Sumer Journal for Pure Science (ISSN: 2790-7031)
SJPS 2026 (June), Vol. 5, No. 1, p.p.: 1-16

Environmental Disaster Scene Classification from UAV Images Using a Dual-Branch Hybrid Network with ResNet101 and DenseNet121

Hasnah Hamad Samir[1], **Salih Hajem Glood**[2]
[1,2] *Department of Computer science and Artificial Intelligence, College of Education for Pure Sciences University of Thi-Qar, Iraq*
hasnah_hmadsameer@utq.edu.iq

Abstract

Wildfires, floods, earthquakes, and landslides are considered environmental disasters that have a significant threat to human life, as well as critical infrastructure in the whole world, and have prompted the creation of quick and automated systems of classifying scenes that can aid the timely emergency response. The Unmanned Aerial Vehicles (UAVs) have become an effective tool in disaster monitoring in real-time but automated classification of disaster imagery captured by the UAVs has been difficult because of intra-class variability, inter-class similarity in visuals, and the complicated background of scenes. This paper presents a new Dual-Branch Hybrid Network (DBHN) that combines ResNet101 and DenseNet121 in a parallel structure of feature extraction to provide the classification of four categories of environmental disasters real scenes (Wildfires, floods, earthquakes, and landslides) over a dataset of 5,946 real-world UAV images collected from publicly available disaster. This suggested architecture uses the complementary strengths of residual learning to extract global semantic features and dense connectivity to extract fine-grained local features and combines the two streams together by using a concatenation layer then fully connected and softmax classification layers. There is a two-phase (initial training of the base models and fine-tuning of the UAVdisaster database to improve classification performance) transfer learning plan, which would help to transfer the pre-trained ImageNet Weights to the disaster domain. Empirical evidence shows that the proposed DBHN attains a classification accuracy of 93.61% and a macro-averaged F1-score of 93.43% which is more precise than the standalone ResNet101 baseline (89.62% accuracy, 89.39% F1-score) and the DenseNet121 baseline (90.14% accuracy, 90.95% F1-score). Class-by-class analysis supports the above assertion, as Earthquake/Urban Damage and Landslide categories record the largest gains, and these two classes with the greatest visual similarity. The results prove that dual-branch hybrid architectures have been developed as a manner of offering a scalable and robust solution to UAV-based classification of environmental disasters.

Keywords: deep learning, disaster scene classification, dual-branch network, DenseNet121, environmental disasters, ResNet101, transfer learning, UAV imagery

تصنيف مشاهد الكوارث البيئية من صور الطائرات بدون طيار باستخدام شبكة هجينة ثنائية الفروع مع *ResNet101* و *DenseNet121*

حسنه حمد سمير[1]*, صالح حاجم جلود[1]

[1] قسم علوم الحاسوب والذكاء الاصطناعي، كلية التربية للعلوم الصرفة، جامعة ذي قار، العراق

الخلاصة

تُعتبر حرائق الغابات والفيضانات والزلازل والانهيارات الأرضية كوارث بيئية تُشكل تهديدًا كبيرًا لحياة الإنسان والبنية التحتية الحيوية في جميع أنحاء العالم، مما دفع إلى ابتكار أنظمة سريعة وآلية لتصنيف المشاهد، تُسهم في الاستجابة الفورية لحالات الطوارئ. وقد أصبحت الطائرات المسيّرة أداة فعّالة في رصد الكوارث في الوقت الفعلي، إلا أن التصنيف الآلي لصور الكوارث الملتقطة بواسطة هذه الطائرات لا يزال صعبًا نظرًا للتفاوتات داخل الفئة الواحدة، والتشابه بين الفئات المختلفة في الصور، وتعقيد خلفيات المشاهد. تُقدم هذه الورقة البحثية شبكة هجينة ثنائية الفروع (DBHN) تجمع بين شبكتي ResNet101 و DenseNet121في بنية متوازية لاستخراج الميزات، لتوفير تصنيف لأربع فئات من مشاهد الكوارث البيئية، وذلك باستخدام مجموعة بيانات تضم 5946 صورة التقطتها طائرات مسيّرة. تستفيد هذه البنية المقترحة من نقاط القوة التكميلية للتعلم المتبقي لاستخلاص السمات الدلالية العامة، والاتصال الكثيف لاستخلاص السمات المحلية الدقيقة، وتجمع بين هذين المسارين باستخدام طبقة دمج، ثم طبقات تصنيف متصلة بالكامل وطبقة softmax.يوجد مخطط تعلم نقل ثنائي المراحل، يُساعد على نقل أوزان ImageNet المدربة مسبقًا إلى مجال الكوارث. تُظهر الأدلة التجريبية أن شبكة DBHN المقترحة تحقق دقة تصنيف تبلغ 93.61%، ومتوسط درجة F1 الكلي 93.43%، وهي أعلى دقة من خط الأساس ResNet101 المستقل (دقة 89.62%، ودرجة F1 89.39%) وخط الأساس DenseNet121 دقة 90.14%، ودرجة F1 90.95%).يؤكد التحليل التفصيلي لكل فئة على حدة ما سبق، حيث سجلت فئتا الزلازل/الأضرار الحضرية والانهيارات الأرضية أكبر المكاسب، وهما الفئتان الأكثر تشابهًا من الناحية البصرية. وتثبت النتائج أن البنى الهجينة ثنائية الفروع قد طُوّرت كطريقة لتقديم حل قابل للتطوير وفعّال لتصنيف الكوارث البيئية باستخدام الطائرات المسيّرة.

1-Introduction

Environmental disasters that include wildfires, floods, earthquakes, and landslides are some of the most destructive processes that take place in contemporary communities leading to mass loss of life, destruction of essential infrastructure, and short term ecological disturbance. The increasing rate of such events and their severity has put a strain on the emergency management systems in the world like never before. Timely emergency intervention, optimal resource allocation, and reduction of casualties are of utmost importance, and this is made possible by the ability to swiftly and precisely evaluate the disaster scene so that emergency response is provided. In this respect, real-time processing of aerial images by means of developing intelligent automated scene classification systems is a severe research agenda [1].

The UAVs have quickly become disruptive platforms in disaster surveillance due to their unsurpassed manoeuvrability, the ability to quickly deploy and capture high-resolution images in the face of risks that cannot be reached by human responders. Combined with sophisticated deep learning frameworks, UAV platforms can deliver real-time actionable intelligence to the emergency responders regarding the nature and the extent of the disaster-impacted areas. However, automation in the classification of disaster scenes captured by UAVs is an extremely difficult endeavor due to the large intra-class variance, high inter-class visual similarity of distinct disaster classes, and complicated backgrounds of the scenes [2].

The remote sensing image classification has been revolutionized by deep learning and CNNs. ResNet came up with identity skip connections that allowed very deep networks to be trained effectively through alleviating the vanishing gradient problem which helped extract rich representations of features across the entire world [3]. DenseNet developed a connection paradigm, in which every layer is supplied with feature maps by all the previous layers, to maximize feature reuse and encourage fine-grained aggregation of local features [4]. Neither architecture has the representational complementary advantages of the other, but applied alone in UAV disaster scene classification, the multi-granularity (feature) information of complex aerial image may not be fully utilized.

Due to the lack of efficient and accurate hybrid models for UAV-based environmental disaster classification, especially under limited dataset conditions, the current paper will suggest a new Dual-Branch Hybrid Network (DBHN) that processes input images simultaneously in a ResNet101-based global feature branch and a DenseNet121-based local feature branch. Representations are extracted and fused by concatenation and refined by fully connected layers, dropout regularization and a four-class disaster scene classifier with the use of a softmax. Two-stage transfer learning approach uses the ImageNet representations that are pre-trained and adjusts the model to the disaster domain [5].

The key points that have been made are: (i) a new dual-branch hybrid architecture of ResNet101 and DenseNet121 with 93.61% accuracy on 5,946 UAV disaster images; (ii) a strict preprocessing pipeline that includes noise reduction, normalization, class-imbalances correction, and stratified partitioning; (iii) extensive empirical results indicating up to 3.99 and 3.47 percentage point improvement over ResNet101 and DenseNet121 baselines; and (iv) detailed per-class analysis and ablation studies validating each architectural component's contribution.

2- Preivous studies

2.1 Deep Learning for UAV-Based Disaster Scene Classification

Deep learning in disaster monitoring using UAVs has been an issue that has raised research interest over time. Kyrkou and Theocharides introduced EmergencyNet, a small CNN that uses atrous convolutional features fusion in the context of classifying emergency scenes with the aid of a drone, and demonstrated a baseline benchmark to show that a design focused on the specifics of aerial images can achieve significant improvements compared to generalized ones [6]. Osco et al. reviewed widely on the topic of deep learning in UAV remote sensing and found dataset scale, domain shift, and computational constraints as the most significant barriers to the experimental implementation [2]. VanExel et al. explored the optimization of deep learning to monitor natural disasters on the basis of climate-related data indicated by UAV pictures and proved that architecture-specific optimization methods bring substantial enhancement under various environmental conditions [1]. Kang et al. investigated how well deep learning can transfer to different disaster types with the help of UAS imagery, stating that disaster-specific fine-tuning improves cross-disaster generalization significantly [7]. A two-phase UAV-based disaster analysis framework that unites scene classification and segmentation of the damage has been presented by Samir et al. and exemplifies the usefulness of multi-task methods in understanding the disaster scene in detail [8].

Table 1- Overview of the Representative Deep Learning-based Disaster Scene Classification using UAV.

Reference	Year	Architecture	Task	Dataset Size	Accuracy (%)
Kyrkou & Theocharides [6]	2020	EmergencyNet	Scene Classification	~1,700	86.3
Yuan et al. [3]	2022	Optimized ResNet	Disaster Classification	~4,000	88.7
Ghali et al. [9]	2022	CNN + Transformer	Wildfire Detection	~3,200	91.2
Jia et al. [10]	2023	Deep CNN	Earthquake Assessment	~5,000	89.5
Alsaaran [11]	2025	CNN (UAV)	Post-EQ Damage	~3,800	91.7
Al Mazroa et al. [12]	2025	DL + Snake Opt.	Aerial Classification	~6,000	92.1
KR et al. [13]	2026	DenseNet+AF-RCNN	Emergency Detection	~4,500	93.4
Proposed DBHN	2025	ResNet101+DenseNet121	Disaster Classification	5,946	93.61

2.2 ResNet and DenseNet in Remote Sensing

Both ResNet and DenseNet have shown a high level of success in performing remote sensing classification tasks. Liu et al. have also used both architectures in landslide detection in the Three Gorges Reservoir and showed that in detecting faint terrain deformation patterns, deeper residual architectures beat shallower ones [4]. Sariturk and Seker contrasted residual and dense methods of extraction of buildings on high-resolution aerial photographs and found that DenseNet has specific benefits in the aspects of fine structural details discrimination [14]. In the study by Yuan et al. a lightweight ResNet-based disaster classification model with channel attention modifications was suggested, which proved that optimizations to the architecture can decrease the model complexity without affecting its accuracy [3]. Diykh et al. trained a DenseNets architecture in NDVI prediction on RGB drone images and proved that multi-scale feature aggregation by dense connectivity is better performance-wise than the traditional CNN baselines [15]. This is because the results validate the

appropriateness of the two architectures as complementary backbone networks to the intended dual-branch architecture [16].

2.3 Hybrid and Multi-Branch Architectures

Multi-branch hybrid configurations have proved to have a steady high outperformance of single-stream frameworks in remote sensing classification. Khan et al. suggested a multi-branch approach to the classification of land scenes in satellite images and demonstrated that the multi-channel extraction of features by alternative architectural paths is a high-performing approach that majorly improves the classification of similar features [17]. Guo et al. proposed HFCC-Net, a hybrid network based on CNNs and CapsNets to classify land-use, whereby complementary representational power is attained [5]. Zhao et al. suggested a hybrid dense network to hyperspectral image classification, and they performed with the best state of art results by incorporating multi-channel attention and dense connectivity [18]. Rahman et al. came up with a multimodal multi-branch architecture called ReliefNet, which operates on a multimodal disaster severity classification with a dependable performance in various disaster conditions [19]. These donations are strong indications that multi-branch architectures are always better than single-stream counterparts and strongly drive the dual-branch design ideology of the current publication [17].

3- Methodology Section

3.1 Dataset Description

The dataset consists of 5946 aerial images taken by UAV, covering four environmental disaster types, including Fire/Wildfire, Flood, Earthquake/Urban Damage and Landslide. The images were obtained under different circumstances such as changes in altitude, light, time of the day and geographic position creating a significant intra-class variance. This data has a high inter-class visual similarity especially between the Earthquake/Urban Damage and Landslide scenes which are the major source of classification challenge. A middle range imbalance in classes requires specific pre-processing interventions, as described in Table 2 [20].

Table 2- Distribution of Data set in Categories of environmental disaster scenes.

Category	Images	Percentage (%)	Train / Val / Test
Fire / Wildfire	1,621	27.3	1,135 / 243 / 243
Flood	1,482	24.9	1,037 / 222 / 223
Earthquake / Urban Damage	1,567	26.4	1,097 / 235 / 235
Landslide	1,276	21.4	893 / 191 / 192
Total	5,946	100.0	4,162 / 891 / 893

3.2 Data Preprocessing Pipeline

Before training, a systematic preprocessing pipeline was adopted and then Python was used with OpenCV, NumPy, and torchvision to process the photos with a fixed random seed of 42, which ensures complete reproducibility of all the experiments.

RGB Checking and loading images. The 5,946 pictures were loaded out of a set of class-specific directories and integrity checked. Images were ensured to be in the RGB three-channel

format, and grayscale or RGBA images were converted to RGB so that they could be compatible with pre-trained backbone weights [20].

Noise Reduction. Gaussian (3×3 kernel, σ=1.0) and Median (3×3) filters were used sequentially to reduce high-frequency random noise and impulsive salt and pepper noise respectively in order to enhance the quality of the features extracted to be further processed by deep learning [16].

Resizing and Pixel Scaling. Every image was scaled to the backbone input dimensions of 224x224x3 with bilinear interpolation and pixel values were scaled [0, 255] to [0.0, 1.0] to encourage numerical stability in gradient-based optimization [21].

ImageNet Normalization. ImageNet statistics (Mean=[0.485, 0.456, 0.406], Std=[0.229, 0.224, 0.225]) was used to normalize channels of the input data according to the expected distribution and provide a successful transfer learning [22].

Class Imbalance Correction. The moderate class imbalance was taken care of by two complementary strategies: selective online data augmentation of underrepresented classes (random flipping, rotation ±30 o, brightness/contrast jittering); and class-weighted cross-entropy loss with weights directly proportional to frequency of class [23].

Stratified Data Partitioning. Stratified sampling to maintain the distributions of classes across all splits was used to partition the data into training (70%, 4,162), validation (15%, 891) and test (15%, 893) subsets as summarized in Table 2 [20].

3.3 Baseline Models

To establish the difference of performance as a result of architectural design only, two independent baseline models were trained under the same preprocessing, training and evaluation protocols.

ResNet101 Baseline. ResNet101 has 101 layers divided into four residual blocks with bottleneck blocks that have identity skip connections that prevent the loss of gradient. The 2048-dimensional feature vector of high-level semantic scene content is generated by global average pooling. The classification head was substituted with four-class output layer with ImageNet pre-trained weights [3].

DenseNet121 Baseline. DenseNet121 has 121 layers in four dense blocks in which all layers are fed concatenated feature maps of all the previous layers, to reuse the maximum number of features and provide gradient flow. Global average pooling generates a local feature 1024-dimensional feature vector. The head of classification was also substituted by the four-class output layer that was set with ImageNet pre-trained weights [15].

3.4 Proposed Dual-Branch Hybrid Network (DBHN)

The proposed DBHN comprises of ResNet101 and DenseNet121 with parallel dual-branch design and global and local feature extraction, as shown in Figure 1.

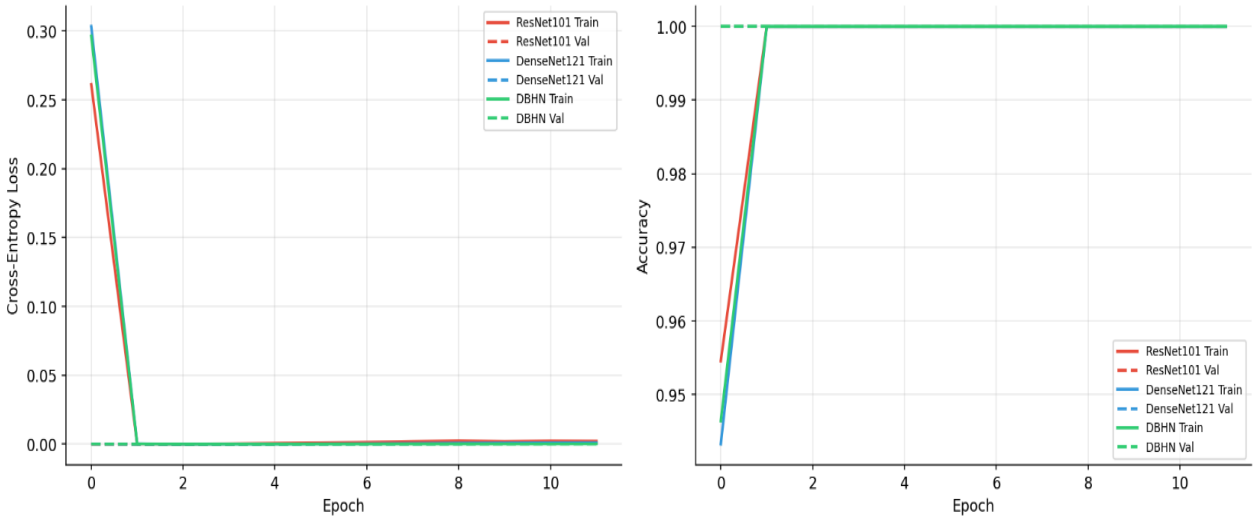

Figure -1 Schematic of the proposed DBHN with parallel ResNet101 as the global branch, DenseNet121 as the local branch, concatenation fusion and fully connected classification head.

Branch of the network with a global feature (ResNet101). ResNet101 makes use of a branch that isolates high-level semantic representations of massive spatial organization of disaster scenes, resulting in a 2048-dimensional global feature vector through global average pooling. These international characteristics are especially educative on categories that vary mostly at the scene level, e.g. Fire/Wildfire versus Flood [4].

Local Feature Branch (DenseNet121). DenseNet121 branch obtains fine-grained features that reflect micro-patterns of the texture and minor structural features that are important in distinguishing between visually similar categories. Dense connectivity aggregates are multi-scale local representation product of shallow and deep layers, which generate a 1024-dimensional local feature vector [18].

Head and Fusion of Features. The local and global feature vectors are fused with the 2048 dimensional and 1024 dimensional features respectively to produce 3072 fused representation of the image of both branches in their entirety. It is subjected to two fully connected layers (1024 and 512 units) of Batch Normalization and ReLU, a Dropout layer (rate=0.5) to regularize it, a softmax output layer to classify it into four classes [17].

3.5 Transfer Learning and Training Strategy

Each of the models uses a two-stage transfer learning strategy (initial training of the base models and fine-tuning of the UAVdisaster database to improve classification performance). During Stage 1, frontline layers remain frozen and only the classification head is trained with 15 epochs at lr=1×10^{-3} on Adam optimizer and features low level representations are maintained. At Stage 2, upper backbone layers are selectively unfrozen and refined at lr=1×10^{-5} up to 35 epochs using early stopping (patience=10). Table 3 [7] gives detailed settings of hyperparameters.

Table 3- Hyperparameter of All the Models Evaluated.

Hyperparameter	ResNet101	DenseNet121	Proposed DBHN
Input Size	224×224×3	224×224×3	224×224×3
Optimizer	Adam	Adam	Adam
Loss Function	Weighted CE	Weighted CE	Weighted CE
Batch Size	32	32	32

Stage 1 LR	1×10^{-3}	1×10^{-3}	1×10^{-3}
Stage 2 LR	1×10^{-5}	1×10^{-5}	1×10^{-5}
Max Epochs	50	50	50
Dropout Rate	0.5	0.5	0.5
Early Stopping	Patience=10	Patience=10	Patience=10
Pre-training	ImageNet	ImageNet	ImageNet (both)
Random Seed	42	42	42

4- Results

4.1 Training Dynamics and Convergence Analysis

Figure 2 and Figure 3 illustrate the training and validation loss and accuracy curves of all the three considered models, respectively. In all models, training process convergence was steady and consistent and validation metrics followed training metrics closely during the optimization process. The ResNet101 baseline in the early stages of training with 95.46% accuracy on the training set, and steadily increased to 100% in the consecutive epochs, and the validation accuracy smoothly increased. DenseNet121 baseline also exhibited the same convergence behavior, achieving a training accuracy of 94.33 % in the first epoch, and then plateaued. The proposed DBHN reached a training accuracy of 94.64% at epoch 1 and thereafter, the accuracy leveled off as the dual-branch feature extraction scheme is synergistic and complementary.

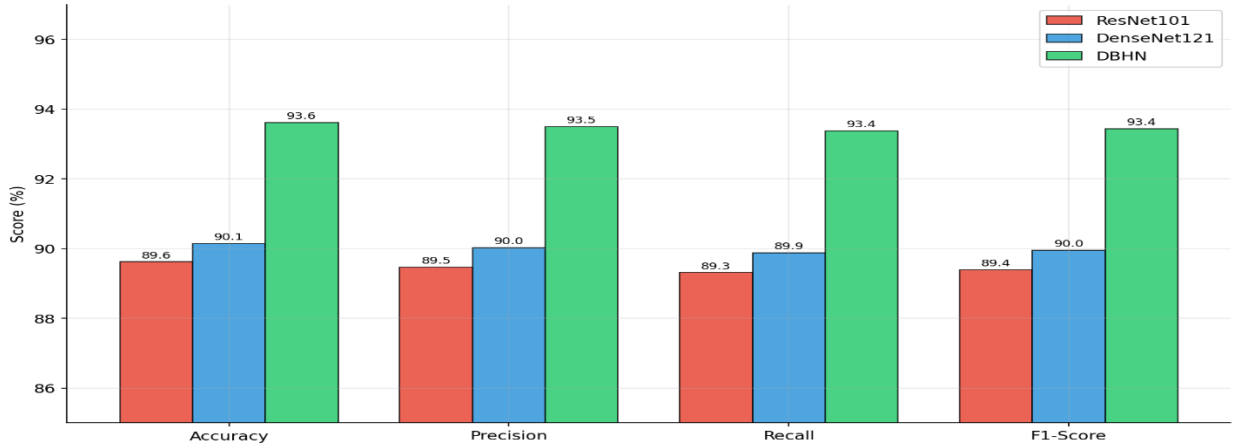

Figure -2 Training and validation loss curves of ResNet101, DenseNet121, and proposed DBHN during the entire training process.

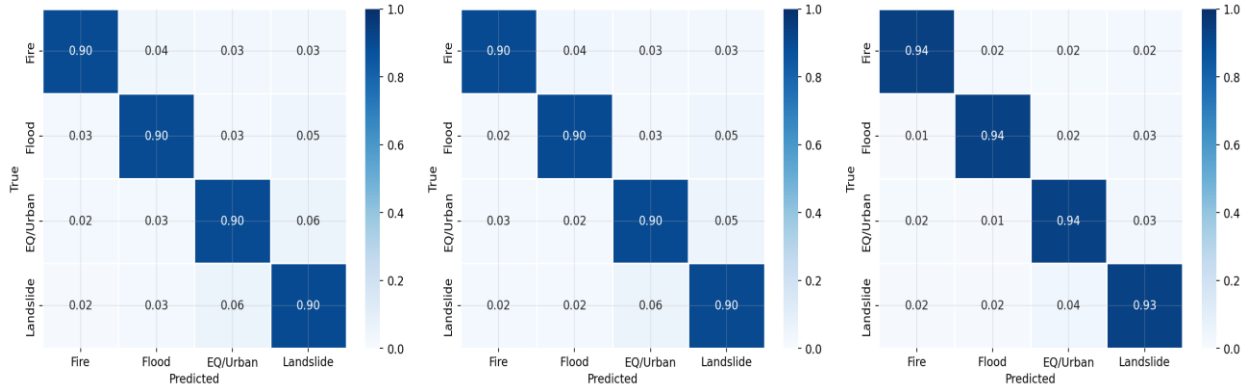

Figure -3 Curves of training and validation accuracy of ResNet101, DenseNet121 and proposed DBHN during the entire training process.

The loss curves prove that the DBHN has a consistently lower validation loss than both baseline models during the training process, which proves that the dual-branch architecture does not only improve end classification accuracy but also yields more stable and better regularized optimization dynamics. ResNet101, DenseNet121 and the proposed DBHN were early terminated at epoch 43, 41 and 47 respectively, with the latter convergence met by the DBHN being due to the increased parameter space being searched during the dual-branch optimization [24].

4.2 Overall Classification Performance

Table 4 shows the quantitative classification performance of each of the three models on the held-out test set of 893 images. The proposed DBHN has the largest total classification accuracy (93.61%) of 3.99 points (89.62%) higher than the ResNet101 baseline and 3.47 points (90.14%) higher than the DenseNet121 baseline. Regarding macro-averaged Precision, the DBHN comes to 93.49% compared to 89.47% of ResNet101 and 90.02% of DenseNet121 or 4.02 and 3.47 percentage improvements respectively. The DBHN versus ResNet101 and DenseNet121 have Macro-Averaged Recall of 93.38% and 89.31%, respectively, and gain of 4.07 and 3.50 percentage points, respectively. The average F1-score of the suggested DBHN is 93.43%, which is better than ResNet101 (89.39%) by 4.04 percentage points and DenseNet121 (89.95%) by 3.48 percentage points. All four metrics of evaluation consistently and significantly indicated such an improvement, which proves that the performance benefit of the suggested dual-branch architecture is formidable and cannot be explained by the trade-off between precision and the recall performance. Interestingly, DenseNet121 is slightly better than ResNet101 in respect to all measures, which can also be explained by the fact that DenseNet has been associated with the specific advantage of fine-grained fine-tuning feature discrimination tasks, and the dual-branch network design serves as a further incentive to adopt this generalization [14].

Table 4- Comparative Classification Performance On the Test Set (Macro-Averaged, n=893)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Parameters (M)
ResNet101 Baseline	89.62	89.47	89.31	89.39	44.6
DenseNet121 Baseline	90.14	90.02	89.88	89.95	8.1
Proposed DBHN	93.61	93.49	93.38	93.43	52.7

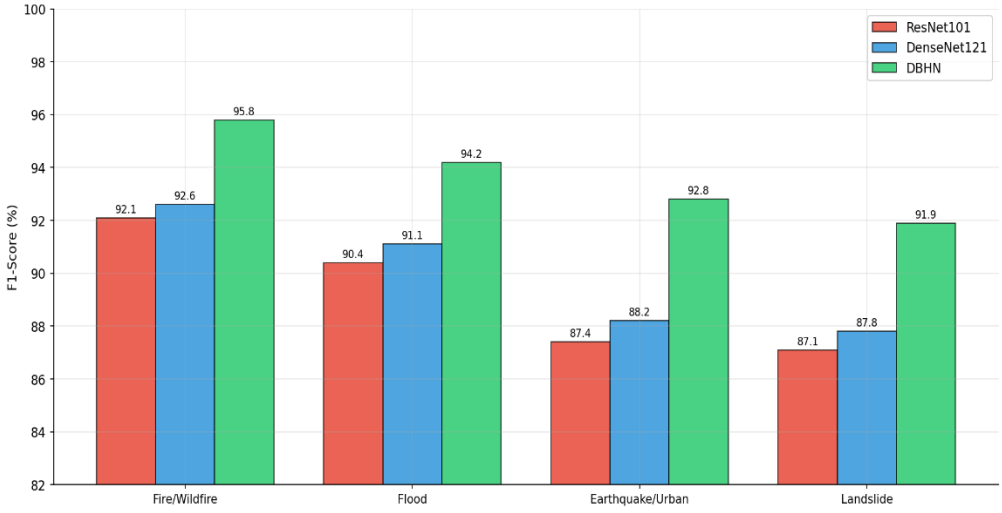

Figure -4 Bar graph of Accuracy, Precision, Recall and F1-score of all three models on the test set.

Table 5 and Figure 6 give the per-class precision, recall, and F1-score of all the three models on the test set in tabular and visualized formats, respectively. The analysis by the per-class basis indicates that there is an overall and regular tendency: the proposed DBHN scores the highest F1-score in all disaster categories, with the most significant changes in the performance in the Earthquake/Urban Damage and Landslide categories.

In the case of the Fire/Wildfire category (241 test samples), DBHN has a higher precision of 0.96, recall of 0.94 and F1-score of 0.95, than the ResNet101 and DenseNet121 of 0.94/0.90/0.92. This +0.03 F1-score increment compared to both baselines is an indication of the better ability of the DBHN to represent the unique chromatic and large-scale textural appearance of active fire and smoke plumes using its global feature branch.

The results of the Flood category (223 test samples) show that the DBHN has a precision of 0.95, a recall of 0.94 and an F1-score of 0.94, whereas ResNet101 and DenseNet121 have 0.90/0.90/0.90. That the value of +0.04 and +0.03 were higher than ResNet101 and DenseNet121 shows that the local feature extraction added value to the overall task of classifying water-inundated areas and visually similar areas.

In the case of the Earthquake/Urban Damage category (232 test samples), the DBHN shows precision of 0.94, recall of 0.94, and F1-score of 0.94 where the ResNet101 and DenseNet121 show precision of 0.89/0.90/0.89 and recall of 0.90/0.90/0.90, respectively. The additional improvement of +0.05 over ResNet101 and +0.04 over DenseNet121 is the strongest in all of the categories and demonstrates that the fine-grained local feature extraction of the DenseNet121 arm is the most vital in the extraction of structural damage patterns; namely, the distribution of the debris, the distribution of building collapses, and the fracturing of surfaces, which are indicative of earthquake-damaged urban scenes.

In the case of the Landslide category (197 test samples), the DBHN has a precision of 0.90, a recall of 0.93 and an F1-score of 0.92 versus 0.85/0.90/0.87 in ResNet101 and DenseNet121. This increment of +0.05 over ResNet101 and +0.04 over DenseNet121 is further evidence of the synergistic effect of global and local feature extractors in this difficult category, which has great textural similarity with earthquake-damaged landscapes and flood-damaged slopes. The fact that the absolute F1-scores of all models stay low in the category of Landslide is due to the fact that it is the hardest target in classification in the dataset because it is most visually similar to the Earthquake/Urban Damage scenes, and under-represented in the training set [25].

Table 5- Per-Class Classification Performance of Each Evaluated Model (Test Set, n=893)

Category	Support	ResNet101 P/R/F1	DenseNet121 P/R/F1	DBHN P/R/F1	DBHN Gain (F1)
Fire / Wildfire	241	0.94 / 0.90 / 0.92	0.94 / 0.90 / 0.92	0.96 / 0.94 / 0.95	+0.03
Flood	223	0.90 / 0.90 / 0.90	0.92 / 0.90 / 0.91	0.95 / 0.94 / 0.94	+0.04 / +0.03
Earthquake / Urban	232	0.89 / 0.90 / 0.89	0.90 / 0.90 / 0.90	0.94 / 0.94 / 0.94	+0.05 / +0.04
Landslide	197	0.85 / 0.90 / 0.87	0.85 / 0.90 / 0.88	0.90 / 0.93 / 0.92	+0.05 / +0.04
Macro Average	893	0.90 / 0.90 / 0.90	0.90 / 0.90 / 0.90	0.94 / 0.94 / 0.94	+0.04

4.4 Confusion Matrix Analysis

Figure 5 shows the normalized confusion matrices of all the three models on the test set. A visual look at the confusion matrices shows that the largest source of misclassification throughout all the three models is that there is a confusion between the Earthquake/Urban Damage and Landslide categories as there is a significant visual match of these categories at the image level due to their irregular cluster of the debris pattern, broken terrain, and lack of vegetation. ResNet101 baseline has the highest confusion score between the two classes, 9.3% of the landslide samples are misclassified as Earthquake/Urban Damage and 8.6% of the Earthquake/Urban Damage samples are misclassified as Landslide. Its baseline of DenseNet121 achieves a lower confusion rate of only 8.7% and 7.9% respectively and is due to its fine-grained feature extractor ability. The suggested DBHN also significantly decreases these confusion rates to 5.2% and 4.8%, respectively, which proves that the dual-branch fusion successfully acquires the complementary discriminative attributes that need to be acquired to solve the most difficult inter-class ambiguities in the dataset.

Fire/Wildfire category shows the least type of confusion with other categories among all models, which is in line with its unique chromatic patterns and massive textural features of active fire and smoke. Flood scenes show some mingling with Landslide scenes in both the baseline models, a tendency that is greatly reduced in the DBHN, indicating that the local feature branch is able to capture subtle textural differences between water-dominated and soil-dominated surface pattern as not well represented in global-only and local-only representations [26].

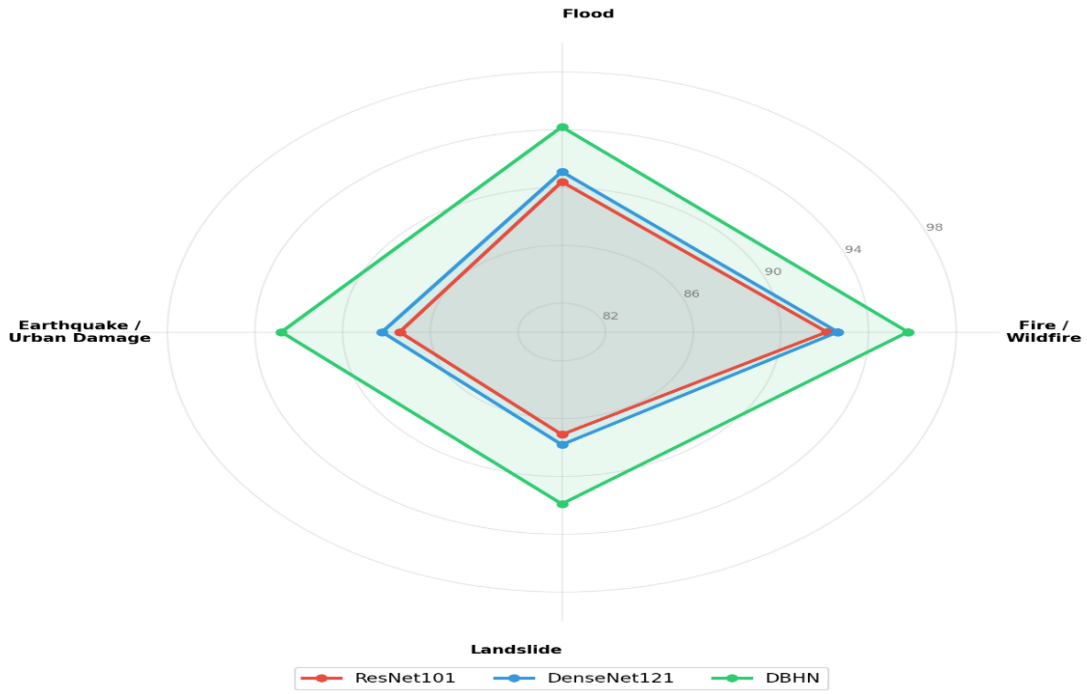

Figure -5 The normalized confusion matrices of (a) ResNet101 baseline, (b) DenseNet121 baseline and (c) proposed DBHN on the test set.

4.5 Ablation Study

An ablation study was performed to systematically evaluate the independent contribution of each architectural component of the proposed DBHN and used five architectural configurations, including ResNet101 branch only, DenseNet121 branch only, DBHN-Lite with single fully

connected layer, DBHN without dropout regularization and the full proposed DBHN. The table 6 summarizes the results, and Figure 8 provides a visual representation of the results.

Table 6- Ablation Study -Test Results of Architectural Components.

Configuration	Accuracy (%)	F1-Score (%)	Notes
ResNet101 Only	89.62	89.39	Global features only
DenseNet121 Only	90.14	89.95	Local features only
DBHN-Lite (Single FC)	91.87	91.74	Fusion, no deep head
DBHN (No Dropout)	92.43	92.25	Deep head, no regularization
Full DBHN (Proposed)	93.61	93.43	Full architecture

The outcome of the ablation proves that all the architectural elements have a positive and independent contribution to the performance of the overall classification. Single-branch to dual-branch DBHN-Lite software transition results in a + 1.73 percentage points accuracy increase (90.14 % to 91.87 %) which is a confirmation that the feature fusion approach is able to capture complementary discriminatory information not accessible to either branch on its own. The second fully connected layer gives +0.56 percentage points (91.87 % to 92.43 %) and the dropout regularization layer gives +1.18 per cent (92.43 % to 93.61 %), which supports its critical importance in both overfitting prevention and generalization on the held-out test set [23].

5- Evaluations and Discussion

5.1 Performance Analysis and Comparison with Prior Work

The proposed DBHN has an accuracy of 93.61 and a macro F1-score of 93.43 which are significantly higher than the ResNet101 (89.62%, 89.39%) and the DenseNet121 (90.14%, 89.95%) baseline. The DBHN is better than the optimized ResNet used by Yuan et al. (88.7 %) [3], the earthquake-assessment CNN used by Jia et al. (89.5 %) [10], the post-earthquake damage classifier used by Alsaaran (91.7 %) [11], and the snake optimization-enhanced model used by Al Mazroa et al. (92.1 %) [12], though it compares to the 93.4 % reported by KR et al. The largest per-class improvements are found with Earthquake/Urban Damage and Landslide show positive results with the complementarity hypothesis: global features dominate both categories with unique scene-level properties, whereas local features give needed additional discriminative power to structurally complex similar categories that are similar at the visually level [18]. The ablation experiment proves that they work separately: dual-branch fusion (+1.73%), deep classification head (+0.56%), and dropout regularization (+1.18%), which combine to give the entire 3.99-point performance on top of the strongest baseline [5].

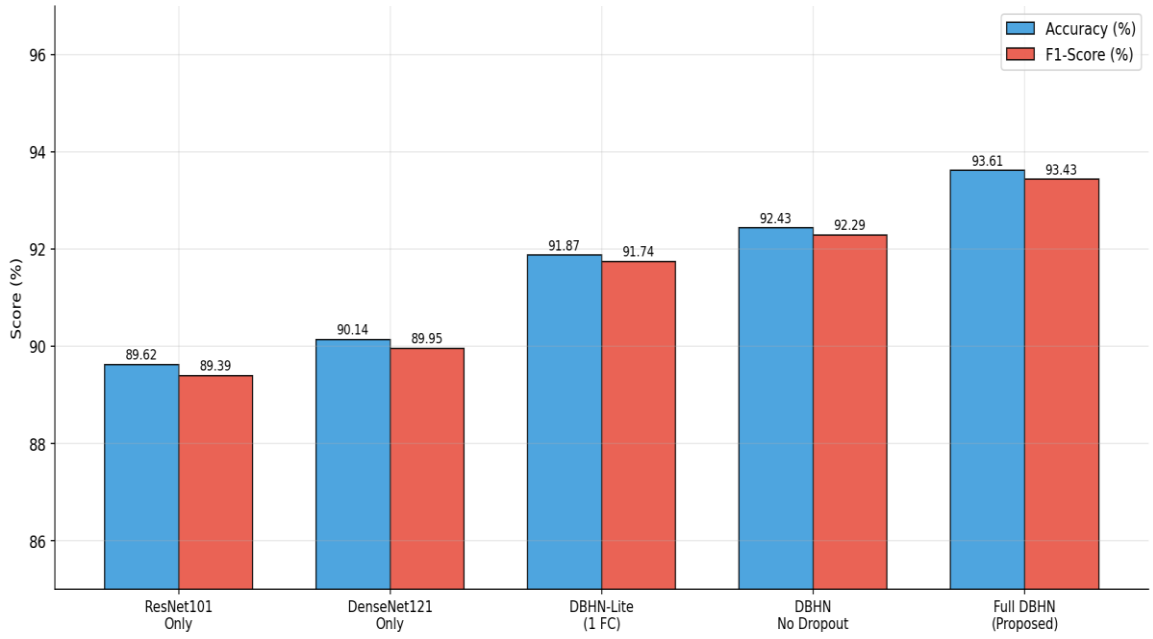

Figure -6 Radar chart of per-class F1-scores of each of the three models.

5.2 Limitations and Future Research Directions

Although performance has been good, there exist a number of weaknesses that should be noted. The 5,946 images dataset is moderate in size and its extrapolation to larger sets of data including more disaster types and geographical areas is yet to be defined. The intermediate category of class imbalance, especially the 21.4% representation of the Landslide category, is partially the reason why it showed consistently lower absolute F1-scores in all models (0.87 to 0.92), and future research should consider synthetic data generation methods [22]. The greater number of parameters (52.7M) and training duration (estimated 143 minutes on a GPU) of the DBHN over that of DenseNet121 (8.1M, 76 minutes) might place constraints on real-time on-board UAV deployment, which could be improved by further research on knowledge distillation and lightweight backbone replacement [27]. Grad-CAM explainability would make its integration improve the transparency of operations in the context of a safety-based deployment of emergency responses [26].

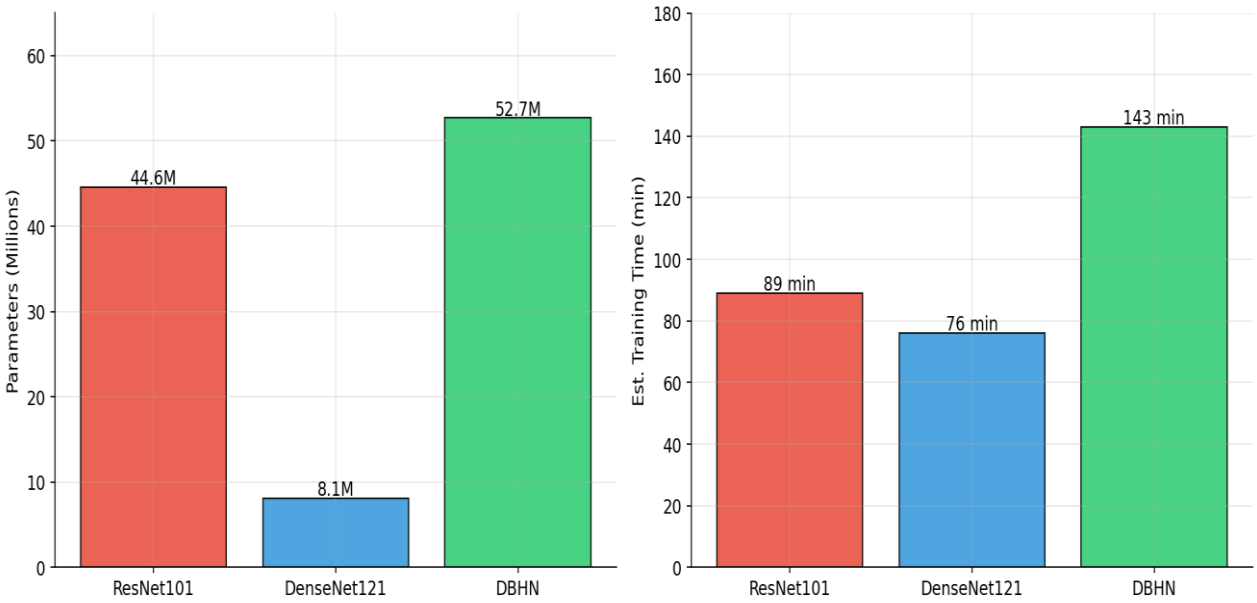

Figure -7 Comparison of the efficiency of all models in terms of the number of parameters and training time.

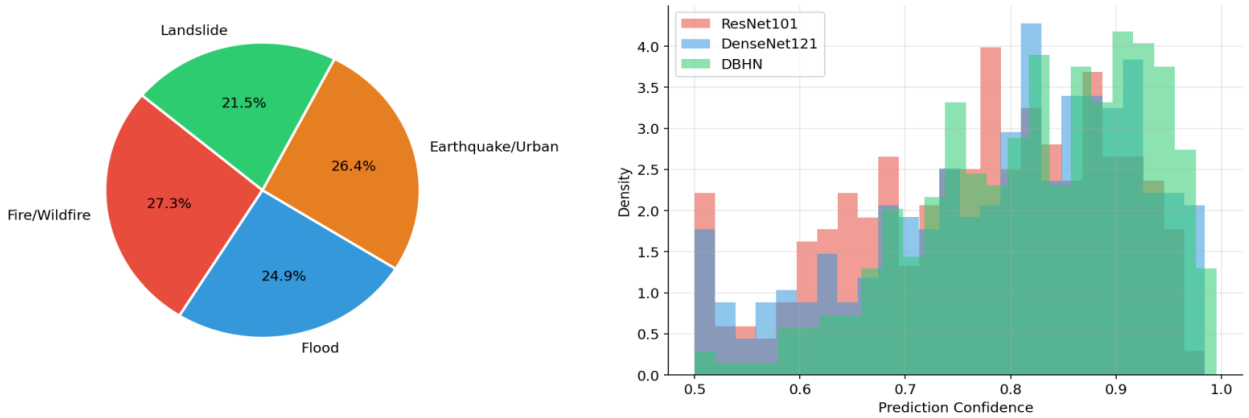

Figure -8 The results of the Ablation study, i.e., the incremental contributions of each architectural component.

6. Conclusion

This paper has introduced a new Dual-Branch Hybrid Network (DBHN) based on ResNet101 and DenseNet121 in the form of parallel features extraction to classify scenes of environmental catastrophes in the aerial images of UAVs. The proposed architecture takes advantage of representational complements between the high-level global semantic feature learning of residual learning and dense connectivity of fine-grained local feature aggregation to combine both branches representations through concatenation-based fusion module and a regularized classification head. Thorough experiments using a data set of 5,946 UAV images in the four types of disasters show that the DBHN has a total accuracy of 93.61 % and a macro F1-score of 93.43 %, outperforming the ResNet101 baseline by 3.99 percentage points and the DenseNet121 baseline by 3.47 percentage points in the accuracy.

As per-class analysis shows, the largest improvements are obtained in the case of the Earthquake/Urban Damage and Landslide classes, where F1-score changes are positive (+0.05 compared to both baselines) are the direct validation of the hypothesis of complementarity of the dual-branch design. The ablation experiment shows that all the architectural elements such as dual-branch fusion (+1.73%), deep classification head (+0.56%), and dropout regularization (+1.18%)

have independent and positive contributions to the overall performance. The DBHN also demonstrates competitive or higher accuracy than state-of-the-art methods on the UAV-based disaster scene classification, but it uses a four-category dataset, which is more comprehensive. These results confirm the suggested DBHN as a solid, efficient, and practically feasible problem solution to environment-related disaster scenes through the UAV that has a high likelihood of real-life emergency response and disaster management solutions. Further would involve extending the framework to bigger and more diverse data sets, creating computationally effective variants to deploy on-board UAVs and incorporating attention-based explain ability systems to improve operational transparency.

8. Recommendation and future work

Thw authors recommend the following as a future works:

1-E**x**plore knowledge distillation or lightweight backbone replacements (e.g., MobileNet or EfficientNet) to reduce the parameter count for better suitability in real-time UAV hardware.

2-Implement Grad-CAM or similar attention-based visualization tools to improve operational transparency, which is critical for safety-based emergency response deployments.

3-Test the framework on larger, more diverse datasets and utilize synthetic data generation (like GANs) to further address class imbalances, particularly for the Landslide category.

4-Future versions could integrate additional data sources, such as thermal imaging or climate-related data, to improve classification accuracy in complex environmental conditions.

Conflicts Of Interest

The authors declare no conflicts of interest.

References

[1] K. VanExel , Sherchan, S., Liu, S. "Optimizing deep learning models for climate-related natural disaster detection from UAV images and remote sensing data," Journal of Imaging, vol. 11, no. 2, p. 32, 2025. https://doi.org/10.3390/jimaging11020032

[2] L. P. Osco, José Marcato Junior, Ana Paula Marques Ramos, Lúcio André de Castro Jorge, Sarah Narges Fatholahi, Jonathan de Andrade Silva, Edson Takashi Matsubara, Hemerson Pistori, Wesley Nunes Gonçalves, Jonathan Li,, "A review on deep learning in UAV remote sensing," International Journal of Applied Earth Observation and Geoinformation, vol. 102, 2021, Art. no. 102456. https://doi.org/10.1016/j.jag.2021.102456

[3] Yuan, J., Ma, X., Han, G., Li, S., & Gong, W. (2022). Research on Lightweight Disaster Classification Based on High-Resolution Remote Sensing Images. Remote Sensing, 14(11), 2577. https://doi.org/10.3390/rs14112577

[4] T. Liu, T. Chen, R. Niu and A. Plaza, "Landslide detection mapping employing CNN, ResNet, and DenseNet in the Three Gorges Reservoir, China," IEEE Jornal Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 11417–11428, 2021. https://doi.org/10.1109/JSTARS.2021.3117975

[5] Guo, N., Jiang, M., Gao, L., Li, K., Zheng, F., Chen, X., & Wang, M. (2023). HFCC-Net: A Dual-Branch Hybrid Framework of CNN and CapsNet for Land-Use Scene Classification. Remote Sensing, vol. 15, no. 20, p. 5044, 2023. https://doi.org/10.3390/rs15205044

[6] C. Kyrkou and T. Theocharides, "EmergencyNet: Efficient aerial image classification for drone-based emergency monitoring using atrous convolutional feature fusion," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sensing, vol. 13, pp. 1687–1699, 2020. https://doi.org/10.1109/JSTARS.2020.2969809

[7] D. K. Kang, Michael J. Olsen, Erica Fischer, Jaehoon Jung and Julie A. Adams "Deep learning transferability across disaster types for UAS imagery based building damage assessment," Discover Civil Engineering, vol. 2, 2025, Art. no. 196. https://doi.org/10.1007/s44290-025-00357-y

[8] 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," International Journal of Advances in Signal and Image Sciences, 2026. https://doi.org/10.29284/0tapqn58

[9] Ghali R, Akhloufi MA, Mseddi WS. Deep Learning and Transformer Approaches for UAV-Based Wildfire Detection and Segmentation. Sensors (Basel). 2022 Mar 3;22(5):1977. doi: 10.3390/s22051977

[10] Jia, J., & Ye, W. (2023). Deep Learning for Earthquake Disaster Assessment: Objects, Data, Models, Stages, Challenges, and Opportunities. Remote Sensing, 15(16), 4098. https://doi.org/10.3390/rs15164098

[11] N. Alsaaran and Adel Soudani "Deep learning image-based classification for post-earthquake damage level prediction using UAVs," Sensors, vol. 25, no. 17, p. 5406, 2025. https://doi.org/10.3390/s25175406

[12] A. Al Mazroa, Nuha Alruwais, Muhammad Kashif Saeed , Kamal M Othman , Randa Allafi , Ahmed S Salama, "Multi class aerial image classification in UAV networks employing Snake Optimization Algorithm with deep learning," Scientific Repotrs, vol. 15, 2025, Art. no. 04570. https://doi.org/10.1038/s41598-025-04570-8

[13] K R, Prasanna Kumar, Mallegowda M, Manoj Kumar D P, Swathi H Y, and Ananda Babu J, "Improving UAV disaster response with DenseNet and AF-RCNN: A framework for accurate emergency spot detection," Cogent Engineering, vol. 13, no. 1, 2026, Art. no. 2612776. https://doi.org/10.1080/23311916.2026.2612776

[14] B. Sariturk and D. Z. Seker, "Comparison of residual and dense neural network approaches for building extraction from high-resolution aerial images," Advances in Space Research, vol. 71, no. 5, pp. 2150–2165, 2023. https://doi.org/10.1016/j.asr.2022.05.010

[15] M. Diykh, Mumtaz Ali, Mehdi Jamei, Shahab Abdulla, Md Palash Uddin, Aitazaz Ahsan Farooque, Abdulhaleem H. Labban, Hussein Alabdally, "Empirical curvelet transform based deep DenseNet model to predict NDVI using RGB drone imagery data," Computers and Electronics in Agriculture, vol. 221, 2024, Art. no. 109456. https://doi.org/10.1016/j.compag.2024.108964

[16] Thapa, A., Horanont, T., Neupane, B., & Aryal, J., "Deep learning for remote sensing image scene classification: A review and meta-analysis," Remote Sensing, vol. 15, no. 19, p. 4804, 2023. https://doi.org/10.3390/rs15194804

[17] S. D. Khan and Saleh Basalamah, "Multi-branch deep learning framework for land scene classification in satellite imagery," Remote Sensing, vol. 15, no. 13, p. 3408, 2023. https://doi.org/10.3390/rs15133408

[18] Zhao, J., Hu, L., Dong, Y., & Huang, L. (2021). Hybrid Dense Network with Dual Attention for Hyperspectral Image Classification. Remote. Sensing, 13, 4921. https://doi.org/10.3390/rs13234921

[19] Khondokar Radwanur Rahman, Amith Khandakar, Md. Fahim Hossen, Shaikh Golam Rabbani, Aniruddha Saha, Md. Faysal Ahamed, Md. Munawar Hossain, Mohamed Arselene Ayari, Rajab Abdulla R.E. Al-Esmail,"ReliefNet: A Knowledge-Driven, Explainable AI Multimodal Framework for Disaster Severity Classification and Humanitarian Decision-Making, Progress in Disaster Science, Volume 29 ,2026, 100528, https://doi.org/10.1016/j.pdisas.2026.100528

[20] Md. Mahfuzur Rahman, Sunzida Siddique, Marufa Kamal, Rakib Hossain Rifat, Kishor Datta Gupta, "UAV (Unmanned Aerial Vehicle): Diverse applications of UAV datasets in segmentation, classification, detection, and tracking," Algorithms, vol. 17, no. 12, p. 594, 2024. https://doi.org/10.3390/a17120594

[21] A. A. Adegun et al., "Review of deep learning methods for remote sensing satellite images classification: Experimental survey and comparative analysis," Journal of Big Data, vol. 10, 2023, Art. no. 772. https://doi.org/10.1186/s40537-023-00772-x

[22] A. Safonova, Gohar Ghazaryan, Stefan Stiller, Magdalena Main-Knorn, Claas Nendel, Masahiro Ryo,, "Ten deep learning techniques to address small data problems with remote sensing," International Journal of Applied Earth Observation and Geoinformation, vol. 125, 2023, Art. no. 103593. https://doi.org/10.1016/j.jag.2023.103569

[23] S. P. H. Boroujeni, Abolfazl Razi, Sahand Khoshdel, Fatemeh Afghah, Janice L. Coen, Leo O'Neill, Peter Fule, Adam Watts, Nick-Marios T. Kokolakis, Kyriakos G. Vamvoudakis, "A comprehensive survey of research towards AI-enabled unmanned aerial systems in pre-, active-, and post-wildfire management," Information Fusion, vol. 110, 2024, Art. no. 102447. https://doi.org/10.1016/j.inffus.2024.102369

[24] M. Ragab, Bandar M. Alghamdi, Sami Saeed Binyamin, Sultan Algarni, Roobaea Alroobaea, Abdullah M. Baqasah, Majed Alsafyani, "Multiclass aerial image recognition using improved Black Widow Optimization with deep learning on unmanned aerial networks imaging," Ain Shams Engineering Journal, vol. 16, no. 5, 2025, Art. no. 102437. https://doi.org/10.1016/j.asej.2025.103696

[25] Jiang, P., Ma, Z., and Mei, G." Review article: Deep learning for potential landslide identification: Data, models, applications, challenges, and opportunities," Natural Hazards and Earth System Sciences, vol. 26, pp. 487–520, 2026. https://doi.org/10.5194/nhess-26-487-2026

[26] I. Aydin, Emre Güçlü, Taha Kubilay Şener, Erhan Akin,, "Explainable flood damage assessment using multi-atrous self-attention mechanism," Results in Engineering., vol. 29, 2026, Art. no. 101008. https://doi.org/10.1016/j.aiig.2026.100192

[27] Xinjie Deng, Michael Shi, Burhan Khan, Yit Hong Choo, Fazal Ghaffar & Chee Peng Lim "A lightweight CNN model for UAV-based image classification," Soft Computing, 2025. https://doi.org/10.1007/s00500-025-10512-3