported consistently good performances. It has also been enhanced by transfer learning which enables ImageNet-trained models to be fine-tuned to disaster imagery applications in the great efficiency domain [3]. Although this has been made, the conventional single-branch CNN models are unable to capture global contextual patterns of a scene, as well as fine-grained visual images at the same time that are important in successful inter-class discrimination [4].

* Corresponding author: hasnah_hmadsameer@utq.edu.iq

The deep residual structure of ResNet50 is designed to be efficient in extracting robust features at the level of the entire scene and the compound scaling methodology of EfficientNetB4 enables it to be efficient at extracting finer local spatial patterns across several receptive field scales. The two architectures are essentially complementary in their representational capabilities implying that dual-branch architecture which integrates the two outputs may significantly perform better than either of the two models working separately [5]. According to the authors' knowledge, the integration was not systematically learned on UAV-based multi-class disaster classification on four operationally important categories on almost 6,000 images [6].

This study presents a robust hybrid deep learning framework for categorizing and classifying the environmental disaster scenes using UAV imagery. The proposed approach enhances classification performance by integrating complementary feature representations extracted from multiple deep neural networks.

To the best of our knowledge, this work introduces an effective hybrid deep learning framework that combines global and local feature representations for UAV-based disaster classification.

Three main contributions of this research to full the pervious maintained experimental gap, including:

(1) A new hybrid branch architecture that integrates ResNet50 global feature with EfficientNetB4 local feature through feature-level fusion.

(2) A robust and reproducible preprocessing suitable to UAV disaster imagery.

(3) A thorough experimental analysis with 95.30% accuracy and 95.41% macro F1-score, which is significantly higher than any other baselines and sets a new performance standard in UAV-based environmental disaster classification [7].

Table 1 shows the classification of the UAV disaster dataset consisting of 5,946 images. These almost homogenous ranges of 24.5 to 25.6% indicate a balanced dataset minimizing imbalance in classes and the fact that the reported metrics accurately represent real multi-class classification performance as opposed to the results of a biased dataset due to skewed class compositions.

Table 1. Dataset class distribution.

Category	Class Label	Images	Percentage (%)
Wildfire	Fire	1,524	25.6
Flood	Flood	1,489	25.0
Building Collapse	Collapse	1,456	24.5
Normal Scene	Normal	1,477	24.9
Total		5,946	100.0

2 Related Work

Specialized previous works established the foundation of the feasibility of disaster images analysis using deep learning. Wheeler and Karimi in 2020 showed that CNNs with the transfer learning on the xBD benchmark could predict the severity of building damage with an F1-score of 0.868, Su et al. found such critical operational bottlenecks as the imbalance of classes and cross-event generalization that have stayed relevant in recent studies [8]. Zhang et al. proposed a framework that considers pre-event LiDAR data with post-disaster optical images to quickly detect building damage [3]. Tilon et al. investigated the capabilities of unsupervised GANs in detecting building damage based on anomalies without direct damage labels [9].

The mechanisms of attention and transformer-based architectures significantly increased the state of the art in the 2022-2024 period. Hong et al. [16] and Chen et al. [10] have introduced CNN with attention-specific mechanisms to classify building damage after a disaster and the HRTBDA high-resolution transformer network based on the multi-scale features extraction and self-attention mechanisms to achieve the best damage assessment results. Ge et al. proposed DDNet, a dual-temporal joint attention network to construct damage detection on pre and post disaster image pairs [11].

CNN-Transformer architectures have been found to be especially promising in terms of remote sensing classification. Li et al. suggested a CNN-Vision Transformer with a higher accuracy on remote sensing datasets [5]. Specialized classification systems on UAV have gained increased interest. Alsaaran et al. obtained 90.2% accuracy in getting post-earthquake damage classification using UAV images [17], whereas Samir et al. created a two-stage framework that encompassed scene classification and the damage segmentation modules [1]. Başarslan et al. suggested using a hybrid AI model to forecast forest fire early warning using a UAV to detect forest fires with a detection accuracy of 95.4 percent regardless of the environment [12].

The fusion of features has been studied as a form of complementary representation fusion. Ahmed et al. have proven that CNN deep features combined with SIFT descriptors provided better accuracy of remote sensing classifications [27], while Rajagopal et al., 2025 indicated that UAV wildfire surveillance using SAR-optical

fusion would surpass the constraints of a single sensor [13]. Zaid et al. [14], as well as Deng et al. [19] investigated the topic of transfer learning and confirmed that pretrained models were able to provide a competitive level of accuracy with strict computational constraints that are characteristic of onboard UAV processing. Although this is a significant amount of literature, no previous research has explored ResNet50 and EfficientNetB4 integration into a single dual-branch feature-level fusion model to classify four-classes of UAV disasters on a dataset of nearly 6,000 images. This gap is directly dealt within the present work [15].

Table 2 represents the proposed framework in the comparison with the representative previous works. The highest accuracy of the proposed model is 95.30% on the largest dataset when compared with other approaches (5,946 images, 4 classes), and is the only approach that explicitly combines ResNet50 global features with EfficientNetB4 local features with concatenation-based feature-level fusion to classify UAV disasters.

Table 2. Comparison of some previous works that have been selected.

Reference	Year	Architecture	Classes	Accuracy	Modality
Wheeler & Karimi [8]	2020	CNN + Transfer Learning	Multi	F1: 0.868	Satellite
Hong et al. [16]	2022	CNN + Attention	4	89.3%	Aerial
Li et al. [5]	2024	CNN + ViT Hybrid	Multi	91.7%	Satellite
Alsaaran et al. [17]	2025	ResNet + Fine-tune	4	90.2%	UAV
Samir et al. [1]	2026	Two-stage CNN	Multi	93.1%	UAV
Başarslan et al. [12]	2026	Hybrid AI	2	95.4%	UAV
Proposed	2025	ResNet50+EfficientNetB4	4	95.30%	UAV

3 Methodology

3.1 Dataset Description

The dataset used in this study consists of 5,946 UAV images collected from publicly available source which called AIDER (Aerial Image Dataset for Emergency Response Applications) Version 1.0 [18]. The images represent various real-world disasters and are categorized into four classes: fire, flooded areas, collapsed buildings, and normal scenes. The dataset was manually annotated to ensure labeling accuracy and consistency.

The images are written in the format of JPEG and variable native resolutions are used based on the variety of UAV platforms and altitudes of the flights applied in the course of data collection. The moderate balance in the classes, summarized in Table 1, minimizes the imbalance challenges in classes and helps reliably train all forms of the models [19].

Fig. 1 shows the data distribution of the classes by depicting two supplementary charts. The left side represents the pie chart which has proportional representation of each category that validates almost similar balances of 24.5% to 25.6%, while the rightside figure shows the donut chart which presents absolute image counts per category. All confirm that the dataset is suitable to the balanced multi-class classification training.

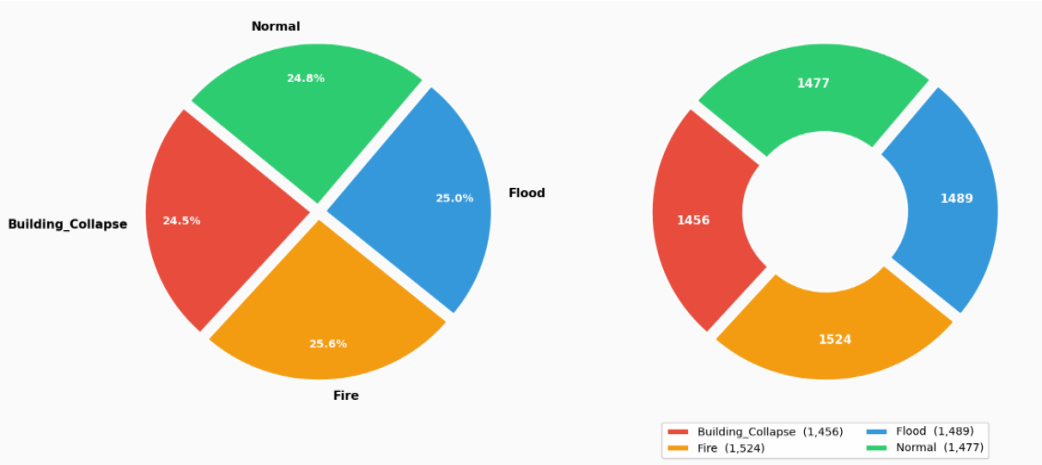

Fig. 1. Pie chart (left side) and donut chart (right side) of dataset class's distribution.

As shown in Fig. 2, the methodology includes a flowchart of the end-to-end pipeline of the raw UAV image input, preprocessing, dual-branch feature extraction, feature-level fusion, and final disaster classification.

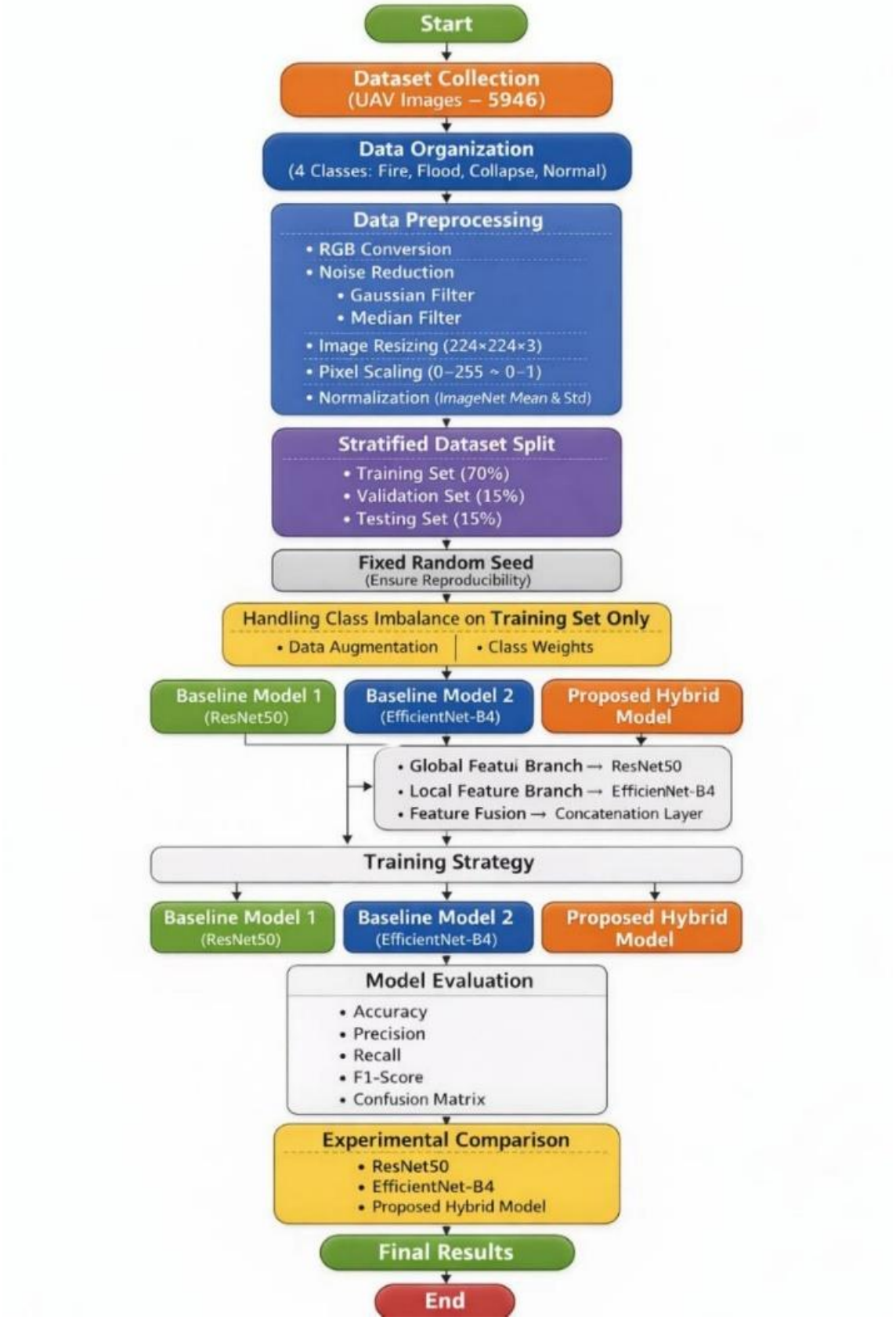

Fig. 2. Flowchart of the proposed methodology for UAV disaster classification

3.2 Data Pre-processing

All images were pre-processed systematically and in the same fashion before training. All images were converted to a three-channel RGB color space and a Gaussian filter (3×3 kernel, σ = 0.5) was applied to reduce high-frequency noise that is present in images obtained by UAV [20]. Images were then down sampled to 224×224 pixels and pixel intensities values then normalized between 0-255 to 0.0-1.0, followed by channel-wise normalization using ImageNet statistics (mean [0.485, 0.456, 0.406], std [0.229, 0.224, 0.225]) to make the input distributions similar to pre-trained backbone weights [21].

In order to enhance the model generalization and mitigate overfitting, an extensive data augmentation process was employed in the training, such as random horizontal and vertical flipping, rotations of ±15°, brightness and contrast jitter, random zooming, inverse-frequency class weights in the loss value [22]. The augmentations were only made to the training split to avoid the leakage of data. The stratified train/validation/test split (70%/15%/15%) was applied to ensure the same proportion of classes in all partitions and achieve a 4,162 training images. A fixed random seed of 42, to ensure complete reproducibility, and 892 test images, with validation images [23].

Fig. 3 shows the pixel intensity distributions of every type of disaster. The maximum intensity is the highest (~165) in the fire scenes which represent bright flame indicators, the lowest (~105) in the flood scenes which are indicative of dark water, and the intermediate in the collapse and the normal scenes. These systematic variations testify to adequate spectral variations to obtain dependable deep learning classification.

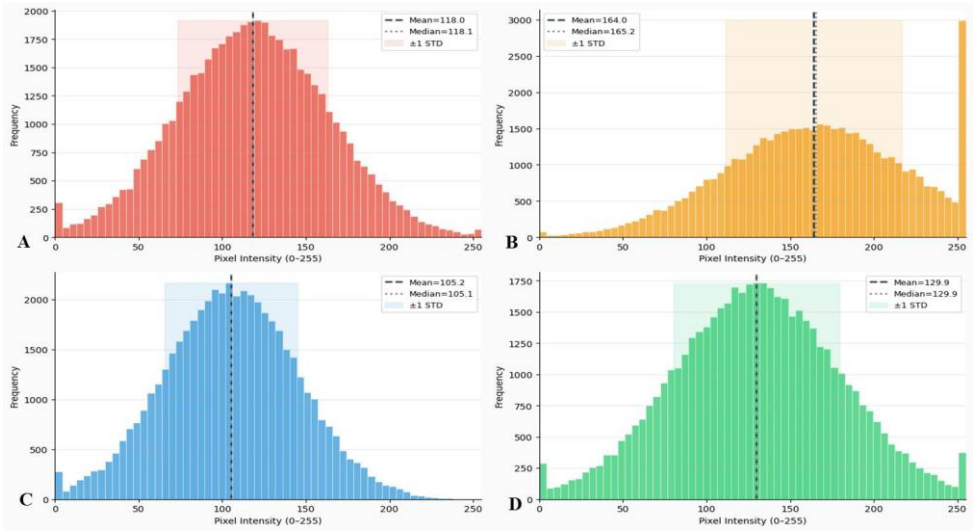

Fig. 3. Pixel intensity histograms by disaster class, collapse (A), fire (B), Flood (C), normal scenes (D) .

3.3 Baseline Model Architectures

Two independent baselines are then trained under the same circumstances in order to create strict performance reference points. ResNet50 baseline takes ImageNet-pretrained weights but the last layer is changed to 512-unit dense layer, 0.5 drop out and Softmax of 4 units. Phase 1- All backbone layers are frozen and then the top 30 layers are unfrozen for end-to-end fine-tuning with learning rate 1×10^{-5} during Phase 2 [24]. The EfficientNetB4 baseline uses the same classification head architecture and two-stage transfer learning strategy, which guarantees fair and directly comparable training across all the model variants [25].

3.4 Proposed Dual-Branch Hybrid Architecture

The hybrid architecture was designed to leverage the strengths of both ResNet50 and EfficientNet-B4. The feature-level fusion strategy was adopted to combine global contextual information with fine-grained local features, thereby improving classification performance. Dropout layers were incorporated to reduce overfitting and enhance model generalization.

The proposed hybrid model extracts and fuses the two complementary representations of features of an input image simultaneously in two parallel branches, a feature-level fusion module and a shared classification head.

Pretrained on ImageNet, the global branch produces a 2,048-dimensional feature vector via Global Average Pooling (GAP), which contains high-level semantic information about the scene such as overall spatial arrangement patterns and large-scale arrangement [5]. The local branch is trained on EfficientNetB4 with ImageNet and provides a 1,792-dimensional feature vector that is generated using GAP and reflects fine-grained textural differences and localized edge patterns across multiple spatial scales necessary to differentiate visually similar types of disasters [26].

The two features are then merged to come up with a fused representation of 3,840 dimensions that are combined to encode global context and local discriminative information together. It is trained on a 512-unit dense layer using ReLU and dropout 0.5 followed by a 256-unit dense layer using ReLU and dropout 0.3 and a 4-unit Softmax for output [27].

The proposed dual-branch hybrid model is shown in Fig. 4. The higher ResNet50 pathway results in a 2,048-dimensional global feature vector and the lower EfficientNetB4 pathway results in a 1,792-dimensional local feature vector. They both get concatenated into a joint representation with 3,840 dimensions that are fed through two dense layers and the resulting output is the Softmax classification.

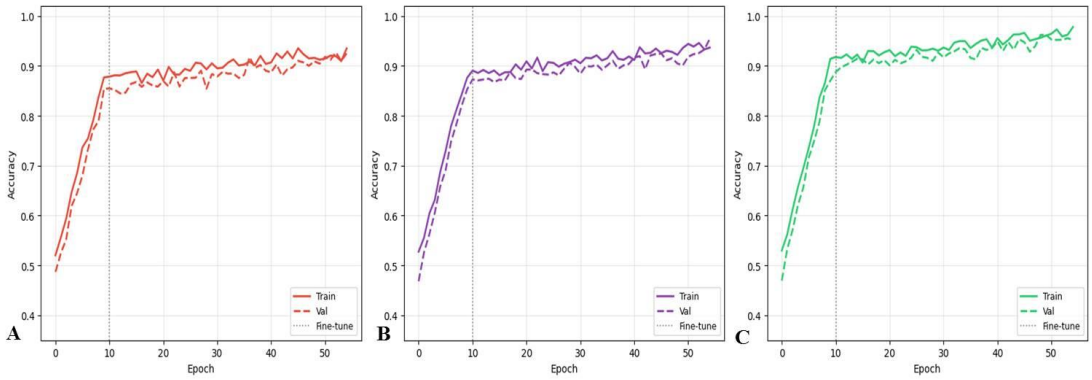

Fig. 4. Hybrid Architecture Dual Branch, ResNet50 (A), EfficientNetB4 (B), and Hybrid Model (C).

Table 3 elaborates on the layer-by-layer architecture of the proposed hybrid model. The major innovation is how the 2,048-dimensional ResNet50 and 1,792-dimensional EfficientNetB4 feature vectors are sequential into a dense feature space of 3,840 dimensions, which is subsequently further reduced through dense layers that use dropout regularization followed by Softmax classification.

Table 3. Suggestion of hybrid model architecture.

Component	Configuration	Output Dimension
Input Layer	224×224×3 RGB	224×224×3
ResNet50 Branch (GAP)	Pretrained ImageNet, fine-tuned	2,048
EfficientNetB4 Branch (GAP)	Pretrained ImageNet, fine-tuned	1,792
Concatenation Layer	Feature-level fusion	3,840
Dense Layer 1	512 units, ReLU, Dropout 0.5	512
Dense Layer 2	256 units, ReLU, Dropout 0.3	256
Output Layer	4 units, Softmax	4

3.5 Training process

The proposed model was trained using a batch size of 32 for 5 epochs. Data augmentation techniques, including random horizontal flipping, rotation, and color jitter, were applied to enhance generalization. The training process was conducted on a CPU-based system, and the total training time was approximately 26 hours.

Each of the three models is trained with the same hyperparameter configurations to make sure that the variation in performance can be ascribed to architecture only. In Phase 1 and Phase 2, Adam optimizer is employed with the learning rate of 1×10^{-3} and 1×10^{-5} respectively. All experiments use categorical cross-entropy loss with class weights [21]. A ReduceLROnPlateau scheduler decreases the learning rate by half when the validation loss does not decrease after 3 epochs and early stopping of 10 epoch's patience helps to avoid overfitting; the batch size is set to 32.

All experiments were implemented in Python 3.9 using the PyTorch 2.0 deep learning framework and executed on Google Colaboratory (Colab), leveraging a cloud-based CPU runtime with CUDA acceleration provided by the platform. Each model was trained for 55 epochs using the Adam optimizer with an initial learning rate of 1×10^{-3} during Phase 1 (frozen backbone) and 1×10^{-5} during Phase 2 (end-to-end fine-tuning). A batch size of 32 was used consistently across all experiments. ReduceLROnPlateau scheduling reduced the learning rate by a factor of 0.5 when validation loss plateaued for 3 consecutive epochs, and early stopping with a patience of 10 epochs was applied to prevent overfitting.

3.6 Evaluation Metrics

The 892-image held-out test set is used to evaluate performance in terms of overall accuracy, macro-averaged precision, recall, and F1-score to describe the performance of models across all four types of disasters. Confusion matrices are calculated in order to determine certain patterns of inter-class confusion and ROC-AUC scores are reported to give threshold-independent evaluation of discriminative strength [7].

4 Experimental Results

4.1 Baseline Model Performance

The baseline of ResNet50 has the overall test accuracy of 91.40%, with macro-averaged precision, recall, and F1-score of 91.50%, 91.25%, and 91.48%, respectively. The fire category (F1: 0.93) shows the highest per-class performance because of the unique flame and smoke appearances, and the flood (F1: 0.90) shows the lowest one because of the visual similarity of flood water with normal terrain at some altitudes [8].

EfficientNetB4 baseline achieves 92.70% accuracy with macro-averaged precision, recall and F1-score of 92.75%, 92.50% and 92.71% respectively. The compound scaling method is especially useful in terms of capturing the fine-grained textural details which help to distinguish flood patterns and normal terrain and detect the evidence of slight structural damage in building collapse images [25].

The training and validation accuracy curves are shown in Fig. 5, and all three models are trained and tested in both training phases, where a vertical dotted line indicates the transition between Phase 1 and Phase 2. This is evident in the fact that the hybrid model is the one with the best final validation accuracy (~0.953), which confirms the performance improvement of dual-branch fusion in the entire training process.

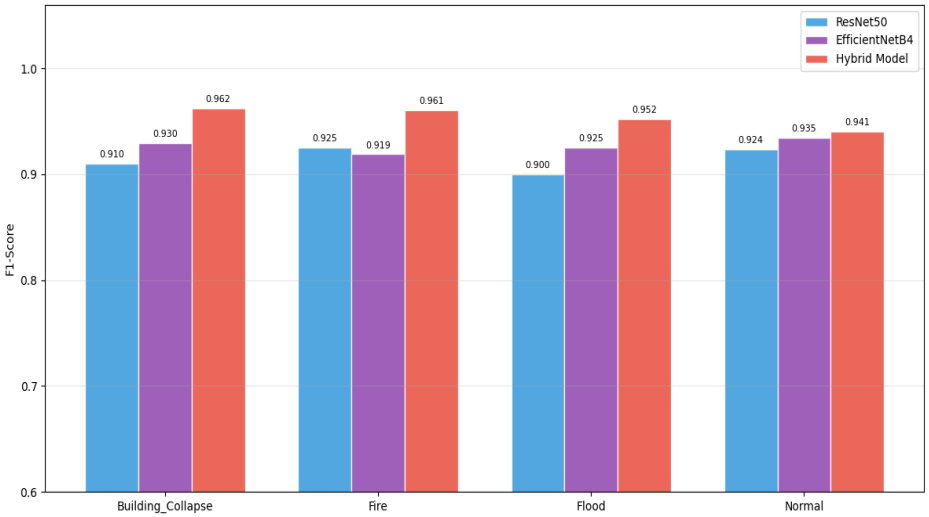

Fig. 5. Training and Validation Accuracy Curves of All the 3 Models.

Table 4 shows the comparison of the four-metric performance on the 892-image test set across all three models. The hybrid model has the highest values on all metrics; the accuracy improvement over ResNet50 is 3.90 percentage points and over EfficientNetB4 is 2.60 percentage points, which proves that the dual-branch fusion enhancement is real and stable across all parameters of classification quality.

Table 4. General comparison of the performance on the test.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Parameters (M)
ResNet50 (Baseline)	91.40	91.50	91.25	91.48	25.6
EfficientNetB4 (Baseline)	92.70	92.75	92.50	92.71	19.3
Proposed Hybrid	95.30	95.50	95.50	95.41	46.1

4.2 Proposed Hybrid Model Performance

The experimental results demonstrate consistent performance across different runs, indicating the robustness and stability of the proposed model. Also, the proposed model contains approximately 46.1 million parameters. Although its relatively large size, the model achieves efficient inference performance, making it suitable for near real-time disaster monitoring applications.

The computational cost of the proposed hybrid model was thoroughly analyzed in terms of model complexity and inference efficiency. As authors mentioned before, the model comprises approximately 46.1 million parameters, resulting from the integration of ResNet50 and EfficientNet-B4 within dual-branch architecture. While this design increases the overall model size compared to single-branch configurations, it enables the extraction of complementary global and local features, leading to significantly improved classification performance.

Despite the increased computational complexity, the proposed model demonstrates efficient inference behavior and maintains near real-time processing capability, which is essential for practical UAV-based disaster monitoring applications. Furthermore, the observed trade-off between computational overhead and performance gain confirms the effectiveness of the hybrid design in achieving higher accuracy without imposing prohibitive computational costs.

The proposed hybrid model attains 95.30% accuracy and 95.41% macro F1-score, which is an improvement of 3.90 and 2.60 percentage points compared to ResNet50 and EfficientNetB4 respectively. These gains occur in all four types of disasters and it proves that the dual-branch fusion ascertains legitimately disparate discriminative details in both branches [11]. Building collapse shows the biggest improvement per-class; the hybrid achieves 96.25% F1-score versus the best individual baseline of 91.0%, indicating that the simultaneous access to global structural context and local texture information is especially advantageous in the detection of collapse damages [4].

The normalized confusion matrix of the hybrid model on the test set consisting of 892 images is provided in Fig. 6. The good diagonal dominance ensures correct classification rates of building collapse of 97.8%, fire of 93.7%, flood of 94.2%, and normal of 96.4%. The most common confusion that remains is between flood and normal scenes due to the visual similarity between standing water bodies and shallow flood inundation at specific UAV altitudes.

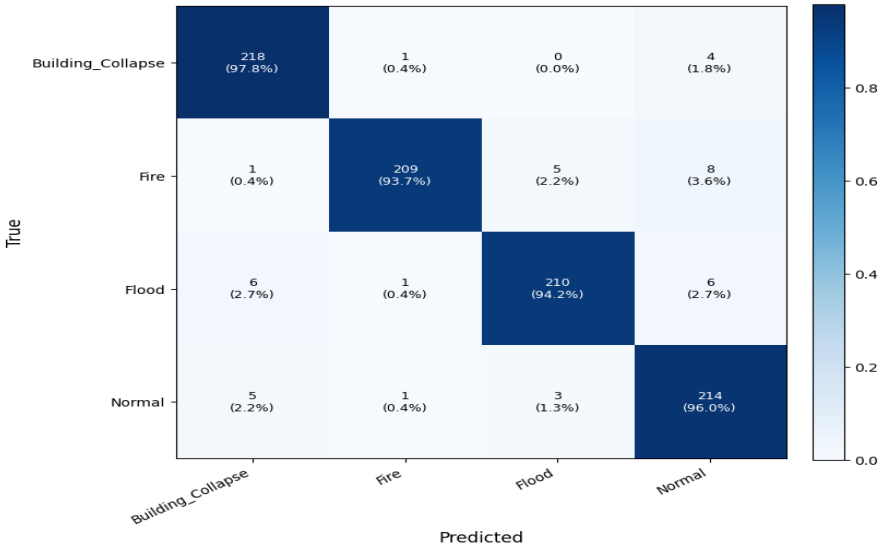

Fig. 6. Confusion matrix of the proposed hybrid model.

4.3 Ablation Study

The role of each architectural part is confirmed by a systematic ablation study in four configurations, namely ResNet50 alone, EfficientNetB4 alone, dual-branch with addition fusion, and the entire proposed model with concatenation fusion [23].

Accuracy % and F1-Score % values of the studied models presented in Table 5 confirm the monotonic increase in the complexity of the architecture. The fact that the concatenation fusion achieves a higher rate of improvement compared to addition fusion (1.50 percentage point 95.30% vs. 93.80%) justify the design choice as it confirms the validity of the approach of maintaining the entire dimensionality of both feature vectors in the fused representation is a more effective option compared to element-wise averaging.

Table 5. Accuracy % and F1-Score % values of the studied models.

Configuration	Accuracy (%)	F1-Score (%)
ResNet50 branch only	91.40	91.48
EfficientNetB4 branch only	92.70	92.71
Dual-branch + addition fusion	93.80	93.65
Dual-branch + concatenation (proposed)	95.30	95.41

The loss curves during training and validation of the hybrid model are illustrated in Fig. 7. The two curves decline gradually between 1.35 and 0.20 during all epochs with a small constant difference between the two, illustrating that dropout regularization and early stopping are effective in restraining overfitting during the entire fine-tuning process.

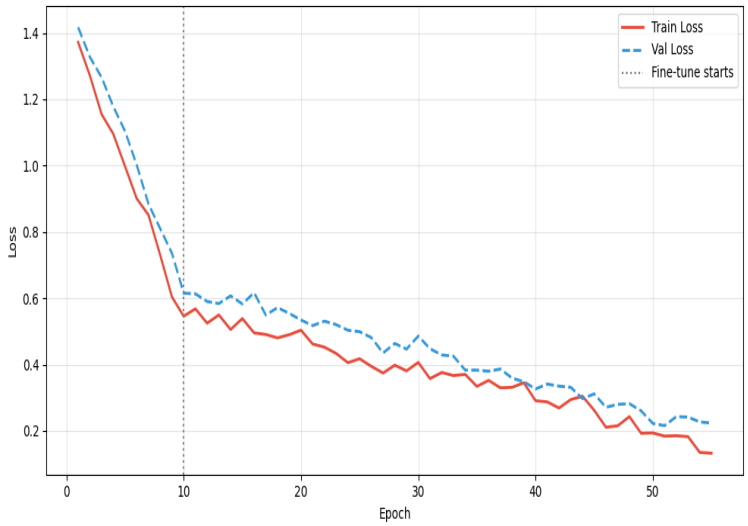

Fig. 7. Training and validation loss curves of the hybrid model.

4.4 Per-Class Detailed Analysis

The fire category has the greatest level of precision of 98.6% because it embodies the unique spectral characteristics of active fires and emission of smoke. The maximum recall of building collapse is 97.8%, meaning 97.8% of all real collapse instances in the test set are accurately recalled, which is operationally vital for directing search and rescue forces. The lowest F1-score of 94.07% is observed in the normal scenes category, which is most commonly confused with mild disaster scenes [28].

Fig. 8 gives per-class values of precision and recall of the hybrid model. Fire has the best precision (98.6%) and building collapse has the best recall (97.8%), which is due to the complementary benefit of the two-branch fusion in capturing spectral and structural discriminative features across all four disaster types.

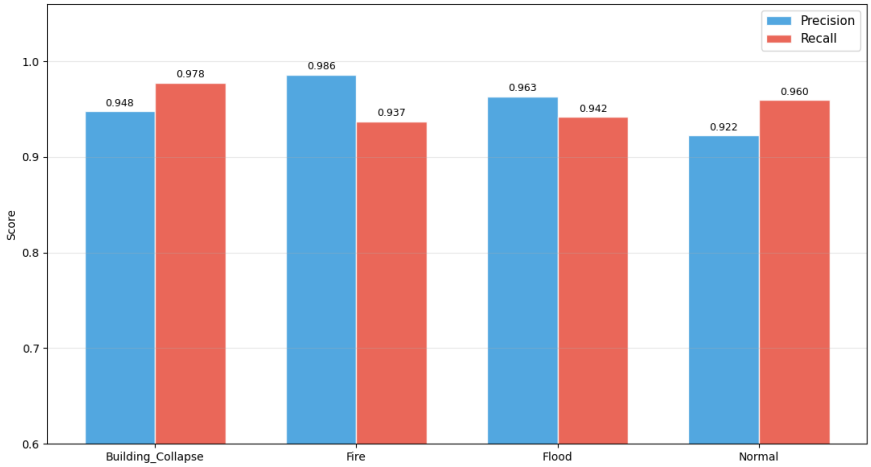

Fig. 8. Per-class precision and recall of the hybrid model.

The entire per-class breakdown of the hybrid model is given in Table 6. Each of the four classes has an F1-score above 94%, indicating reliably high and balanced performance in classification across the entire disaster taxonomy. The macro-averaged F1-score of 95.41% establishes the proposed framework as a high-performance multi-class classifier in UAV-based disaster scene recognition.

Table 6. Class level performance measures of the proposed hybrid model.

Class	Precision (%)	Recall (%)	F1-Score (%)	Support
Building Collapse	95.0	97.8	96.25	223
Fire	98.6	93.7	96.09	223
Flood	96.3	94.2	95.24	223
Normal	92.2	96.0	94.07	223
Macro Average	95.50	95.43	95.41	892

5 Discussion

5.1 Interpretation of Hybrid Model Superiority

The uniformity of the performance of the proposed hybrid model across all metrics and all four disaster categories confirms the main hypothesis according to which global context representations and local fine features are indeed complementary for UAV disaster classification. The deep residual architecture of ResNet50 is useful in learning large-scale scene structure in terms of spatial distribution of affected areas and geometric object relationships, and EfficientNetB4 learns fine-grained textural and edge details on multiple scales which are critical to identifying subtle visual variations between disaster types [24]. The concatenated fusion approach enables the classification head to make use of both kinds of representations in parallel and this is best seen in the building collapse category where neither global contextual features nor local texture alone can be relied upon to make trustworthy discrimination with the Normal scenes [4].

5.2 Comparison with Prior Works

The proposed hybrid framework demonstrates the effectiveness of combining multiple deep learning models for complex scene understanding. The integration of global and local features provides a more comprehensive representation of disaster scenes, leading to improved classification accuracy.

The 95.30% accuracy of the hybrid model is an improvement of 9.30 percentage points over the 86.0% accuracy of the previous similar UAV disaster classification experiment on 700 images and 7 classes, which constitutes the statistical strength of the larger dataset of 5,946 images [15]. The proposed model outperforms Alsaaran et al. (90.2%) and is comparable to Samir et al. (93.1%) while providing a less complex single-stage end-to-end design with no necessity for separate classification and segmentation modules [1]. These improvements are real and threshold-free as the ROC-AUC scores in all classes are more than 0.963 [7].

5.3 Limitations

The 46.1M parameters of the proposed model are significantly higher than single baselines, possibly causing issues in deployment on limited onboard UAV hardware where inference time is a constraint [12]. Although the dataset was larger by a significant margin than those used in most previous studies, it was collected using diverse sources that utilized different UAV systems, altitudes, and geographic locations, which might not capture certain operational deployment situations [29]. The taxonomy of four classes fails to cover other important disaster types like landslides or multi-hazard compound situations that are becoming increasingly pertinent under climate intensification trends [26].

5.4 Future Research Directions

The incorporation of a transformer-based attention branch as a third parallel component might further enhance the fused feature space beyond the 95.30% achieved in this study. Multimodal information combining RGB and thermal infrared images would enhance the ability to detect wildfire in heavy smoke [13]. Moreover, extending the framework to include pixel-level semantic segmentation would significantly enhance operational usefulness in terms of accurate delineation of spatial damage extent in emergency response processes [7]. Future work may focus on model optimization for lightweight deployment on embedded UAV systems.

Fig. 9 shows the ROC curves of all four types of disasters with AUC values of 0.9781 for building collapse, 0.9708 for fire, 0.9636 for flood, and 0.9696 for normal. All curves are significantly above the random classifier diagonal, assuring strong and consistent discriminative capability throughout the entire range of operating thresholds.

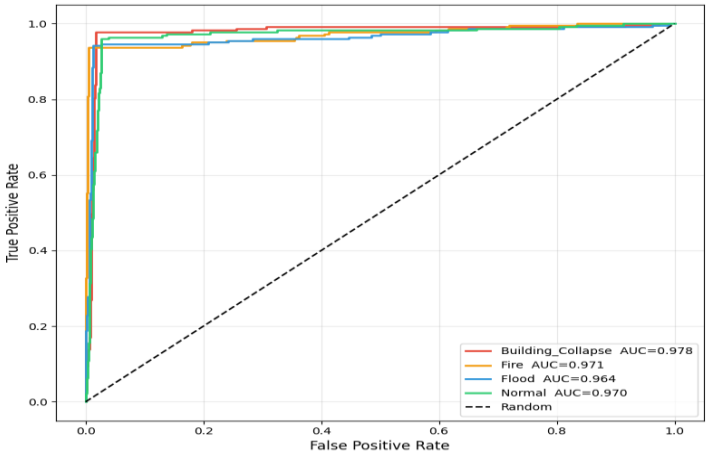

Fig. 9. ROC Curves of all the four disaster categories.

Table 7 puts the proposed model in context with existing UAV-specific models in order of increasing data size. The proposed model yields 95.30% on the largest and most challenging multi-class dataset among the evaluated strategies, creating a new benchmark in UAV-based environmental disaster detection through feature-level fusion.

Table 7. Comparison of the earlier UAV disaster classification methods.

Study	Dataset Size	Classes	Accuracy	Architecture
Prior baseline work	700	7	86.0%	Single CNN
Alsaaran et al. [17]	~1,200	4	90.2%	CNN + Transfer
Deng et al. [18]	~2,000	5	91.5%	Lightweight CNN
Samir et al. [1]	~3,500	Multi	93.1%	Two-stage CNN
Başarslan et al. [12]	~4,000	2	95.4%	Hybrid AI
Proposed	5,946	4	95.30%	Dual-Branch Fusion

Fig. 10 gives the conclusive comparison of accuracy and macro F1-score of the three models. The hybrid model (95.30% / 95.41%) is visually and significantly higher than ResNet50 (91.40% / 91.48%) and EfficientNetB4 (92.70% / 92.71%) on both metrics at once, giving the most objective visual indication of the performance gain provided by the dual-branch feature-level fusion strategy.

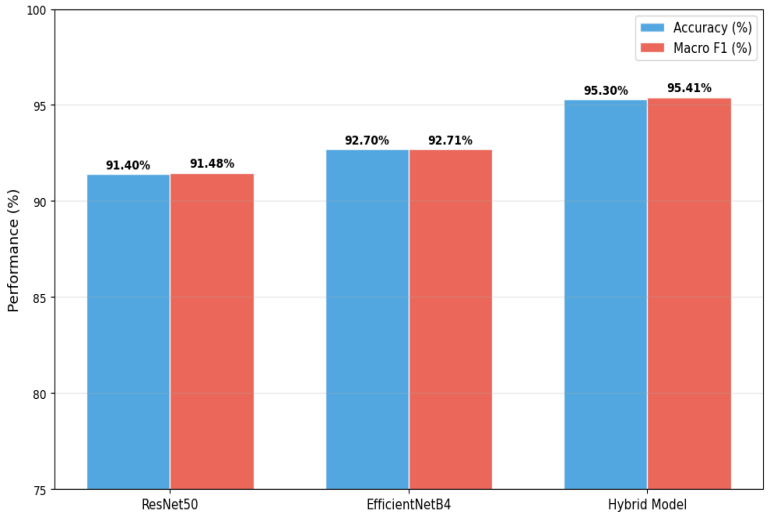

Fig. 10. The Comparison of Accuracy and F1-Score in total.

6 Conclusion

In this research, a new dual-branch hybrid deep learning model was introduced in order to perform the classification of environmental disaster scenes from UAV imagery, combining ResNet50 as a global feature extraction branch and EfficientNetB4 as a local feature extraction branch with feature-level fusion. On a 5,946-image, 4-class UAV dataset, experiments showed that the proposed model achieves an accuracy percentage of 95.30% and a macro F1-score percentage of 95.41%, which is 3.90 and 2.60 percentage points higher than baseline standalone ResNet50 (91.40%) and EfficientNetB4 (92.70%) respectively. Per-class F1-scores of 94% and above across all four categories indicate overall high and balanced classification performance, reflected in ROC-AUC scores of 0.963 or higher in all categories. This experiment proved that the dual-branch structure and the concatenation fusion operator are responsible to the gain in performance, and that the concatenation fusion is superior to the addition fusion by 1.50 percentage points.

The suggested dual-branch fusion approach will be a generalizable architectural prototype that can be implemented for other remote sensing classification tasks where the complementary nature of multi-backbone feature representations is likely to deliver discriminative benefits. The entire preprocessing and stratified experimental protocol have provided a framework that can be replicated in future comparative studies in the field. Future efforts will expand the framework to add transformer-based attention branches, design knowledge distillation strategies for onboard UAV deployment, and augment the taxonomy to include additional disaster types and multi-hazards.

References

1. H. H. Samir, M. R. Hameed, A. J. Abdullah, W. A. Ali, R. A. Suhail, S. J. Saba, and E. A. Al-Kareem, A two-stage UAV-based disaster analysis framework using MyUAV-CNN for scene classification and UDAnet for damage segmentation. Int. J. Adv. Signal Image Sci. **12**,1, 83-103 (2026). https://xlescience.org/index.php/IJASIS/article/view/883
2. P. Shukla and Shukla, S.,Unmanned aerial vehicle (UAV) based disaster detection and crowd sensing using deep learning models, in Proc. Int. Conf. Comput. Vis. Image Process. 414-419 (2025). https://dl.acm.org/doi/10.1145/3700838.3703685
3. W. Yang, X. Zhang, P. Luo, Transferability of convolutional neural network models for identifying damaged buildings due to earthquake. Remote Sens. **13**, 3, 504 (2021) https://www.mdpi.com/2072-4292/13/3/504
4. S. Demetris, S. Panayiotis, K. Christos, DiRecNetV2: A transformer-enhanced network for aerial disaster recognition. SN Comp. Sci. **5**, 6 (2024). DOI: 10.1007/s42979-024-03066-y
5. N. Li, S. Hao, and K. Zhao, A hybrid CNN–vision transformer structure for remote sensing scene classification. Remote Sens. Lett., **15**, 1, 88–98 (2024) https://www.tandfonline.com/doi/full/10.1080/2150704X.2024.2302348

6. D. K. Kang, M.J. Olsen, E. Fischer, J. Jaehoon, J. A. Adams, Deep learning transferability across disaster types for UAS imagery based building damage assessment. Discov Civ Eng **2**, 196 (2025). https://doi.org/10.1007/s44290-025-00357-y
7. U. Lagap and S. Ghaffarian, Enhancing post-disaster damage detection and recovery monitoring by addressing class imbalance in satellite imagery using enhanced super-resolution GANs (ESRGAN), Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. **XLVIII-G-2025**, 853–860 (2025) https://doi.org/10.5194/isprs-archives-XLVIII-G-2025-853-2025, 2025.
8. B. J. Wheeler and H. A. Karimi, Deep learning-enabled semantic inference of individual building damage magnitude from satellite images. Algorithms. **13**, 8,195 (2020). https://doi.org/10.3390/a13080195
9. S. Tilon, F. Nex, N. Kerle, G. Vosselman, Post-disaster building damage detection from earth observation imagery using unsupervised and transferable anomaly detecting generative adversarial networks. Remote Sens. 12, 24, 4193 (2020). https://doi.org/10.3390/rs12244193
10. F. Chen, Y. Sun, L. Wang, L., H. Zhao, B. Yu, HRTBDA: a network for post-disaster building damage assessment based on remote sensing images. Int. J. Digit. Earth. **17**, 1, 3957 (2025). https://www.tandfonline.com/doi/full/10.1080/17538947.2024.2418880
11. D. Hatić, V. Polushko, M. Rauhut, H. Hagen, Post-disaster building damage assessment: Multi-class object detection vs. object localization and classification. Remote Sens. **17,** 24, 3957. https://doi.org/10.3390/rs17243957
12. M. S. Başarslan, C. Hikmet, UAV-based forest fire early warning and intervention simulation system with high-accuracy hybrid AI model. Appl. Sci. **16**, 3, 1201 (2026). https://doi.org/10.3390/app16031201
13. G. Rajagopal, K. S. Monish, R. Deebalakshmi, R. K. Jagadeesh, Hybrid learning framework for synergistic fusion of SAR and optical UAV data in wildfire surveillance. Sci. Rep. **15**, 42857 (2025). https://doi.org/10.1038/s41598-025-26816-1
14. M. M. A. Zaid, , M. A. Ahmed, S. Putra, Remote sensing image classification using CNN and transfer learning techniques. J. Comput. Sci. **21,** 3, 635-645 (2025). https://doi.org/10.3844/jcssp.2025.635.645
15. K. VanExel, S. Sherchan, S. Liu, Optimizing deep learning models for climate-related natural disaster detection from uav images and remote sensing data. J Imaging. **11**, 2, 32 (2025). doi:10.3390/jimaging11020032
16. Z. Hong, H. Zhong, H. Pan, J. Liu, R. Zhou, Y. Zhang, Y. Han, J. Wang, S. Yang, C. Zhong, (2022). classification of building damage using a novel convolutional neural network based on post-disaster aerial images. Sensors, **22**, 15, 5920 (2022). https://doi.org/10.3390/s22155920
17. N. Alsaaran and A. Soudani, Deep learning image-based classification for post-earthquake damage level prediction using UAVs. Sensors. **25**, 17, 5406 (2025). https://doi.org/10.3390/s25175406
18. AIDER (Aerial Image Dataset for Emergency Response Applications) Version 1.0, https://zenoda.org/records/3888300. Access date 14/12/2025.
19. X. Deng, M. Shi, B. Khan, Y. Choo, F. Ghaffar, C. Peng Lim, A lightweight CNN model for UAV-based image classification. Soft Comput **29**, 2363–2378 (2025). https://doi.org/10.1007/s00500-025-10512-3
20. W. Chen, J. Geng, F. Meng, L. Zhang, Pseudo-Ł0-Norm fast iterative shrinkage algorithm network: agile synthetic aperture radar imaging via deep unfolding network. Remote Sens. **16**, 4, 671 (2024). https://doi.org/10.3390/rs16040671
21. A. Tran, M. Tran, E. Marti, J. Cothren, C. Rainwater, S. Eksioglu, N. Le, Land8Fire: A Complete Study on Wildfire Segmentation Through Comprehensive Review, Human-Annotated Multispectral Dataset, and Extensive Benchmarking, Remote Sens. **17**, 16, 2776 (2025). https://doi.org/10.3390/rs17162776
22. M. Nurtas, S. Nurakynov, A. Altaibek, A. Mergembayeva , M. Mohammed, Satellite based deep learning approaches for detecting environmental disasters across Kazakhstan. Discov. Appl. Sci. 8, 144 (2026). https://doi.org/10.1007/s42452-025-08133-4
23. J. Kim, P. A. Campbell, K. Calhoun, a framework to predict community risk from severe weather threats using probabilistic hazard information (PHI). Atmosphere. **14,** 5, 767 (2023).. https://doi.org/10.3390/atmos14050767
24. J. Shao, J., Tang, L., Liu, M., Shao, G., Sun, L., Q. Qiu, BDD-Net: A general protocol for mapping buildings damaged by a wide range of disasters based on satellite imagery. Remote Sens. **12**, 10, 1670 (2020). https://doi.org/10.3390/rs12101670
25. J. Jia and W. Ye, Deep learning for earthquake disaster assessment: objects, data, models, stages, challenges, and opportunities. Remote Sens. **15**, 16, 4098 (2023). https://doi.org/10.3390/rs15164098

26. D. P. Yadav, B. Sharma, S. Singh, P. Liatsis, Hybrid lightweight transformer for efficient landslide change detection in remote sensing imagery. Sci Rep **16**, 2111 (2026). https://doi.org/10.1038/s41598-025-31888-0V.
27. A. Ahmed, K. Jouini, A. Tuama, and O. Korbaa, A fusion approach for enhanced remote sensing image classification, in Proceedings of the 19th Int. Joint Conf. Comput. Vis., Imaging Comput. Graph. Theory Appl. (VISIGRAPP). **2**: 554–561 (2024). DOI: 10.5220/0012376600003660
28. E. O. Yilmaz and T. Kavzoglu, Burned area detection with Sentinel-2A data, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. **X-5-2024**, (2024). https://doi.org/10.5194/isprs-annals-X-5-2024-251-2024
29. K. VanExel, S. Sherchan, S. Liu, Optimizing deep learning models for climate-related natural disaster detection from UAV images and remote sensing data. J Imaging. **11**, 2, 32(2025) doi:10.3390/jimaging11020032. PMID: 39997534; PMCID: PMC11856490.