Gümüşhane University Journal of Science and Technology

GUFBD / *GUJS* (2026) 16(1): 267-282 doi: 10.17714/gumusfenbil.1800580 Research Article

Comparative analysis of deep feature fusion and machine learning classifiers for UAV imagery in post-earthquake building damage assessment

Deprem sonrası bina hasar tespitinde İHA görüntüleri için derin öznitelik füzyonu ve makine öğrenmesi sınıflandırıcılarının karşılaştırmalı analizi

Uğur ŞEVİK*[1], Aleyna YILMAZ[1]

[1]*Karadeniz Teknik Üniversitesi, Fen Fakültesi, Bilgisayar Bilimleri Bölümü, 61050, Trabzon*

• Received: 10.10.2025 • Accepted: 05.03.2026

Abstract

The rapid and accurate assessment of building damage after earthquakes is critically essential for search-and-rescue and humanitarian-aid operations. This study proposes a comprehensive hybrid intelligent system to classify buildings into three categories—intact, damaged, and destroyed—using the UAV-TEBDE dataset, which comprises high-resolution Unmanned Aerial Vehicle (UAV) images collected after earthquakes in Türkiye. The proposed methodology is based on the fusion of deep features extracted from five different pretrained Convolutional Neural Network (CNN) models, including ResNet50, EfficientNetB4, VGG16, DenseNet121, and MobileNetV2, using a transfer learning approach. These enriched, high-dimensional combined feature vectors were systematically used to compare the performance of 12 machine learning classifiers, including ensemble learning methods, support vector machines, and discriminant analyses. The experimental results, validated through a robust 10-fold Stratified Group Cross-Validation, demonstrated that the proposed feature-level (early) fusion strategy achieved outstanding success. The Quadratic Discriminant Analysis (QDA) model exhibited the highest performance, attaining a mean Weighted F1 Score of 99.53% (±0.09%), surpassing more complex ensemble models and neural networks. The exceptionally low standard deviation observed across the validation folds confirmed that the superior performance of the QDA model was statistically robust and consistent. This study revealed that CNN-based feature fusion yields a highly distinctive feature space for post-disaster damage assessment, thereby enabling rapid near-perfect automatic damage mapping.

Keywords: Deep learning, Earthquake damage assessment, Image fusion, Machine learning

Öz

Depremlerden sonra hızlı ve doğru bina hasar tespiti, arama-kurtarma ve insani yardım operasyonları için kritik öneme sahiptir. Bu çalışma, Türkiye'de meydana gelen depremlerden sonra toplanan yüksek çözünürlüklü İnsansız Hava Aracı (İHA) görüntülerinden oluşan özgün UAVs-TEBDE veri setini kullanarak, binaları sağlam, hasarlı ve yıkılmış olarak üç kategoride sınıflandırmak için yeni bir hibrit zeki sistem önermektedir. Önerilen metodoloji, transfer öğrenimi yaklaşımı kullanılarak ResNet50, EfficientNetB4 ve VGG16 dahil olmak üzere beş farklı önceden eğitilmiş Evrişimli Sinir Ağı (CNN) modelinden çıkarılan derin özniteliklerin birleştirilmesine (füzyonuna) dayanmaktadır. Bu zenginleştirilmiş ve yüksek boyutlu birleşik öznitelik vektörleri; topluluk öğrenmesi, destek vektör makineleri ve diskriminant analizi dahil olmak üzere 12 makine öğrenmesi sınıflandırıcısının performansını sistematik olarak karşılaştırmak için kullanılmıştır. Sağlam bir 10-katlı Katmanlı Grup Çapraz Doğrulama (Stratified Group Cross-Validation) ile doğrulanan deneysel sonuçlar, önerilen öznitelik füzyonu stratejisinin üstün bir başarı elde ettiğini göstermiştir. İkinci Dereceden Diskriminant Analizi (QDA) modeli, daha karmaşık topluluk modellerini ve sinir ağlarını geride bırakarak %99,53 (±0,09) ortalama Ağırlıklı F1 Skoru ile en yüksek performansı sergilemiştir. Doğrulama katmanları genelinde gözlemlenen son derece düşük standart sapma değerleri, QDA modelinin üstün performansının istatistiksel olarak kararlı ve tutarlı olduğunu doğrulamıştır. Bu çalışma, CNN tabanlı öznitelik füzyonunun afet sonrası hasar tespiti için son derece ayırt edici bir öznitelik uzayı oluşturduğunu, böylece hızlı ve neredeyse mükemmel doğrulukta otomatik hasar haritalarının üretilmesini mümkün kıldığını ortaya koymaktadır.

Anahtar kelimeler: *Derin öğrenme, Deprem hasar değerlendirmesi, Görüntü birleştirme, Makine öğrenimi*

*Uğur ŞEVİK; usevik@ktu.edu.tr

1. Introduction

Earthquakes are among the most devastating natural disasters, causing sudden and massive destruction in cities and leaving deep socioeconomic scars on societies. Severe seismic events in densely populated urban areas cause significant damage to buildings and infrastructure, resulting in substantial loss of life, economic collapse, and mass displacement (Yu et al., 2024). The Kahramanmaraş-centered earthquakes with magnitudes of 7.8 and 7.5 that struck southeastern Türkiye on February 6, 2023, and caused massive destruction in 11 provinces are the most recent and striking example of this tragic reality. In this disaster, hundreds of thousands of buildings were severely damaged or destroyed, resulting in tens of thousands of deaths and creating an urgent need for shelter for millions of people (Xia et al., 2023). After such a large-scale catastrophe, it is vital to quickly and accurately determine the locations, numbers, and damage levels of affected buildings to effectively coordinate search and rescue operations, ensure the proper and prioritized allocation of emergency resources (medical aid, heavy machinery, personnel), and plan humanitarian aid logistics (Dell'Acqua & Gamba, 2012).

Although the currently available damage assessment methods are widely accepted as the primary basis for detailed structural damage analysis, they have significant limitations in disasters affecting large geographic areas. Field studies are slow and costly, require intensive manpower, and are hazardous owing to safety risks in debris zones (e.g., aftershocks and dangerous materials) (Korkmaz, 2009). Roads closed and infrastructure damaged after a disaster prevent field teams from reaching the most urgently needed areas promptly, delaying the creation of a comprehensive damage assessment that decision-makers require during the critical first 72-hour "golden window." Therefore, Remote Sensing (RS) technologies have become indispensable tools in disaster management for overcoming these challenges (Dong & Shan, 2013).

In recent years, Unmanned Aerial Vehicles (UAVs) have demonstrated revolutionary potential in post-disaster damage assessment owing to their flexibility, low cost, and ability to collect centimeter-level data on demand (Kerle et al., 2019). Unlike satellites, which are constrained by fixed orbits and revisit intervals, UAVs can collect data at desired times and angles without interference from atmospheric obstacles such as cloud cover. Thus, UAVs enable the acquisition of oblique images that capture not only roofs but also building facades, facilitating the analysis of details such as partial collapses or wall damage that are difficult to detect in satellite imagery (Nex et al., 2019).

The increasing volume and complexity of UAV data have made automation essential for extracting meaningful information from these data sets. Breakthroughs in Artificial Intelligence (AI), particularly Deep Learning (DL) and Machine Learning (ML) algorithms, offer robust solutions for automatic damage detection from UAV images. DL models, such as Convolutional Neural Networks (CNNs), demonstrate superior performance compared to traditional methods owing to their ability to learn hierarchical and distinctive features (deep features) from images without human intervention (Zhu et al., 2017). Although various approaches, such as CNN-based classification (Ji et al., 2018), object detection (Li et al., 2022), and image segmentation (Zhan et al., 2022), have been successfully applied in the literature for earthquake damage detection, most existing studies focus on the binary classification problem (damaged/undamaged) or evaluate the performance of a single DL architecture. However, in disaster management, damage grading is more critical and functional for determining intervention priorities. Therefore, a multi-class approach that categorizes damage into three or more classes–intact, damaged, and collapsed–provides richer and more actionable information (Gürer & Karslıgil, 2024). In this context, this study aims to fill a significant gap in the literature by systematically comparing the performance of deep features extracted from different DL architectures (ResNet, EfficientNet, and VGG16) on a three-class damage detection problem across a diverse set of 12 ML classifiers.

The main hypothesis of this study is that combining features extracted from different CNN models (feature fusion) and classifying the resulting enriched feature set, particularly with powerful ensemble learning models such as XGBoost and LightGBM, will yield higher classification accuracy and generalization performance than using a single model. To achieve this goal, a unique dataset, UAVs-TEBDE, comprising very high-resolution UAV images acquired after earthquakes in Türkiye was employed. Within the scope of this study, feature extraction was performed using five pretrained CNN models, and the extracted features were evaluated using 12 ML classifiers. The findings are expected to make significant scientific contributions to the development of more accurate, faster, and more reliable automated systems for post-disaster damage assessment and strengthen operational decision-support mechanisms for emergency response teams.

While recent advancements in Deep Learning have introduced transformer-based architectures (e.g., Vision Transformers) and multimodal approaches that leverage the global context, Convolutional Neural Networks (CNNs) remain exceptionally strong in capturing local textural details. Identifying structural damage often relies on the detection of specific local patterns, such as rubble piles and cracks. Therefore, this study focuses on maximizing the potential of established CNN architectures through a robust feature-level (early) fusion strategy that balances high accuracy and computational feasibility.

2. Related works

Over the past two decades, post-earthquake building damage assessment has undergone a significant evolution within the fields of remote sensing and computer vision. Rapid technological advancements have fundamentally reshaped the entire pipeline, spanning data acquisition platforms, analytical algorithms, and image interpretation paradigms. This section positions the current study within the broader literature by examining established data sources, analytical frameworks, and the state-of-the-art in machine learning and deep learning approaches for damage evaluation. Furthermore, Table 1 outlines the methodologies, study objectives, and key findings of prominent studies in this domain.

Table 1. Related works

Author/Year	Study Objective	Methods	Key Results / Conclusions
Tong et al. (2012)	Building damage detection through 3D geometric changes using satellite stereo images	HRSI, DEM difference, RPC, SGM matching	1.1 m in planimetry, higher accuracy at 1.5 m elevation; 90%+ overall accuracy; 100% object-based accuracy (at certain thresholds)
Du et al. (2024)	Single building damage detection based on multi-feature fusion	UAV + satellite imagery, nDSM, ROI, PCA, feature fusion	93.4% OA, 0.87 Kappa (UAV model); 89% OA in the portability test
Azizi (2023)	Multi-label building damage classification using machine learning methods for UAV imagery	CLIP model, RescueNet, image enhancement	YOLOv8 (Top-5) achieved the highest accuracy (98%). CLIP (92%) outperformed YOLOv8 (Top-1), EfficientNet, and MobileNetV2.
Ghahrloo & Mokhtarzade (2025)	Damage detection based on smart systems using optical and radar data	CNN, SVM, fuzzy inference, pixel/decision-level fusion	The object-based CNN achieved 91.7% overall accuracy (OA) and 87.4% kappa coefficient (KC), whereas the pixel-based approach yielded 87.4% OA and 81.0% KC. Augmenting a CNN with an SVM yielded no additional improvement. The object-based CNN+FIS configuration delivered the highest performance, attaining 92.7% OA and 88.7% KC.
Matsuoka & Yamazaki (2004)	Detection of building damage after an earthquake using SAR intensity images	ERS SAR, speckle reduction (Lee, Frost), discriminant analysis	78.5% discrimination random rate; optimum window size 21×21 pixels
Park & Jung (2020)	Damage assessment with Polarimetric SAR (POLSAR)	PALSAR-2, polarimetric scattering variations, fuzzy-based fusion	90.9% detection rate, 1.3% false alarm, 0.813 Kappa, 69.7% FOM
Saito et al. (2005)	Visual damage assessment of the earthquake using QuickBird images	100×100 m grid visual interpretation and comparison	53.5% consistency between comments; objective accuracy measurement could not be performed
This Study (2026)	Three-class building damage classification (Intact, Damaged, Destroyed)	Fusion of 5 CNN Features + Regularized QDA (10-Fold Group CV)	99.53% Mean Weighted F1-Score; Statistically robust validation.

As shown in the final row of Table 1, the proposed method outperforms existing approaches by employing a comprehensive feature-level (early) fusion strategy and rigorous 10-fold group cross-validation, achieving a mean F1-score of 99.53%.

2.1. Remote sensing data sources for damage assessment

The UAV data used in damage assessment are generally obtained from satellites, airplanes, and UAVs. Initial studies were based on medium-resolution satellite imagery, such as Landsat imagery. Using these data, it was only possible to detect general damage conditions at the regional scale and at the level of city blocks (Yusuf et al., 2001). Since the early 2000s, the commercial availability of very high-resolution (VHR) optical satellites such as QuickBird, IKONOS, WorldView, and Pleiades has enabled analyses at the scale of individual buildings (Dong & Shan, 2013). Although VHR optical images provide comprehensive visual content for detecting damage indicators such as roof integrity loss, debris piles, and significant structural deformations, they are adversely affected by atmospheric conditions such as cloud cover, fog, and insufficient illumination, which are frequently encountered in the aftermath of disasters (Yamazaki et al., 2005). To overcome this limitation, Synthetic Aperture Radar (SAR) systems offer a strong alternative, as they can collect data day and night and under all weather conditions, owing to active microwave sensors (Ge et al., 2020). SAR-based damage assessment is typically based on coherence analysis, which measures the degradation of phase information between pre- and post-earthquake image pairs or by examining changes in the intensity of the backscatter signal (Yun et al., 2015; Rao et al., 2023). VHR SAR satellites, such as TerraSAR-X, COSMO-SkyMed, and ALOS-2, have enhanced the potential of this field by enabling more detailed analyses of structural changes at the building scale (Brunner et al. 2010; Kim et al. 2023).

In the past decade, UAVs made available for civilian and commercial use have been able to collect data not only from vertical perspectives but also from oblique angles that reveal damage to building facades (Vetrivel et al., 2018). This multi-angle imaging capability allows for the identification of critical types of damage, such as soft-story collapse or deep cracks on facades, which are nearly impossible to detect from satellite imagery. Additionally, Dense Point Clouds obtained through UAV photogrammetry and the derived Digital Surface Models (DSM) were used to analyze changes in the three-dimensional structures of buildings. In particular, the analysis of normalized DSMs (nDSM), obtained from the differences between pre- and post-earthquake DSMs, is a widely used and effective method for quantitatively measuring the number of building collapses and accurately detecting specific types of damage, such as "pancake collapse" (Menderes et al., 2015; Erdogan & Yilmaz, 2019). In this study, UAVs were chosen as the data source to enable more precise classification by leveraging the detailed observations offered by this platform.

2.2. Machine learning and deep learning approaches in damage assessment

Automatic damage detection from UAV data has evolved from traditional image processing techniques to shallow learning (SL) and deep learning (DL) approaches. In conventional SL methods, distinctive features are usually extracted from images manually or semi-automatically. These features include Gray-Level Co-occurrence Matrix (GLCM)-based texture features (Haralick et al., 1973), shape indices (e.g., roundness and elongation), and spectral indices (e.g., NDVI). The extracted feature vectors are then used to train classifiers, such as Support Vector Machines (SVMs), random forests (RFs), or Artificial Neural Networks (ANNs) (Takhtkeshha et al., 2023; Kaya, 2024). Although these methods can be effective, their success largely depends on the quality of feature engineering, and finding the most suitable set of features for each geography or sensor is a laborious process.

DL, especially CNNs, has significantly transformed the field by overcoming the limitations of feature engineering. CNNs learn hierarchical feature representations, ranging from raw image pixels to simple features, such as edges and corners, and complex object parts, such as building roofs or piles of debris (LeCun et al., 2015). This ability to learn features automatically and in a data-driven manner renders CNNs highly effective for damage assessment applications. In the literature, CNN-based models have been applied within the framework of three main tasks:

Image classification: In this approach, small image patches centered on each building are assigned to classes such as “damaged,” “slightly damaged,” or “intact.” This method is particularly advantageous for labeling large datasets and fast model training. Xu et al. (2019) demonstrated the potential of this approach by

comparing different CNN architectures for classifying damaged buildings in satellite images after the Haiti earthquake.

Object detection: This method involves enclosing damaged buildings in an image within a bounding box to determine their locations and classify them accordingly. One-stage detectors, such as You Only Look Once (YOLO) and Single Shot MultiBox Detector (SSD), are preferred for rapid post-disaster analyses because of their speed and efficiency (Li et al., 2022).

Image segmentation: This approach provides the most detailed spatial information regarding the images. In semantic segmentation, each pixel in an image is assigned to a class (e.g., building, debris, or road). In contrast, in instance segmentation, different objects of the same class (e.g., each damaged building) are distinguished from one another. Mask R-CNN is one of the most well-known and successful architectures in the field of instance segmentation (He et al., 2017). Zhan et al. (2022) demonstrated the effectiveness of this method using a modified Mask R-CNN model to detect damaged buildings in aerial photographs after the Kumamoto Earthquake. Similarly, studies conducted after earthquakes in Türkiye have successfully utilized the Mask R-CNN and other segmentation architectures (Gürer & Karslıgil, 2024; Kaya, 2025).

2.3. Two-Stage approach: deep feature extraction and machine learning classification

While end-to-end deep learning models demonstrate remarkable success, a two-stage hybrid approach presents a highly effective alternative, particularly in scenarios where labeled data is limited. In the primary stage of this methodology, a Convolutional Neural Network (CNN) pre-trained on a massive dataset, such as ImageNet, functions as a robust feature extractor. Through transfer learning, the terminal classification layer of the model is discarded, enabling the extraction of a high-level, dense feature vector from the penultimate layers for each image patch. This mechanism allows deep learning models to transfer generalized visual representations to a domain-specific application—in this context, post-earthquake damage detection (Kaur et al., 2022).

In the subsequent stage, these extracted deep feature vectors serve as inputs for conventional machine learning classifiers. This hybrid methodology synergistically couples the profound representation learning capabilities of deep neural networks with the robust generalization attributes of traditional classifiers on smaller datasets. Consequently, it establishes a highly flexible framework for systematically comparing the performance of diverse machine learning algorithms.

2.4. Research gap and contribution of this study

A critical review of the existing literature reveals several notable research gaps in post-disaster building damage assessment. Primarily, the majority of studies rely on a single deep learning architecture, lacking a systematic comparison of feature extraction capabilities across diverse models. Furthermore, research utilizing machine learning classifiers trained on deep features frequently limits its scope to a few conventional algorithms, such as Support Vector Machines (SVM) or Random Forests (RF). To date, a comprehensive evaluation encompassing a broad spectrum of tree-based, kernel-based, linear, probabilistic, and ensemble classifiers remains absent. Consequently, performing an exhaustive comparative analysis on a challenging multi-class problem (intact, damaged, and destroyed) using a unique dataset of ultra-high-resolution UAV imagery represents a substantial contribution to the domain.

This study directly addresses these limitations. By fusing features extracted from five CNN models and evaluating this high-dimensional representation across 12 distinct classifiers, we seek to establish the most robust and accurate methodological pipeline for earthquake damage detection. Ultimately, this systematic framework offers a highly reliable foundation for both future academic research and real-world operational deployments.

3. Materials and methods

This study proposes a two-stage hybrid methodology that systematically compares deep feature fusion and ML classifiers for detecting building damage in three classes (intact, damaged, and destroyed) from post-earthquake UAV images. The workflow includes data preparation, deep feature extraction, ML classification, and performance evaluation. All analyses were designed with reproducibility and efficiency in mind.

3.1. Study area and dataset

The dataset utilized in this study, designated as UAVs-TEBDE, comprises ultra-high-resolution imagery acquired via UAVs following severe seismic events in Türkiye. This unique dataset captures a wide array of structural perspectives within the affected urban landscapes. To facilitate a robust classification framework, the data was categorized into three primary damage levels: intact, damaged, and collapsed. To ensure class balance and bolster the generalization capabilities of the models, diverse data augmentation techniques—including spatial rotation, flipping, and brightness adjustments—were systematically applied to the original images. The final dataset utilized for the analysis comprised 18,136 images, with the detailed class distribution outlined in Table 2. Each image was meticulously cropped to isolate individual buildings or prominent structural segments, establishing these localized patches as the fundamental input for the proposed methodology.

The UAVs-TEBDE dataset encompasses images captured under highly variable and uncontrolled illumination (e.g., sunny, overcast, and shadowed) and across varying flight altitudes, accurately reflecting the chaotic environmental conditions typical of rapid disaster response. This inherent heterogeneity is crucial for mitigating model overfitting to specific lighting or scale conditions. Furthermore, the dataset spans a broad spectrum of urban densities—ranging from detached masonry structures to high-density reinforced concrete complexes—ensuring that the algorithms learn the fundamental visual semantics of structural failure rather than merely memorizing specific architectural typologies.

Table 2. Class distribution of the UAVs-TEBDE dataset

Damage Class	Number of Images	Class Percentage (%)
Intact	5,998	33.07
Damaged	6,164	33.99
Collapsed	5,974	32.94
Total	18,136	100.00

3.2. Experimental setup

All experiments were conducted on the Google Colab Pro platform using a high-performance NVIDIA L4 Tensor Core Graphics Processing Unit (GPU). The hardware infrastructure of the system was supported by 52.96 GB of system memory (RAM) and an x86_64 processor. For the software environment, Python 3.11 was used, along with TensorFlow (v2.19.0) and Keras for deep learning models, and commonly used libraries such as Scikit-learn (v1.5.1), XGBoost, LightGBM, and CatBoost for machine learning classification and metric computation. This configuration ensured the efficient execution of computationally intensive tasks and intense feature extraction.

3.3. Proposed methodological framework

A flowchart of the proposed two-stage framework is shown in Figure 1. The first stage involved extracting deep features from five pretrained CNN models using transfer learning and fusing them. In the second stage, these combined feature vectors were used to train and test 12 different machine learning classifiers.

3.3.1. Deep feature extraction and fusion

The purpose of this stage is to obtain high-level, semantically rich representations (feature vectors) from raw image pixels that are effective in distinguishing the damage status of buildings.

Feature extraction with transfer learning: Pre-trained CNN models trained on large-scale datasets such as ImageNet possess an exceptional ability to recognize general visual patterns. In this study, five state-of-the-art CNN models with different architectures, depths, and complexities were used as feature extractors: ResNet50, EfficientNetB4, VGG16, DenseNet121, and MobileNetV2. These architectures were selected to represent diverse structural philosophies in deep learning: ResNet50 for residual learning, DenseNet121 for feature reuse, VGG16 for its deep yet simple architecture, EfficientNetB4 for compound scaling efficiency, and MobileNetV2 for lightweight, depthwise separable convolutions.

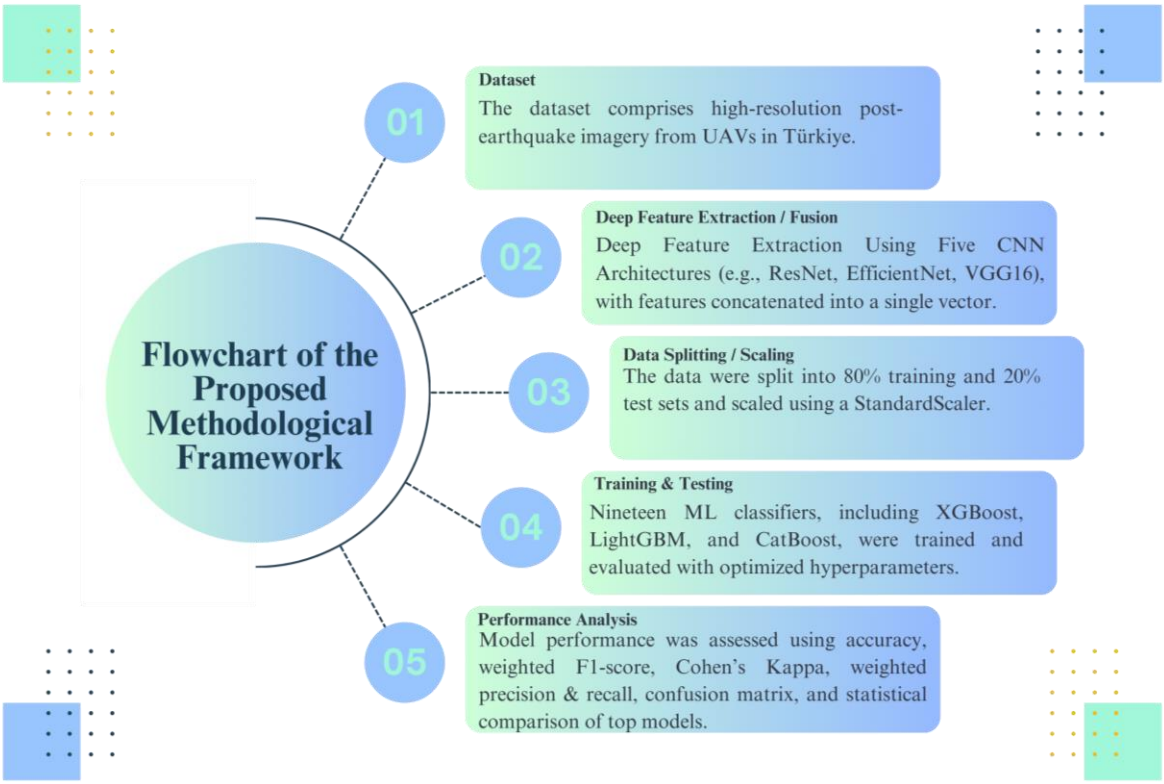

Figure 1. Flowchart of the proposed methodological framework

This diversity ensures that the fused feature vector captures a comprehensive range of visual patterns from low-level textures to high-level semantic abstractions. The primary purpose of selecting these models was to create a more robust feature set by combining complementary information from different architectures. For each model, the final classification layer (fully connected layer) was removed and replaced with a GlobalAveragePooling2D layer. Thus, each model produces a fixed-size 1D feature vector for the building image. During this phase, the weights of all pretrained CNN models were kept completely frozen, and no fine-tuning was applied. This approach ensured that the extracted features captured the general visual patterns learned from ImageNet, thereby preventing the models from overfitting to specific building instances in the dataset.

Feature fusion: For the image, features from the five models were combined into a single, long, comprehensive feature vector. This "early fusion" strategy aims to provide the classifier with richer information by combining features learned by different models, such as the deep features learned by ResNet via residual connections, feature reuse in DenseNet, and scale efficiency of EfficientNet, into a single representation. Consequently, a unified feature vector with 6,656 dimensions was obtained for each image.

Caching for efficiency: The deep feature extraction process was run once for the entire dataset, and the extracted feature vectors, along with their labels and file paths, were saved in a pickle file. This caching mechanism significantly accelerated the analysis process by allowing the skipping of steps that could take hours in subsequent experiments.

3.3.2. Classification Strategy

At this stage, the effectiveness of the obtained combined feature set was tested using a wide range of ML classifiers.

Data preparation and Validation Strategy: To rigorously evaluate the models and prevent data leakage, we employed a 10-fold stratified group cross-validation strategy. Unlike standard random splitting, this method ensures that all images derived from the same building (original and augmented versions) are maintained in the same fold, either entirely in the training set or entirely in the test set. This prevents the model from "memorizing" specific building structures seen during training. In each fold, the data were split, and a

StandardScaler was fitted only to the training subset to scale the test subset, thereby preventing any statistical leakage.

Comparative analysis of classifiers: A total of 12 ML classifiers were trained and evaluated using optimized hyperparameters. These classifiers were selected from Ensemble Models, Kernel-Based Models, Neural Network-Based Models, Linear Models, and other machine learning methods to represent different learning methodologies. To ensure reproducibility, standard hyperparameters were employed with minor optimization. For instance, the **SGD Classifier** was configured with 'hinge' loss to function as a linear SVM, providing a computationally efficient alternative to the standard SVM-RBF, which was excluded because of its high complexity $\left(O(n^2.d)\right)$ in the 6,656-dimensional feature space. The Random Forest and Extra Trees models utilized 100 estimators, whereas the QDA model was regularized (reg_param=0.5) to ensure stability.

Model training and timeout control: Training for each model was automatically terminated if it exceeded the specified TIMEOUT_SECONDS (900s). This prevented the analysis process from being prolonged by models that can take a very long time to train, particularly on large, high-dimensional datasets, such as Gradient Boosting. All trained models were saved in a pickle file for further analysis.

3.4. Performance evaluation and statistical analysis

The performance of the models was evaluated using a comprehensive set of metrics on the test dataset. For this problem, which is not imbalanced but multi-class, the most suitable metrics were selected as the Overall Accuracy (Sokolova & Lapalme, 2009), Weighted F1-Score (Chicco & Jurman, 2020), Cohen's Kappa Coefficient (Cohen, 1960), Weighted Precision, and Weighted Recall. In addition, the confusion matrix visually demonstrated which classes were misclassified, and through failure analysis, misclassifications in the best-performing model were assessed to identify areas for improvement. The statistical stability of the models was evaluated by analyzing the standard deviation of the performance metrics obtained from the 10-fold cross-validation.

4. Results

In this section, the experimental results obtained using the proposed method are presented. The results include a comparative performance analysis of the classifiers, a detailed examination of the best model, an investigation of the cases in which the model failed, and a statistical stability analysis of the best model.

4.1. General performance comparison of classifiers

Fused deep feature vectors derived from five distinct pretrained CNN architectures were used to train 12 machine-learning classifiers. Table 3 presents the performance metrics of the successfully trained models, excluding those that exceeded the computational time limit. To facilitate comparison, the classifiers were ranked in descending order by the Weighted F1-Score, which served as the primary evaluation metric.

Table 3. Comparative performance metrics (Mean ± Std) and training times of the classifiers over 10-fold cross-validation.

Model	Accuracy (Mean)	Accuracy (Std)	F1 Score (Mean)	F1 Score (Std)	Kappa (Mean)	Kappa (Std)	Time(s) (Mean)	Time(s) (Std)
QDA	0.9953	0.0009	0.9953	0.0009	0.9929	0.0014	166.61	3.45
MLP	0.9942	0.0019	0.9942	0.0019	0.9912	0.0029	49.16	15.70
KNN	0.9917	0.0027	0.9917	0.0027	0.9876	0.0040	3.66	0.35
LDA	0.9913	0.0017	0.9913	0.0017	0.9869	0.0025	64.56	0.65
Ridge	0.9880	0.0020	0.9880	0.0020	0.9820	0.0031	4.46	0.34
XGBoost	0.9814	0.0040	0.9815	0.0040	0.9721	0.0061	47.55	1.15
CatBoost	0.9804	0.0036	0.9804	0.0036	0.9706	0.0054	105.65	0.43
LightGBM	0.9779	0.0045	0.9780	0.0045	0.9669	0.0068	129.42	1.39
Linear SVM	0.9747	0.0034	0.9746	0.0034	0.9620	0.0050	13.58	2.90
Extra Trees	0.9473	0.0066	0.9474	0.0066	0.9209	0.0099	6.71	0.69
Random Forest	0.9440	0.0069	0.9441	0.0068	0.9159	0.0104	36.38	0.78
Naive Bayes	0.8392	0.0099	0.8402	0.0099	0.7586	0.0149	0.89	0.06

As shown in Table 3, the proposed deep feature fusion approach is highly effective, with most classifiers achieving mean accuracies of over 94%. The QDA model exhibited the best performance, achieving nearly perfect results with a mean Accuracy and F1-score of 0.9953 and a very low standard deviation (±0.0009), indicating high stability. It was closely followed by the MLP Classifier (0.9942 F1-Score) and K-Nearest Neighbors (0.9917 F1-Score). Furthermore, the analysis of Cohen's Kappa coefficient strongly corroborates these findings. The QDA model achieved an exceptional mean Kappa value of 0.9929 (±0.0014), indicating an almost perfect agreement between the predicted damage classes and the true ground-truth labels, well beyond what would be expected by chance. This outstanding Kappa score is highly consistent with the observed Overall Accuracy and Weighted F1-Scores, further validating the reliability and robustness of the proposed feature fusion methodology across all three damage categories.

Powerful gradient boosting algorithms, namely XGBoost, CatBoost, and LightGBM, also demonstrated strong performance in the 97-98% range but fell slightly behind the QDA and MLP models. Traditional ensemble models, such as Extra Trees and Random Forest, maintained performance levels of approximately 94%. The Naïve Bayes model exhibited the lowest performance.

Training time is a critical factor regarding operational feasibility, as shown in Table 3. Although ensemble models such as XGBoost achieved high accuracy, they required competitive training times (47.55s). The QDA model required a longer training time (166.61s) owing to the computation of large covariance matrices in the high-dimensional space. However, this incurs a one-time computational cost during training. In post-disaster scenarios, where model inference speed is paramount, QDA is highly advantageous because its prediction phase is fast and computationally efficient once the parameters have been estimated.

Figure 2 presents a comparative visual representation of the key performance metrics (Accuracy, F1 Score, Kappa) for the 12 classifiers. The chart clearly demonstrates that QDA, MLP, and KNN stand out from the other models and exhibit the highest performance.

4.2. In-Depth analysis of the best model: quadratic discriminant analysis (QDA)

The results of the best-performing QDA model were subjected to detailed analysis. The model performance by class is shown in Table 4 (Classification Report) and Figure 3 (Confusion Matrix).

Table 4. Detailed classification report presenting class-wise Precision, Recall, F1-Score, and Support for the QDA model (Cumulative 10-Fold CV)

Classes	Precision	Recall	F1-score	Support
Intact	1.00	0.99	1.00	**5,998**
Damaged	0.99	1.00	0.99	**6,164**
Collapsed	1.00	0.99	1.00	**5,974**
Accuracy			**1.00**	**18,136**
Macro Avg	1.00	1.00	1.00	**18,136**
Weighted Avg	1.00	1.00	1.00	**18,136**

The classification report indicated that the QDA model achieved outstanding performance across all three classes. For the "Intact" class, both the precision and recall values were 1.00, indicating that no errors were made in this class. For the "Damaged" and "Destroyed" classes, the metrics were also above 99%. This demonstrates that the model can distinguish among the levels of damage with a very high accuracy.

As shown in Figure 3, the confusion matrix visually confirms these results. It is important to note that this matrix represents the cumulative results of 10-fold stratified group cross-validation, aggregating the predictions for every image in the dataset (total: 18,136) used as test data. The high values along the diagonal (5,957, 6,149, and 5,944) indicate consistent and correct classification across the entire dataset. The very low values outside the diagonal indicate that the model made minimal errors. For example, out of 6,164 'Damaged' buildings, only 13 were misclassified as 'Collapsed' and two as 'Intact.' Overall, across the entire dataset of 18,136 samples, only 86 were misclassified, underscoring the robustness and reliability of the model.

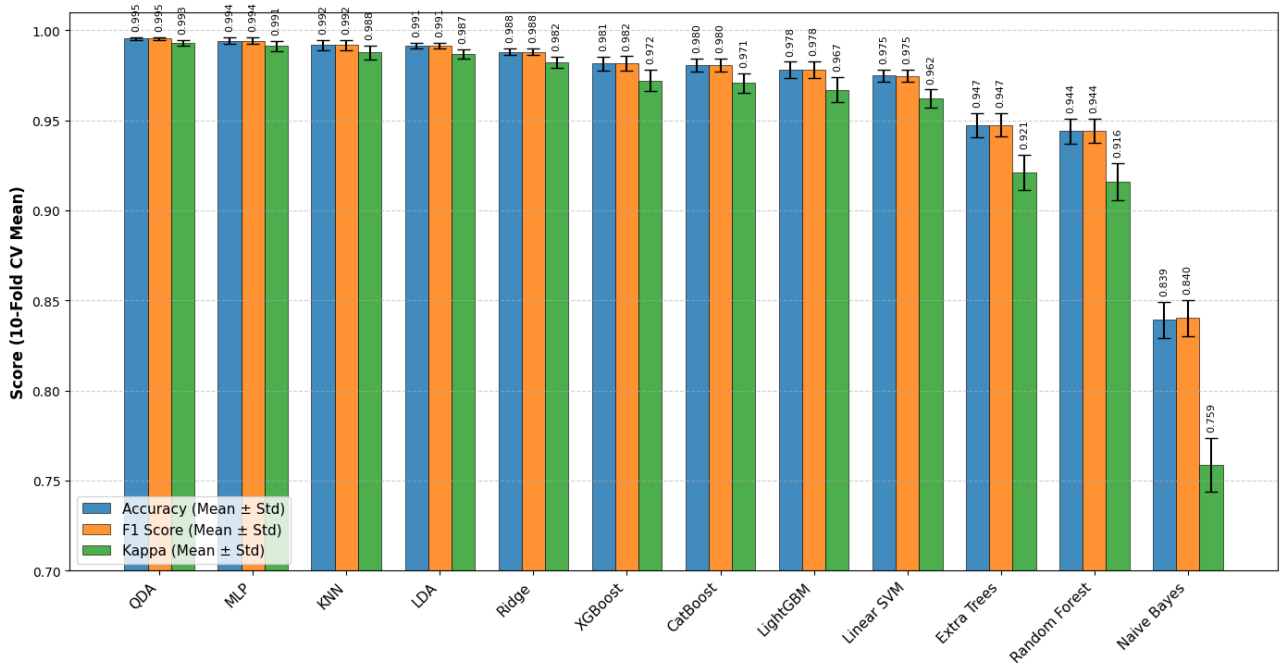

Figure 2. Comparative performance of 12 classifiers based on 10-Fold Stratified Group Cross-Validation. The bar heights represent the mean Accuracy, Weighted F1-Score, and Kappa values, whereas the error bars indicate the standard deviation across the folds.

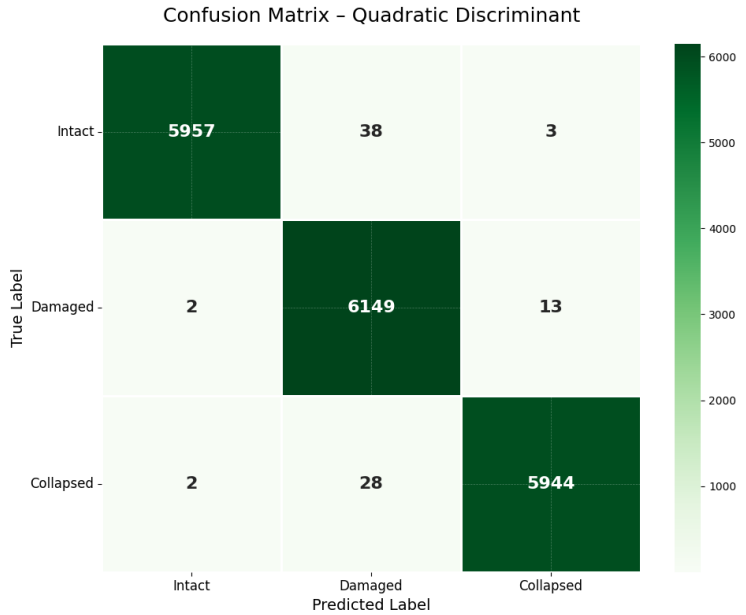

Figure 3. Cumulative confusion matrix of the QDA model obtained via 10-Fold Stratified Group Cross-Validation across the entire dataset (18,136 samples)

4.3. Failure analysis

To systematically investigate the rare instances of misclassification, an in-depth visual assessment of the failure cases was conducted (Figure 4). This analysis revealed that prediction errors predominantly occurred in scenarios characterized by high visual ambiguity or subtle structural transitions between damage categories. For instance, a structure exhibiting partial roof collapse while retaining intact load-bearing walls might be erroneously predicted as "Destroyed" rather than "Damaged." Conversely, a severely compromised building on the verge of total collapse might be mislabeled as "Damaged." These borderline cases underscore the inherent continuum of structural degradation, which poses a significant challenge for discrete classification algorithms. While 'easy' samples—such as completely flattened debris—are generally detected with near-absolute certainty across the dataset, Figure 4 specifically highlights the 'hard' samples (borderline cases).

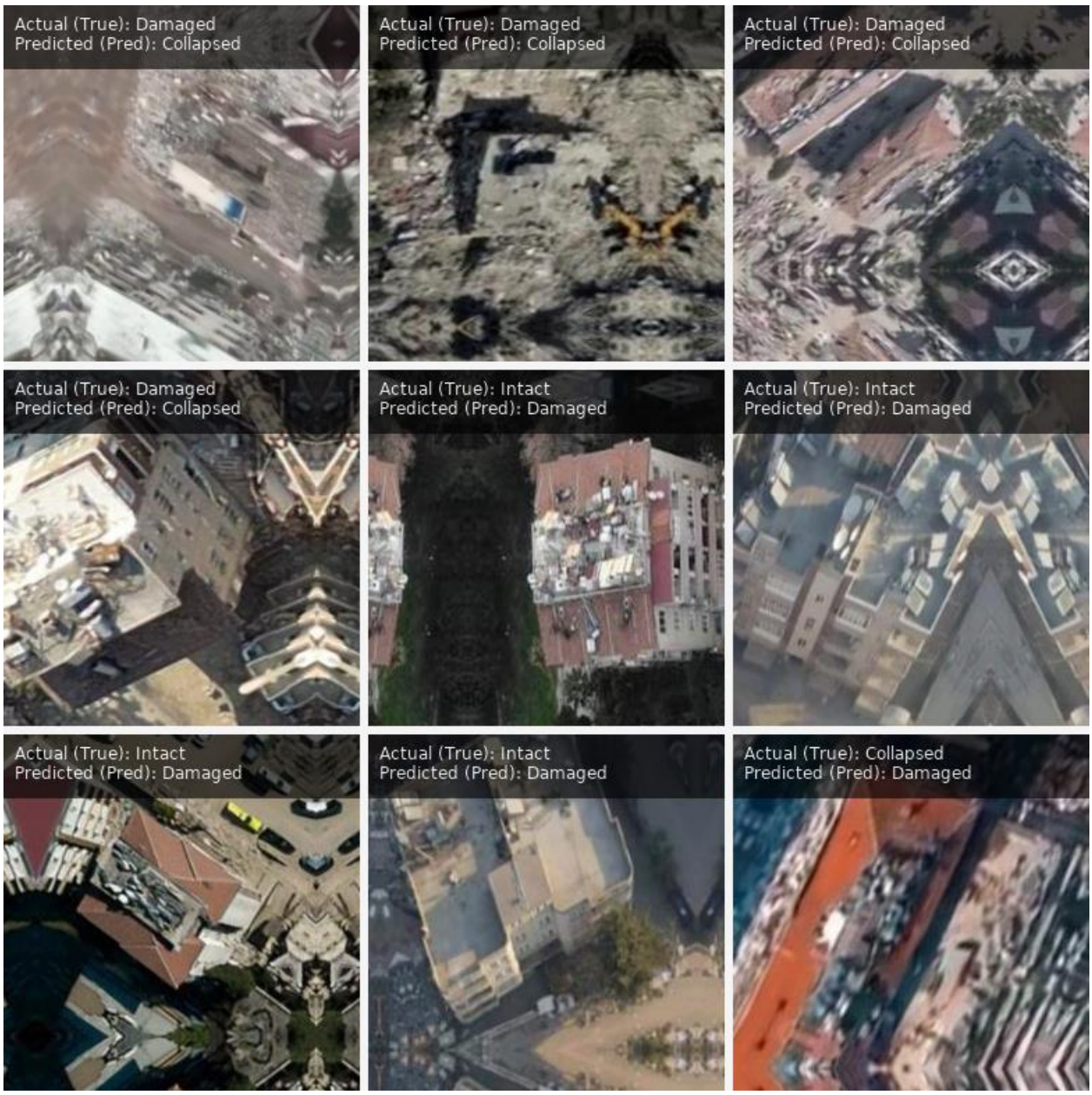

Figure 4. Examples of 'hard' samples (borderline cases) misclassified by the QDA model. The figure demonstrates the visual ambiguity in transitioning damage states (e.g., True: Damaged vs. Predicted: Collapsed), where the model struggles to distinctly separate the classes.

As illustrated, algorithmic confusion typically occurs in highly ambiguous scenarios, such as pancake collapses obscured by intact roofs or partially damaged structures exhibiting characteristics of multiple damage classes.

4.4. Statistical validation and stability analysis

Because 10-fold cross-validation yields a distribution of performance metrics, we assessed model stability using the standard deviation of the F1 Scores. The QDA model not only achieved the highest mean score but also exhibited the lowest standard deviation (0.0009) among the top-tier classifiers, compared with MLP (0.0019) and XGBoost (0.0040). This extremely low variance confirms that the superior performance ofQDA is statistically robust and consistent across different subsets of the data, rather than being coincidental to a specific train-test split.

5. Discussion

The results obtained in this study demonstrate the effectiveness of the proposed two-stage hybrid approach for detecting building damage in UAV images captured after earthquakes. The synergistic fusion of features

extracted from five CNN models, followed by rigorous evaluation across a diverse array of machine learning classifiers, establishes a robust methodological framework that demonstrates exceptionally high performance relative to the existing literature.

The most striking finding of this study is that the QDA model trained on combined deep features slightly outperformed the MLP but significantly surpassed the ensemble models. A mean accuracy rate of 99.53% is an outstanding achievement for complex classification problems. There may be several possible explanations for this. First, the 6,656-dimensional feature space constructed by combining features from five different CNN models may exhibit a highly discriminative structure between the classes. QDA assumes that the data for each class are Gaussian-distributed and determines the decision boundaries between classes using quadratic functions. QDA is expected to be successful if the classes form quadratically separable clusters in the combined feature space. In other words, feature fusion may have made the problem so "easy" for classical classifiers that even a model based on a simpler assumption, such as QDA, could achieve near-perfect results.

The exceptional performance of QDA in such a high-dimensional feature space (d=6656) warrants a theoretical explanation, as classical QDA typically struggles when the number of features exceeds the number of samples (d>n) owing to the singularity of the covariance matrix. In this study, we employed a Regularized QDA. By introducing a regularization parameter to the covariance matrix estimation, we ensured the numerical stability. This regularization enables the model to leverage the high separability of the fused deep features by modeling class distributions as distinct Gaussians, thereby mitigating the 'curse of dimensionality' without overfitting.

Despite the high dimensionality of the fused feature vector (6,656 dimensions), we deliberately avoided dimensionality-reduction techniques such as PCA in this phase. The results from the 10-Fold CV demonstrate that discriminative models, particularly QDA, could effectively handle this high-dimensional space without overfitting, as evidenced by the consistent performance across folds. This suggests that the feature space induced by fusion is inherently well-separated, making dimensionality reduction unnecessary and potentially detrimental to information preservation.

The fact that nonparametric models, such as MLP and KNN, also achieve performance very close to that of QDA validates the quality of the feature space. However, the training time of some powerful ensemble models, such as Gradient Boosting and AdaBoost, demonstrates how computationally expensive training these algorithms can be in high-dimensional feature spaces (when n_features > n_samples). This highlights that in model selection, not only accuracy but also training time and computational resources should be considered essential criteria.

Failure analysis is critical for understanding the limitations of the model. The fact that the few misclassified examples are usually located in the transition regions between the “damaged” and “destroyed” classes reflects the inherent uncertainty in the damage grading. Even in visual interpretation, determining when a building transitions from “severely damaged” to “destroyed” can be subjective and challenging. The model’s difficulty in these borderline cases is inherent to the problem and is one of the main obstacles to achieving 100% accuracy.

5.1. Comparison with literature

The 99.5% accuracy achieved in this study is significantly higher than that reported in many existing studies in the earthquake-damage-assessment literature. Kaya (2025) reported an accuracy of 70% using Mask R-CNN. Takhtkeshha et al. (2023) achieved 82% accuracy with SVM. The superior performance observed in this study can be attributed to several key elements of the methodological approach.

1. **Feature fusion:** Instead of using a single DL model, features from five different CNN architectures (ResNet, EfficientNet, VGG, DenseNet121, and MobileNetV2) were combined to create a more diverse and complementary information set. This increases the likelihood of capturing patterns that may be overlooked by a single model.
2. **Comprehensive classifier comparison:** Systematically testing 12 classifiers maximized the likelihood of identifying the most suitable ML tool for a specific feature set. While many studies are limited to a few popular models, this study shows that a less commonly used model, such as the QDA, may be the best fit for this problem.

3. **Data quality and quantity:** With more than 18,000 images from both the original and augmented versions of the UAVs-TEBDE dataset, there was a sufficient foundation for robust model training.

5.2. Practical applications and effects of the study

The methodology developed and validated in this study has significant practical applications in disaster management. An automated system operating with nearly 99% accuracy can enable the rapid production of damage maps in the first few hours after an earthquake. These maps can guide search and rescue teams by making effective use of time, and by understanding the geographical distribution and intensity of the damage, they can facilitate logistics and aid planning for the establishment of temporary field hospitals and other aid distribution points. Additionally, it can accelerate the estimation of economic damage and enable the rapid identification of collapsed and damaged buildings in disaster areas.

5.3. Limitations and future studies

Despite its high success rate, this study has some limitations. First, the methodology was developed using a dataset that reflects the architectural characteristics of the buildings in Türkiye. It should be tested whether the model performs equally well across different geographic regions (e.g., regions with standard wooden or steel structures). Second, this study was based solely on 2D optical images obtained from UAVs. In future studies, integrating 3D point clouds or nDSM data (Erdogan & Yilmaz, 2019) or SAR images (Russo et al., 2025) obtained from UAVs into the feature set could further improve accuracy, especially in detecting types of damage such as pancake collapse, which cannot be understood from a vertical perspective. Furthermore, while the current research employed standard CNN models exclusively for feature extraction, exploring advanced architectures like Vision Transformers (ViT) or Siamese Networks could potentially yield even greater accuracy. Lastly, due to the substantial computational overhead of executing 10-fold cross-validation across 12 classifiers, a detailed ablation study isolating the individual performance of each CNN architecture with the fused vector was omitted. Subsequent research will therefore focus on quantifying the specific contributions of each feature stream.

6. Conclusion

This study presents a comprehensive and systematic methodology for classifying building damage into three categories—undamaged, damaged, and destroyed—using ultra-high-resolution UAV imagery captured after earthquakes. The methodological framework, key findings, and contributions of this study can be summarized as follows:

- **Hybrid Methodology:** The foundation of the study is a two-stage hybrid approach that combines deep features extracted from five different pretrained CNN models and tests this enriched feature set with 12 machine learning classifiers.
- **Outstanding Performance:** The findings show that the proposed methodology achieved outstanding success. In particular, the QDA model trained on the combined deep features outperformed all other models tested, achieving a mean F1 Score of 99.53%.
- **Discriminative Feature Space:** This result demonstrates that feature fusion based on deep learning creates a highly discriminative feature space for machine learning classifiers, allowing even relatively more straightforward models to outperform complex ensemble learning algorithms.
- **Statistical Robustness:** Instead of relying on a single train-test split, the stability of the models was confirmed through 10-Fold Stratified Group Cross-Validation. The QDA model exhibited the lowest standard deviation among the top classifiers, indicating that its superior performance was statistically robust and consistent across different data subsets.
- **Scientific Contribution:** The main contribution of this study is the presentation of a robust and reproducible scientific workflow that achieves higher accuracy rates than existing approaches in the literature for post-earthquake damage assessment in buildings.
- **Operational Impact:** This methodology, based on feature fusion and a comprehensive classifier comparison, has the potential to provide decision-makers with rapid, accurate, and reliable damage maps during the initial response phase after a disaster. This constitutes an essential step towards enhancing the

effectiveness of search-and-rescue operations, reducing the loss of life, and optimizing humanitarian aid efforts.

- **Future Directions:** Future studies should focus on testing the generalizability of this methodology to earthquakes in different geographies and structures, integrating additional data sources such as 3D data and SAR, and experimenting with next-generation transformer architectures such as ViT as feature extractors. Such improvements would be significant steps toward developing a fully autonomous and universal earthquake-damage detection system.

Acknowledgement

The authors declare no acknowledgments. This study received no external funding, and the authors are solely responsible for the content and writing of the manuscript.

Author contribution

Both authors contributed equally to all phases of the study, including dataset preprocessing, model development, statistical analysis, and interpretation of the results. Both authors have read and approved the final manuscript and agree to be accountable for all aspects of the work.

Declaration of ethical code

The dataset used in this study was open-source and publicly available. No human participants, patient data, animal subjects, or personally identifiable/sensitive information were used in this study. Accordingly, the materials and methods employed in this study did not require ethics committee approval or special legal permission. No administrative permission was required to access or use the data.

Conflicts of interest

The authors declare no conflicts of interest.

References

Azizi, A. (2023). A clip-based method for efficient building damage assessment using UAV images. *Journal of Emerging Technologies and Innovative Research, 10*(7), a678-a686.

Brunner, D., Lemoine, G., & Bruzzone, L. (2010). Earthquake damage assessment of buildings using VHR optical and SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 48(5), 2403–2420. https://doi.org/10.1109/TGRS.2009.2038274

Chicco, D., Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* **21**, 6. https://doi.org/10.1186/s12864-019-6413-7

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37-46. https://doi.org/10.1177/001316446002000104

Dell'Acqua, F., & Gamba, P. (2012). Remote sensing and earthquake damage assessment: experiences, limits, and perspectives. *Proceedings of the IEEE*, 100(10), 2876–2890. https://doi.org/10.1109/JPROC.2012.2196404

Dong, L., & Shan, J. (2013). A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. *ISPRS Journal of Photogrammetry and Remote Sensing*, 84, 85–99. https://doi.org/10.1016/j.isprsjprs.2013.06.011

Du, H., Lin, X., Jiang, J., Lu, Y., Du, H., Zhang, F., Yu, F., Feng, T., Wu, X., Peng, G., Deng, S., He, S., & Bai, X. (2024). A single-building damage detection model based on multi-feature fusion: A case study in Yangbi. *iScience, 27*(1), 108586.

Erdogan, M., & Yilmaz, A. (2019). Detection of building damage caused by Van earthquake using image and digital surface model (DSM) difference. *International Journal of Remote Sensing, 40*(10), 3772–3786. https://doi.org/10.1080/01431161.2018.1552816

Ge, P., Gokon, H., & Meguro, K. (2020). A review on synthetic aperture radar-based building damage assessment in disasters. *Remote Sensing of Environment, 240*, 111693. https://doi.org/10.1016/j.rse.2020.111693

Ghahrloo, M., & Mokhtarzade, M. (2025). Earthquake-induced building damage detection using the fusion of optical and radar data in intelligent systems. *Earth Science Informatics, 18*, 62.

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6), 610-621.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). *Mask R-CNN*. Proceedings of the IEEE international conference on computer vision, 2980-2988.

Ji, M., Liu, L., & Buchroithner, M. F. (2018). Identifying collapsed buildings using post-earthquake satellite imagery and convolutional neural networks: A case study of the 2010 Haiti Earthquake. *Remote Sensing*, 10(11), 1689. https://doi.org/10.3390/rs10111689

Kaur, S., Gupta, S., Singh, S., Hoang, V. T., Almakdi, S., Alelyani, T., & Shaikh, A. (2022). Transfer learning-based automatic hurricane damage detection using satellite images. *Electronics*, 11(9), 1448.

Kaya, A. Y. (2025). Detection of structural damage after an earthquake using GIS and remote sensing methods. *Turkish Journal of Agriculture - Food Science and Technology*, 13(3), 688-696.

Kaya, Ö. (2024). *Comparative Analysis of Multiple Machine Learning Algorithms for Post-Earthquake Building Damage Assessment in Hatay City Following the 2023 Earthquake* [Master's thesis, Middle East Technical University].

Kerle, N., Nex, F., Gerke, M., Duarte, D., & Vetrivel, A. (2019). UAV-based structural damage mapping: A review. *ISPRS International Journal of Geo-Information*, 9(1), 14. https://doi.org/10.3390/ijgi9010014

Kim, M., Park, S. E., & Lee, S. J. (2023). Detection of damaged buildings using temporal SAR data with different observation modes. *Remote Sensing*, 15(2), 308.

Korkmaz, K. A. (2009). Earthquake disaster risk assessment and evaluation for Türkiye. *Environmental Geology*, 57, 307-320.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.

Li, X., Yang, J., Li, Z., Yang, F., Chen, Y., Ren, J., & Duan, Y. (2022). *Building Damage Detection for Extreme Earthquake Disaster Area Location from Post-Event UAV Images Using Improved SSD*. 2022 IEEE International Geoscience and Remote Sensing Symposium, 2674-2677.

Matsuoka, M., & Yamazaki, F. (2004). Use of satellite SAR intensity imagery for detecting building areas damaged due to earthquakes. *Earthquake Spectra*, 20(3), 975–994. https://doi.org/10.1193/1.1774182

Menderes, A., Erener, A., & Sarp, G. (2015). Automatic detection of damaged buildings after earthquake hazard by using remote sensing and ınformation technologies. *Procedia Earth and Planetary Science*, 15, 257–262. https://doi.org/10.1016/j.proeps.2015.08.063

Nex, F., Duarte, D., Tonolo, F. G., & Kerle, N. (2019). Structural building damage detection with deep learning: assessment of a state-of-the-art CNN in operational conditions. *Remote Sensing*, 11(23), 2765. https://doi.org/10.3390/rs11232765

Park, S. E., & Jung, Y. T. (2020). Detection of earthquake-induced building damages using polarimetric SAR data. *Remote Sensing*, 12(1), 137.

Rao, A., Jung, J., Silva, V., Molinario, G., & Yun, S. H. (2023). Earthquake building damage detection based on synthetic-aperture-radar imagery and machine learning. *Natural Hazards and Earth System Sciences*, 23(2), 789–807. https://doi.org/10.5194/nhess-23-789-2023

Russo, L., Tapete, D., Ullo, S. L., & Gamba, P. (2025). A deep learning framework for building damage assessment using VHR SAR and geospatial data: demonstration on the 2023 Türkiye Earthquake. arXiv preprint.

Saito, K., Spence, R., & de C Foley, T. A. (2005). Visual damage assessment using high-resolution satellite images following the 2003 Bam, Iran, earthquake. *Earthquake Spectra*, 21(S1), S309–S318.

Sokolova, M. and Lapalme, G. (2009). A Systematic Analysis of Performance measures for Classification Tasks. Information Processing & Management, 45, 427-437. https://doi.org/10.1016/j.ipm.2009.03.002

Takhtkeshha, N., Mohammadzadeh, A., & Salehi, B. (2023). A rapid self-supervised deep-learning-based method for post-earthquake damage detection using UAV data (Case Study: Sarpol-e Zahab, Iran). *Remote Sensing*, 15(1), 123. https://doi.org/10.3390/rs15010123

Tong, X., Hong, Z., Liu, S., Zhang, X., Xie, H., Li, Z., Yang, S., Wang, W., & Bao, F. (2012). Building-damage detection using pre- and post-seismic high-resolution satellite stereo imagery: a case study of the May 2008 Wenchuan earthquake. *ISPRS Journal of Photogrammetry and Remote Sensing*, 68, 13–27. https://doi.org/10.1016/j.isprsjprs.2011.12.004

Vetrivel, A., Gerke, M., Kerle, N., Nex, F., & Vosselman, G. (2018). Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 140, 45–59. https://doi.org/10.1016/j.isprsjprs.2017.03.001

Xia, H., Wu, J., Yao, J., Zhu, H., Gong, A., Yang, J., Hu, L., & Mo, F. (2023). A deep learning application for building damage assessment using ultra-high-resolution remote sensing ımagery in Türkiye Earthquake. *International Journal of Disaster Risk Science*, 14, 947-962.

Xu, J. Z., Lu, W., Li, Z., Khaitan, P., & Zaytseva, V. (2019). Building damage detection in satellite imagery using convolutional neural networks. arXiv preprint arXiv:1910.06444.

Yamazaki, F., Yano, Y., & Matsuoka, M. (2005). Visual damage interpretation of buildings in Bam city using QuickBird images following the 2003 Bam, Iran, earthquake. *Earthquake Spectra*, 21(S1), 329–336.

Yu, X., Hu, X., Song, Y., Xu, S., Li, X., Song, X., Fan, X., & Wang, F. (2024). Intelligent assessment of building damage of 2023 Türkiye-Syria Earthquake by multiple remote sensing approaches. *npj Natural Hazards*, 1(1), 3.

Yusuf, Y., Matsuoka, M., & Yamazaki, F. (2001). Damage assessment after 2001 Gujarat earthquake using Landsat-7 satellite images. *Journal of the Indian Society of Remote Sensing*, 29, 17-22.

Yun, S. H., Hudnut, K., Owen, S., Webb, F., Simons, M., Sacco, P., Gurrola, E., Manipon, G., Liang, C., Fielding, E., Milillo, P., Hua, H., & Coletta, A. (2015). Rapid damage mapping for the 2015 Mw 7.8 Gorkha Earthquake using synthetic aperture radar data from COSMO–SkyMed and ALOS-2 Satellites. *Seismological Research Letters*, 86(6), 1549–1556. https://doi.org/10.1785/0220150152

Zhan, Y., Liu, W., & Maruyama, Y. (2022). Damaged building extraction using modified Mask R-CNN model using post-event aerial images of the 2016 Kumamoto Earthquake. *Remote Sensing*, 14(4), 1002.

Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8-36.