EEFIT Research Grant Scheme 2024

Project Report

Identification of Seismic Damage States in Reinforced Concrete Residential Buildings: A Deep Learning Approach

Dr. Mihaela Anca CIUPALA

Mr. Kaveesh Guwanindu ABEYSURIYA

Department of Engineering and Construction

School of Architecture, Computing and Engineering

University of East London

United Kingdom

Table of Contents

1. Introduction …… 01
1.1. Problem Statement …… 01
1.2. Project Impact …… 01
1.3. Project Context …… 02
1.4. Intervention Strategy …… 03
2. Methodology …… 04
2.1. Overview of Methodology …… 04
2.2. Unified Element-level Damage State Identification Criteria …… 05
2.3. Database Development …… 05
2.4. Multi-modal Deep Learning-based Damage Assessment Model …… 06
2.5. Performance Evaluation of the Damage Assessment Model …… 07
3. Unified Element-level Damage State Identification Criteria …… 09
3.1. Stakeholder Survey …… 09
3.2. Defining Unified Damage State Identification Criteria…… 09
3.3. Proposed Rapid Inspection Criteria and Visual Guide …… 11
4. Database Development …… 13
4.1. Data Collection …… 13
4.2. Dataset Enhancement …… 14
5. Multi-modal Deep Learning-based Damage Assessment Model …… 16
5.1. Model Architecture …… 16
5.2. Model Hyperparameters …… 17
5.3. Model Validation …… 19
5.4. Model Performance Evaluation …… 20
6. Conclusion …… 23
Acknowledgement …… 25
References …… 25

List of Figures

Figure 1: Seismic zone map of Türkiye ……... 03
Figure 2: Proposed AI intervention strategy of this project …….. 04
Figure 3: Methodology adopted in this project …… 05
Figure 4: Database development steps followed in this project …….. 06
Figure 5: Model architecture adopted in this project ……... 07
Figure 6: Transfer learning strategy adopted in this project …… 07
Figure 7: Damage limits in TBEC 2018 …….. 11
Figure 8: Dataset enhancement techniques adopted in this project ……. 15
Figure 9: Performance Indicators (Accuracy, Precision, Recall and F1 Score) of pre-trained CNN models, a) Feature extraction, b) Fine-tuning ……. 17
Figure 10: Ablation experiments for model hyperparameters, a) Learning rate, b) Model regularisation factor, c) Dropout rate, d) Number of unfreezing layers …… 18
Figure 11: Confusion matrix of the damage assessment model testing ……. 20
Figure 12: Receiver operating characteristic (ROC) curve for model testing ……... 22
Figure 13: Precision-recall (PR) curve for model testing …….. 23

List of Tables

Table 1: The visual indicators of the proposed damage state identification criteria .. 12
Table 2: Overview of the dataset developed in this project ……... 14
Table 3: Overview of the enhanced dataset developed in this project ……... 15
Table 4: Model validation performance indicator results …….. 19
Table 5: Model testing performance indicator results …… 21

1. Introduction

1.1 Problem Statement

In recent years, major earthquakes have caused partial to complete collapse of residential buildings worldwide, resulting in substantial loss of life and severe economic impacts, and producing exceptionally large caseloads of structures requiring seismic structural damage assessments (Suzuki, 2002; Alexander, 2010; Kazama and Noda, 2012; Alberto *et al.*, 2018). Rapid and accurate seismic structural damage assessment is essential in disaster-prone regions because it undermines life-safety determinations, re-occupancy decisions, the allocation of resources and the sequencing of recovery operations. An example of the scale of disruption earthquakes can make is demonstrated by the 2011 Great East Japan Earthquake, which resulted in approximately 122,000 buildings completely destroyed, 283,000 severely damaged, 748,000 partially damaged and generated approximately 25 million metric tons of disaster waste, imposing a significant environmental management burden (Reconstruction Agency, 2025). Moreover, in the 2023 Türkiye–Syria earthquake, direct physical damages in Syria alone were estimated at US$5.1 billion and displacement, while in Türkiye exceeded 2.7 million people displaced, proving how the unmet assessment demand can rapidly translate into social disruption (Ishizawa Escudero *et al.*, 2023). Consequently, it is clear that when seismic structural damage assessments are delayed or poorly coordinated, recovery process slows down, increasing stakeholder uncertainty and obstruct debris clearance, sheltering, re-occupancy and business resumption, a range of challenges that the Post-Disaster Needs Assessment (PDNA) framework seeks to mitigate through structured, timely and coordinated assessment protocols (PDNA, 2017). Therefore, the need for a rapid and accurate seismic structural damage assessment method not only determine immediate occupant safety and usability, but also capture structural integrity data that feed retrofitting strategies and the iterative improvement of building codes and damage assessment guidelines, strengthening resilience before the next event. (Rossetto and Elnashai, 2003).

1.2 Project Impact

This research project introduces a novel approach to assessing seismic structural damages in local structural elements in reinforced concrete (RC) residential buildings by utilising artificial intelligence (AI) techniques. To facilitate AI-based damage assessment, this project develops firstly a unified element-level damage state criterion that aligns with the limit state design principles established in current seismic design codes and damage assessment guidelines. The criterion is intended to enable rapid and accurate classification of earthquake-induced damage

from image data collected during post-earthquake inspections, so that an AI-based damage assessment model can analyse these images to deduce the governing failure mode at element level, thereby providing actionable evidence to inform engineering decisions on demolition, repair or retrofitting.

The following key areas where this project will contribute to impact of the project have been identified:

- The development of a unified criteria for element-level seismic damage state identification when using image-based data, aligning with relevant provisions in current seismic design codes and damage assessment guidelines. These unified criteria will serve as the basis for the classification of the damage assessment model;
- The deep learning-based damage assessment model will rapidly assess seismic structural damage in RC framed residential buildings by processing image-based data depicting local element structural failure modes (e.g. shear failure and flexural failure). This capability allows for the identification of these structural failure modes in RC residential buildings during seismic structural assessments, which can be performed at a faster rate compared to currently used traditional onsite visual methods that are time-consuming, and prone to errors;
- The deep learning-based damage assessment model, once trained on image-based data from previous structural failures, can serve as a valuable tool for identifying vulnerabilities in existing structures that have not yet experienced seismic events. By analysing the failure patterns exhibited in various structural forms, the model can provide insights into the inherent weaknesses and subsequent failure mechanisms of different building typologies. This understanding is useful for academics, researchers and industry experts who can incorporate these insights to refine and modify current seismic design codes and guidelines.

1.3 Project Context

To implement this project, Türkiye was selected as a case study owing to the high seismic activity with an unusually rich data environment in the country (Figure 1). National monitoring centres in Türkiye routinely process about 30,000 earthquakes each year, and regional networks record roughly 1,500 events per month, reflecting sustained activity along the North and East Anatolian fault systems. The 6th February 2023 Kahramanmaraş earthquakes (Mw 7.8 and 7.5) generated more than 33,000 aftershocks in the first three months and affected 11 provinces in Türkiye, affecting around 16% of the national population, highlighting the operational need for

fast, scalable triage of building damage to be tested and implemented (Disaster and Emergency Management Authority, 2023). This project focuses on RC framed residential buildings with unreinforced masonry (URM) infill walls, which is the dominant structural typology across Türkiye's building stock, hence making this system an appropriate target for method development and validation (Morales-Beltran, 2025).

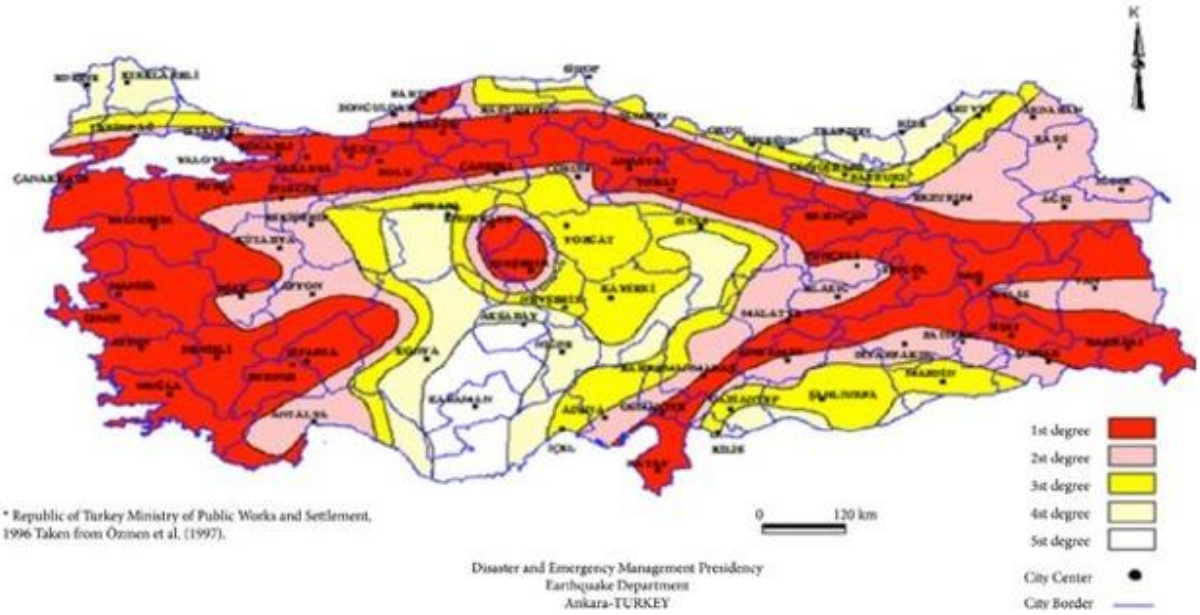

Figure 1: Seismic zone map of Türkiye (Disaster and Emergency Management Authority, 2019)

1.4 Intervention Strategy

Recent advancements in the seismic structural damage assessment field have emphasised improving accuracy and efficiency through the intervention of AI-based techniques. Figure 2 illustrates the AI intervention strategy of this project. Deep learning (DL) techniques, particularly pre-trained convolutional neural networks (CNN) have been widely employed for damage identification and classification tasks using image-based data derived from post-earthquake field inspections (Oktavianus, Chen and Lin, 2024; Abeysuriya *et al.*, 2025; Wang *et al.*, 2025; Wen *et al.*, 2025; Zou *et al.*, 2025). By extracting and interpreting complex features from images, these CNN models analyse the image-based data and classify images according to damage indicators, such as cracking, spalling, bucking and rebar exposure, with promising precision, therefore they reduce reliance on subjective human-based assessments while expediting decision-making processes (Zou *et al.*, 2022). Thus, this project focusses on identifying the failure modes in local elements of post-earthquake RC framed residential buildings with URM infill walls, in particular flexural failure and shear failure modes. One of these two failure modes is the flexural failure (ductile response), where the element can bend

and deform but still carry some load after sustaining failure. Another failure mode is the shear failure (brittle response), where the element breaks suddenly without much warning. By distinguishing between these two failure modes, engineers can better understand whether an element still has some capacity to bend and absorb energy, or if it has lost its strength completely. This clear understanding helps engineers make better decisions about whether to repair, strengthen or demolish the damaged structure.

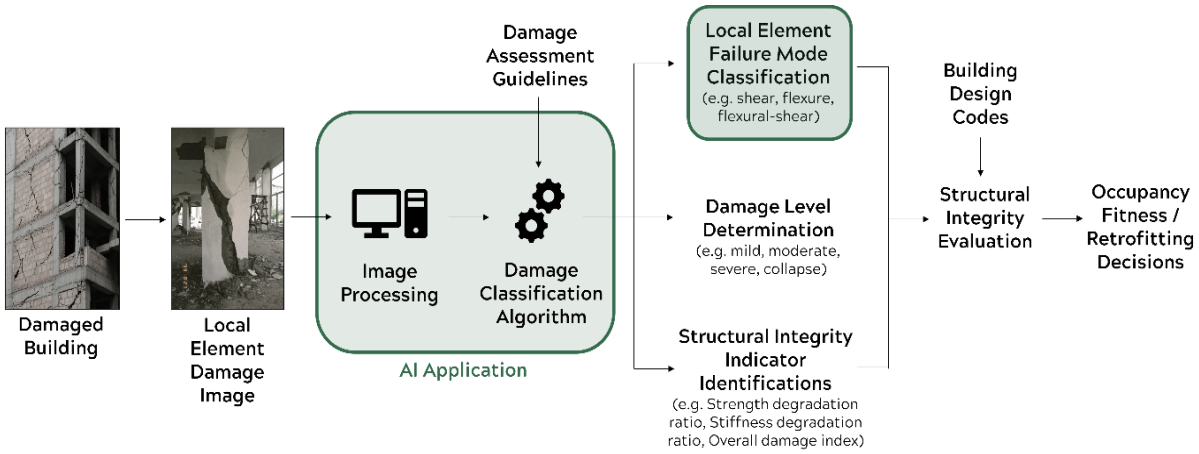

Figure 2: Proposed AI intervention strategy of this project

2 Methodology

2.1 Overview of Methodology

The methodology of this project adopts an interdisciplinary approach to develop and validate a practical, code-aligned DL-based model for element-level seismic structural damage assessment in RC framed residential buildings with URM infill walls. The principal idea is to identify the governing failure mode at the element level, which is flexural failure versus shear failure. Flexural failure happens slowly and gives warning signs, representing a ductile behaviour, whereas, shear failure happens suddenly without much warning, representing a brittle behaviour. Understanding the failure mode that occurred in the local element is very important because it tells engineers how much strength the element still has, whether it can be repaired, and whether it is safe to use. The methodology comprises the four integrated stages below (Figure 3):

- Establishment of unified element-level damage state identification criteria
- Damage Image database development
- Development of multi-modal DL-based damage assessment model
- Performance evaluation of the multi-modal DL-based damage assessment model

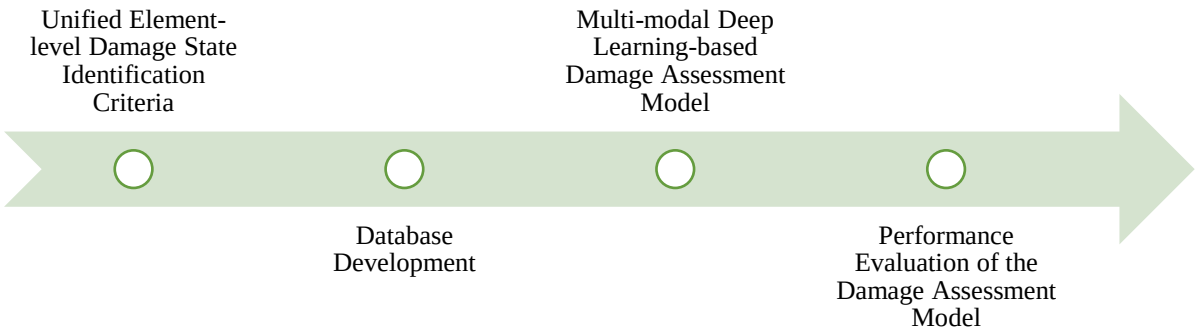

Figure 3: Methodology adopted in this project

2.2 Unified Element-level Damage State Identification Criteria

The unified criteria were derived from a survey conducted in Türkiye with two expert stakeholder focus groups: a technical focus group of civil/structural engineers, and a policy and planning focus group of local/central government officials and insurance assessors. The method of capturing and documenting damages, the prevalent building materials, structural typologies, the specific inspection parameters recorded in the field, the methods practicing engineers use to translate visual observations of damage into useful parameters for structural integrity assessments, typical assessment durations, operational constraints, and the parts of the damage assessment workflow most feasible to automation with AI were recorded. These survey responses were synthesised to produce an element-first taxonomy for RC columns, RC beams and URM infill walls. For each element, observable indicators that can be captured reliably during rapid inspection, such as type, pattern, location and width of cracks, along with presence of concrete spalling, core crushing, reinforcement exposure, reinforcement buckling and cracking in RC elements and joints, were identified. These indicators were organised along a behaviour axis that distinguishes ductile response from brittle response via the failure modes. The resulting unified element-level damage state identification criteria provide clear definitions designed to map directly into screening measures to determine repair and retrofit class and advisory safety status.

2.3 Database Development

Structural damage is documented using photographic evidence during onsite rapid inspections by engineers who produce datasets that record key indicators, such as crushed regions, crack patterns, and measurements of crack length and crack width. A domain specific dataset covering the target failure modes of this project was assembled from open-source web

repositories and from first hand field imagery gathered after recent earthquakes in Türkiye by collaborating practitioners. As explained in Figure 4, data acquisition and curation followed a multi-stage, auditable pipeline that included web scraping, segregation by structural element and failure mode, cleaning and quality control through de-duplication and exposure and blur screening with privacy review, database structuring, versioning and documentation to ensure reproducibility, and expert labelling that assigned the final failure mode classes for model training and testing. To mitigate scarcity and class imbalance that are common in post-earthquake structural damage imagery, targeted data augmentation and synthetic data generation techniques were applied while preserving orientation and texture indications from the original images that are critical for distinguishing ductile and brittle behaviours in the elements.

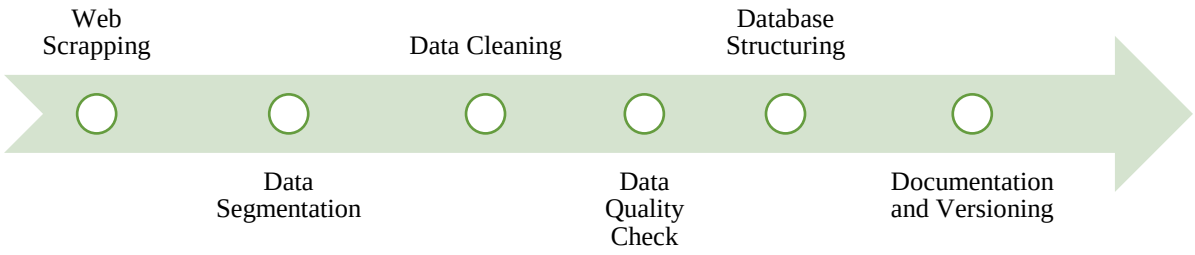

Figure 4: Database development steps followed in this project

2.4 Multi-modal Deep Learning-based Damage Assessment Model

CNNs are the most accurate DL technique for damage classification using image-based data as evidenced by current research within the seismic structural damage assessment field (Oktavianus, Chen and Lin, 2024; Abeysuriya *et al.*, 2025; Wang *et al.*, 2025; Wen *et al.*, 2025; Zou *et al.*, 2025). The architecture of the DL model classifier developed in this project is presented in Figure 5. The multi-modal DL classification model of this project inputs together an image (RGB) and a concise engineer-written textual description of the failure mode, which is used to train the model. The input image is processed using a CNN architecture, which then passes through layers of convolution and pooling to produce a classified result. In parallel, textual descriptions are computed with the contrastive language image pre-training (CLIP) text encoder, which contains layers for textual tokenisation, token embedding, transforming, normalising and text projecting. The image and text feature vectors are concatenated and fed

to a compact multi-layer fusion module that produces class scores for the failure mode classes relevant to this model. This project focuses on two failure modes: shear failure and flexural failure, which are the output classes of this model, with the potential to explore additional classes in future.

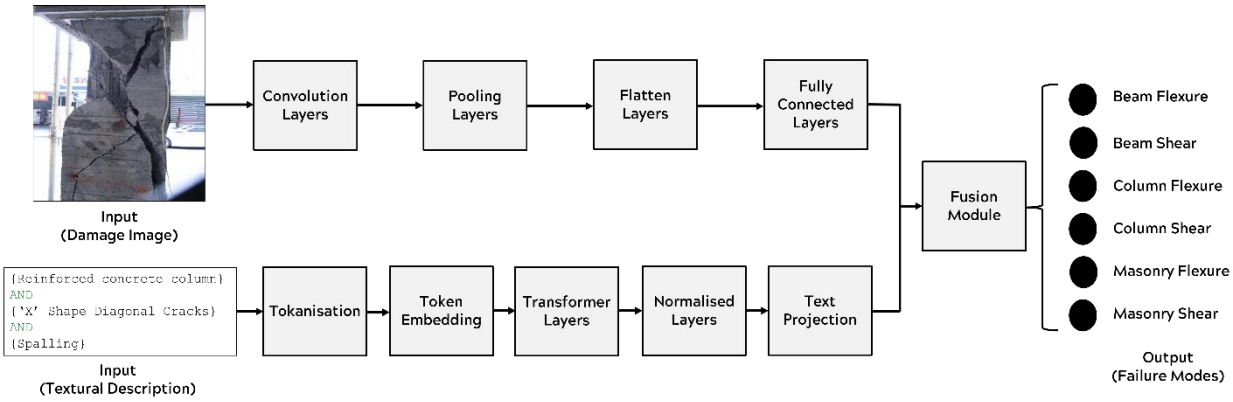

Figure 5: Model architecture adopted in this project

A transfer-learning strategy is adopted within the CNN architecture of this model, in which visual features are extracted with a pretrained CNN backbone, MobileNetV3-Large (Howard *et al.*, 2017), trained on the 'ImageNet' dataset. The backbone's classification layer is removed so it serves purely as a feature extractor. The network is then fine-tuned by unfreezing a small number of the final layers, allowing adaptation to the target dataset while the earlier layers remain fixed as illustrated in Figure 6.

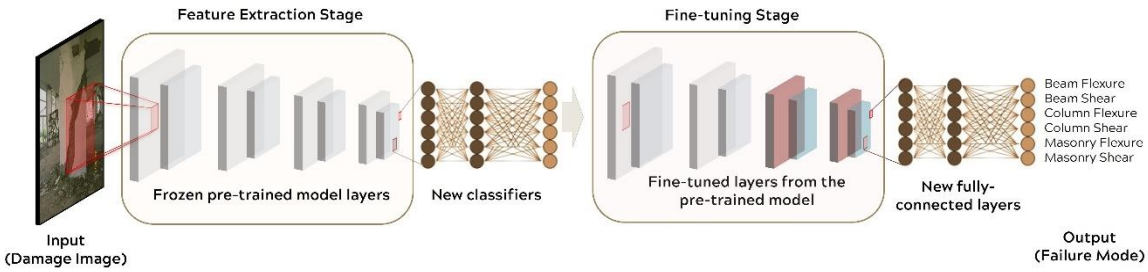

Figure 6: Transfer learning strategy adopted in this project

2.5 Performance Evaluation of the Damage Assessment Model

Evaluating the performance of the DL-based damage assessment model is essential for assessing how effectively the model generalises to unseen data that will be provided during field investigations. A confusion matrix serves as a critical tool in this process, particularly in classification tasks, as demonstrated in this project. It provides a comprehensive breakdown of the model's predictions by comparing them to the actual outcomes (Sokolova and Lapalme, 2009). The matrix consists of four key elements:

- True Positives (TP): The model correctly predicted the positive class
- True Negatives (TN): The model correctly predicted the negative class
- False Positives (FP): The model incorrectly predicted the positive class
- False Negatives (FN): The model incorrectly predicted the negative class

A confusion matrix helps derive several performance metrics, such as accuracy, precision, recall, F1 score, Area under Curve (AUC) and Average Precision (AP), which provide deeper insights into model performance (Powers, 2008).

- Accuracy measures the proportion of correct predictions out of the total, giving an overall sense of the model's performance. However, it can be misleading if the dataset is imbalanced, with significantly more instances of one class than the other:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (1)$$

- Precision, or positive predictive value, is the ratio of correctly predicted positives to all predicted positives. It is important when minimizing false positives is critical:

$$Precision = TP/(TP + FP) \quad (2)$$

- Recall measures the model's ability to correctly identify all positive instances, making it essential in cases where missing positive instances (false negatives) is costly:

$$Recall = TP/(TP + FN) \quad (3)$$

- The F1 score, the harmonic mean of precision and recall, provides a balanced metric, especially useful when class distribution is uneven. It balances the trade-off between precision and recall, offering a comprehensive view of model performance:

$$F1\ Score = 2 \times (Precision \times Recall)/(Precision + Recall) \quad (4)$$

- The Area under Curve (AUC) summarises the model's ability to distinguish between classes across thresholds:

$$AUC = \int TPR(FPR)d(FPR), \qquad TPR = \frac{TP}{TP + FN}, \qquad FPR = \frac{FP}{FP + TN} \quad (5)$$

- Average Precision (AP) summarises the Precision–Recall curve into a single value, averaging precision across recall levels:

$$AP = \sum_{n} (R_n - R_{n-1}).P_n \quad (6)$$

Where, P_n and $R_n$are the precision and recall at the $n - th$ threshold, and $(R_n - R_{n-1})$ is the increase in recall.

3 Unified Element-level Damage State Identification Criteria

This section defines the theoretical context for developing a unified criteria for identifying damage states at element level in post-earthquake RC framed residential buildings with URM infill walls. The damage state identification criteria are developed by element type, then by governing behaviour described as ductile or brittle pertaining to the observable failure modes. It relies on indicators that can be observed during rapid inspection including type, pattern, location and width of cracks, along with presence of concrete spalling, core crushing, reinforcement exposure, reinforcement buckling and cracking in RC elements and joints (Refer to Table 1). The criteria are designed to be read directly into screening outputs that support repair or retrofit choice and advisory safety status.

3.1 Stakeholder Survey

Evidence for the damage state identification criteria was gathered through structured workshops in Türkiye with two high-level stakeholder focus groups held in July 2025. The first focus group comprised civil/structural engineers who carry out technical assessments of post-earthquake structures. The second focus group comprised local and central government officials as well as insurance assessors who are involved with policy making and managing disaster responses. Participants completed standardised questionnaires during these in person sessions. Thus, these instruments captured the prevalent building materials and typologies in the affected residential building stock, the inspection steps and parameters that are normally recorded, the way visual observations are translated into judgements of structural integrity, the typical time required for assessments, the main operational challenges, and the parts of their respective workflow that could be assisted by AI.

3.2 Defining Unified Damage State Identification Criteria

Survey responses were first synthesised to determine how best to define damage state criteria that meet stakeholder needs for rapid and accurate post-earthquake decisions. This synthesis led to an element-first ontology in which each RC structural element is classified by its failure modes. This approach enables distinguishing two distinct governing failure modes in RC structural elements. The first is flexural failure, representing ductile response, where the element exhibits capacity for bending and deformation while maintaining load-carrying capability following failure initiation. The second is shear failure, representing brittle response, where the element experiences sudden failure with minimal early warning. Through diagnosis of these failure modes, engineers can assess whether an element retains residual ductility and energy absorption capacity or has experienced complete strength loss. This representation

directly informs engineering decisions regarding repair, retrofit or demolition of the damaged structure.

3.2.1 Ductile Behaviour via Flexural Failure Mode

Ductile behaviour is indicated by flexural hinging in beams or columns, seen as distributed flexural cracking aligned with bending demand, stable crack spacing concentrated near expected plastic regions, limited and recoverable stiffness loss and confined spalling, which together imply considerable rotation capacity and stable energy dissipation consistent with performance-based assessment principles and capacity design intent in modern standards such as ASCE 41 (American Society of Civil Engineers, 2017), Eurocode 8 (European Committee for Standardisation, 2005) and in the Turkish practice under TBEC 2018 (Turkish Building Earthquake Code Committee, 2018).

3.2.2 Brittle Behaviour via Shear Failure Mode

In contrast, brittle behaviour is signalled by diagonal cracking and X-shaped crack patterns in columns and joints, diagonal rupture of joint panels, sliding or splitting at lap splices, short-column diagonal damage, rapid spalling that exposes the core, concrete crushing of the confined zone and reinforcement buckling or pullout, all of which are associated with abrupt strength loss and potential reduction of axial load capacity as documented in recognised treatments of RC seismic behaviour and in element-level acceptance criteria in ASCE 41 (American Society of Civil Engineers, 2017).

TBEC 2018 (Turkish Building Earthquake Code Committee, 2018) treats seismic loading actions that are prone to brittle failure as force controlled and therefore subjects brittle elements to strict capacity checks at the element level. In practical terms this means that shear, joint shear, bond and anchorage, and confinement-related resistances must not be exceeded by the computed seismic demand for the target performance level. If the seismic demand to element capacity ratio for any primary element is greater than one, TBEC 2018 (Turkish Building Earthquake Code Committee, 2018) does not permit the structure to be classified at that performance target and the building must either be reassigned to a lower performance level or be strengthened before compliance can be claimed. This philosophy is explicitly reflected in the damage limit chart in Figure 7 that distinguishes following stages: the limited damage (SH), controlled damage (KH) and pre-collapse damage (GO).

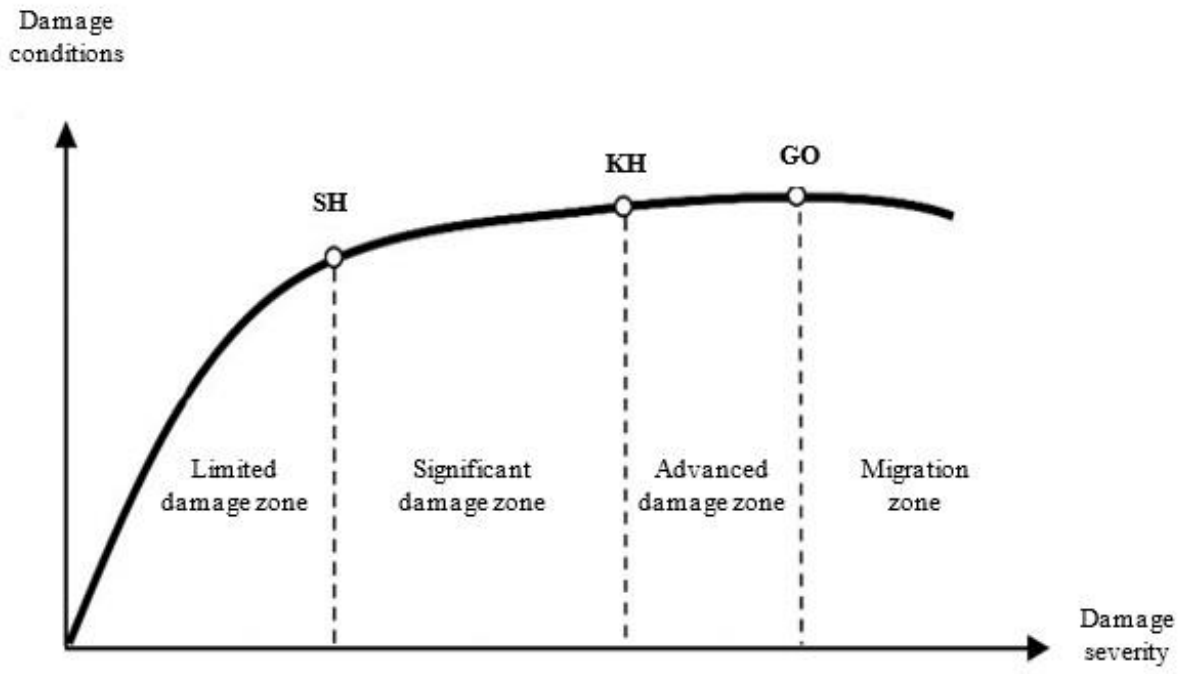

Figure 7: Damage limits in TBEC 2018 (Turkish Building Earthquake Code Committee, 2018)

When a structural element reaches its brittle capacity limit, it immediately moves out of the limited damage (SH) or controlled damage (KH) ranges and into the pre-collapse damage (GO) range. This happens regardless of how much the building has moved or drifted. In contrast, ductile flexural actions are evaluated based on how much the element can bend and rotate. These elements can handle some permanent bending within the controlled range without immediate danger.

Therefore, when engineers observe or predict that columns or joints will exceed their brittle capacity limits, this finding is critical for checking if the building meets safety requirements. This is because brittle exceedance prevents the building from achieving the desired performance level required by TBEC 2018 (Turkish Building Earthquake Code Committee, 2018). When this happens, engineers must immediately consider adding temporary supports (shoring), strengthening the structure or completely retrofitting it.

3.3 Proposed Rapid Inspection Criteria and Visual Guide

The guideline explaining the visual indicators of the proposed damage state identification criteria is presented in Table 1, which aims to support consistent field of practice and traceable decision making during rapid screening of element damages in RC framed residential buildings with URM infill walls.

Table 1: The visual indicators of the proposed damage state identification criteria

Failure Mode	**Visual Indicator**	**Damage Level**	
		Minor	**Major**
Flexural Failure	Perpendicular cracks to the member axis • Beams – Horizontal cracks • Columns – Vertical cracks Concrete Spalling • Plastic hinge regions (cover concrete) • Flaking off at beam bottom/top and column base/top Reinforcement bar yielding Exposed bars and buckling in advanced stages	• Thin cracks (<0.5mm) • No concrete chunks have fallen • Little to no rebar exposure Member has yielded but retains most axial load capacity	• Wide flexural cracks (0.5-3.0mm) • Significant cover spalling • Reinforcement bar buckling/fracture • Sagging/residual deformations in beams and/or yield penetration into core • Shortening in columns Member has lost substantial moment capacity and stiffness approaching/exceeding life safety limit
Shear Failure	Diagonal cracks at 30^0-45^0 to the axis of the element Concrete crushing along crack paths (particularly at column corners and beam sides) No significant yielding of reinforcement bars (rebar may appear intact but concrete shattered around) Shear slip (offset in the crack indicating sliding off the two sides)	• Hairline cracks (<0.5mm) • No concrete chunks dislodged Indicated brittle response of the element	• Wide, open cracks crossing the section (0.5-3.0mm) • Concrete wedges fallen out • Coarse aggregates visible • Column takes an hour-glass shape • Little observable yielding of longitudinal bars (still straight, but shear links have opened up) • Classic ‘X’ pattern with crushed concrete in short columns Columns: Element has lost most lateral load capacity and possible axial capacity (if core concrete is compromised) Beams: Precipitator mechanism if multiple beams failed

4 Database Development

4.1 Data Collection

Structural element damage is presently documented through images obtained during building inspections, conducted either on-site or remotely. These image-based datasets encompass key indicators of physical damage, including cracked/crushed regions, patterns and locations. The following modalities within these image-based datasets collectively contribute to qualitatively characterising the essential inputs, such as cracked/crushed regions, patterns and locations required for the damage assessment model, which plays a critical role in classifying local element structural failure modes:

- Visual (Shatford, 1986): The principal modality of the image contains details, such as colour (RGB/Grayscale), shapes (e.g. crack patterns, structural element shapes) and textures (e.g. in crushed areas)
- Spatial (Shatford, 1986): The encoded spatial relationships between objects (e.g. structural elements), their orientation (e.g. crack pattern propagation), and scale (e.g. crack width and length)
- Contextual (Lev Manovich, 2017): Information about the location (e.g. damaged building equipment, furniture that will be captured in the image)
- Metadata (Terry Barrett, 2005): Data, Camera settings and geotags encored in the image's metadata
- Photographic (Terry Barrett, 2005): Photographic technique used (e.g. lens, framing and exposure)

The initial step involves identifying relevant keywords for conducting web searches to locate open-source datasets. This is done either through manual searches using search engines or by employing Python® for web scraping to collect uniform resource locators (URLs) of dataset sources available on internet. Some of the key terms utilised in this project includes 'earthquake damaged buildings', 'reinforced concrete building damage during earthquakes', 'seismic structural damage to buildings', 'column damage due to earthquakes', and 'beam damage from earthquakes' among others. The open-source datasets published by the below organisations are utilised in developing this damage image database:

- The UC San Diego Earthquake Engineering Research Centre (EERC)
- The Pacific Earthquake Engineering Research) Centre at University of California Berkeley (PEER)

- Earthquake Engineering Research Institute (EERI)
- Natural Hazards Engineering Research Infrastructure program (NHERI)
- International Association for Earthquake Engineering (IAEE)
- The Global Earthquake Model (GEM)
- Purdue University database on earthquake damage records, such as 2011 New Zealand Christchurch earthquake, 2015 Nepal earthquake, 2016 Ecuador earthquake and 2016 Taiwan Meinong earthquake

The damage images from these datasets are then downloaded and segregated according to the local element structural failure modes outlined for RC framed residential buildings with URM infill walls. Table 2 provides an overview of the resulting dataset developed in this project, including the addition data of field imagery gathered after recent earthquakes in Türkiye by collaborating practitioners. The resulting dataset is a clear indication of small in scale and unevenly distributed traits often seen in image datasets related to earthquake and structural engineering, which is a limitation that poses significant challenges for developing DL models, which depend on large and diverse data for effective training.

Table 2: Overview of the dataset developed in this project

Element Type	**Local Element Failure Mode**		**Total**
	Shear Failure	**Flexural Failure**	
RC Columns	895	1,873	2,768
RC Beams	2,084	113	2,197
URM Infill Walls	1,109	401	1,510

4.2 Dataset Enhancement

The development of a domain-generalisable DL-based damage assessment model in the earthquake and structural engineering field is primarily hindered by inherent dataset constraints and imbalances, as seen in Table 2. To expand the size and variability within the training dataset, a range of dataset enhancement strategies were employed. Thus, dataset augmentation was used, which included geometric transformations (e.g. rotation, translation, scaling and flipping), colour space adjustments (e.g. brightness, contrast and saturation), noise injection (e.g. Gaussian, Speckle and Salt-and-pepper) and elastic modifications (e.g. cropping, padding, stretching and compressing). These broaden the model's exposure to diverse arrangements, image quality variations, lighting conditions and shadows caused due to image capture methods (e.g. digital cameras, phone cameras and drone cameras) as well as ambient noise. Alongside

these, synthetic data generation methods, namely CutMix and MixUp, produced composite images by merging attributes from existing samples, distinctly enlarging the dataset. Together, these augmentation and synthetic processes improved dataset balance, diversity and scale, thereby reducing overfitting and fostering stronger domain generalisation in DL-based models. Representative examples of the data augmentation and synthetic data generation methods applied in this work are shown in Figure 8. After applying these dataset enhancement techniques, each class was set to contain 1,000 images as seen in Table 3, and each image contained a textual description of the damage features as corresponding to the failure mode.

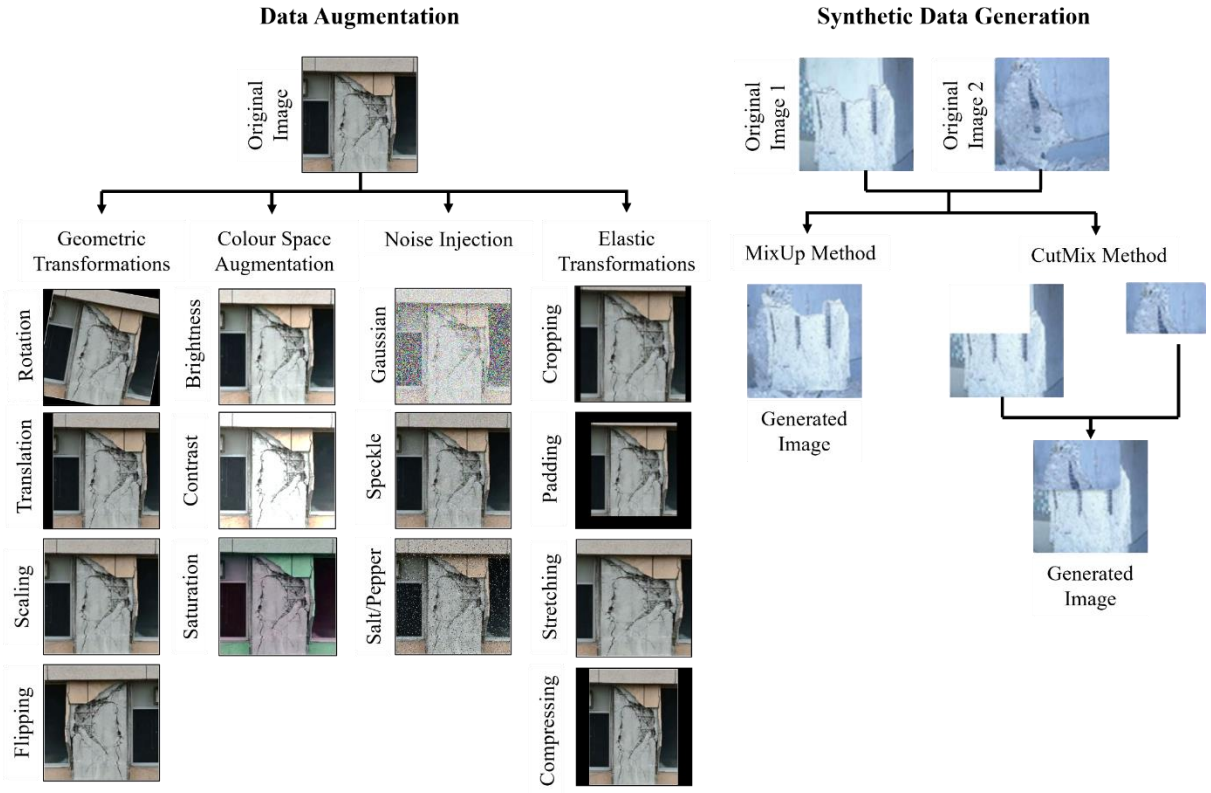

Figure 8: Dataset enhancement techniques adopted in this project

Table 3: Overview of the enhanced dataset developed in this project

Element Type	**Local Element Failure Mode**		**Total**
	Shear Failure	**Flexural Failure**	
RC Columns	1,000	1,000	2,000
RC Beams	1,000	1,000	2,000
URM Infill Walls	1,000	1,000	2,000
	Total Images in the enhanced dataset		**6,000**

5 Multi-modal Deep Learning-based Damage Assessment Model

This section describes in detail the training protocol and hyperparameter tuning and validation of the multi-modal classifier for the DL-based damage assessment model. The training phase of the model accepts a field photographical image of a structural element together with a short, free-text description written by an experienced engineer, and it predicts a discrete damage class in new and unseen field images that are processed through the trained model as explained in Section 2.4. All experiments were implemented in Python and executed on a cloud graphical processing unit (GPU) environment using an NVIDIA L4 Tensor Core device with 22.5 GB of random access memory (RAM) via the Google Colab Pro platform.

5.1 Model Architecture

The architecture of this multi-modal DL model combines images and textual descriptors to classify images more accurately. The architecture processes two parallel streams of data: an RGB image and a descriptive text caption of the local element failure mode observed in the image. The image undergoes several pre-processing steps, resizing, center-cropping and normalising using the parameters set by ImageNet dataset, which is used to pre-train this multi-modal model during the feature extraction and fine-tuning stages as explained in Section 2.4. These pre-processed images are then fed into MobileNetV3-Large pretrained CNN that extracts dimensional feature vectors from these images to represent the visual content of the image. Simultaneously, the text description is tokenised using CLIP and processed through CLIP's text encoder, which is a transformer-based model, to produce a dimensional semantic feature vector.

The key innovation occurs in the next stage where the image and textual vectors are fused to create a dimensional representation that captures both visual (image) and textual information. This fused representation then passes through a classification neural network with model hyper parameters explained in section 5.2. This maps the combined features to class predictions, which is the failure mode of the structural element. The model learns by minimising cross-entropy loss during training, effectively teaching it to leverage both what it "sees" in the image and what it "understands" from the text to make more informed classification decisions. This multi-modal approach allows the model to distinguish visually between similar objects using textual descriptions, making it more robust than single-modality approaches.

5.2 Model Hyperparameters

To identify the most suitable pretrained CNN model for the target dataset among the pretrained CNN models considered in this project, as shown in Figure 9a, a two-stage comparative experiment was conducted in which each candidate model was trained, validated and tested under identical datasets, where their performance indicators were compared to guide the final model selection. In Stage I, all convolutional layers of each pretrained model were frozen so that the networks operated strictly as feature extractors on the target domain. Only models achieving at least 0.75 on accuracy, precision, recall, and F1 score simultaneously were advanced to Stage II (Figure 9a).

In Stage II, the shortlisted models were fine-tuned by unfreezing the last few layers, enabling the network to adapt to domain-specific characteristics of the dataset, consistent with established transfer learning practice (Figure 9b). Among the evaluated architectures, MobileNetV3-Large achieved the strongest overall results, with an accuracy of 0.86, a precision of 0.88, a recall of 0.87, and an F1 score of 0.87, while also exhibiting the shortest end-to-end runtime of 4 hours and 26 minutes.

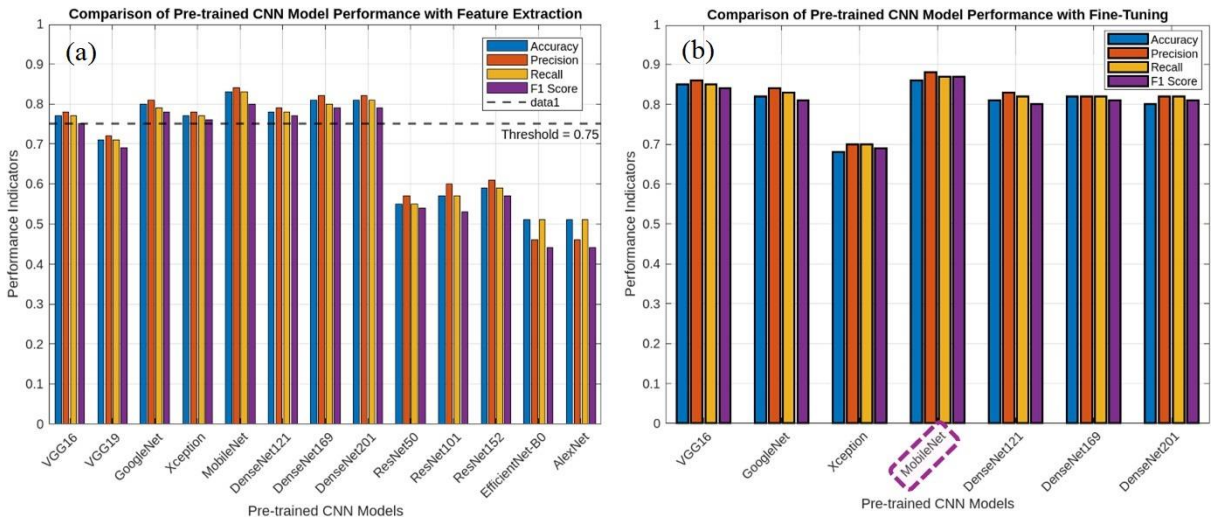

Figure 9: Performance Indicators (Accuracy, Precision, Recall and F1 Score) of pre-trained CNN models, a) Feature extraction, b) Fine-tuning

Using MobileNetV3-Large as the convolutional backbone, the model's regularisation was optimised through an ablation experiment that varied the learning rate (Figure 10a), the L2 weight-decay regularisation factor (Figure 10b), the dropout probability (Figure 10c) and the

fine-tuning depth (Figure 10d) measured as the number of unfrozen layers. The following are the optimised model hyperparameters used in the model of this project:

- Learning rate: 1.00E-05
- Model regularisation factor: 0.02
- Dropout rate: 0.6
- Number of unfreezing layers in fine-tuning: 10

The batch size considered was 32, while the number of epochs was retained at 100.

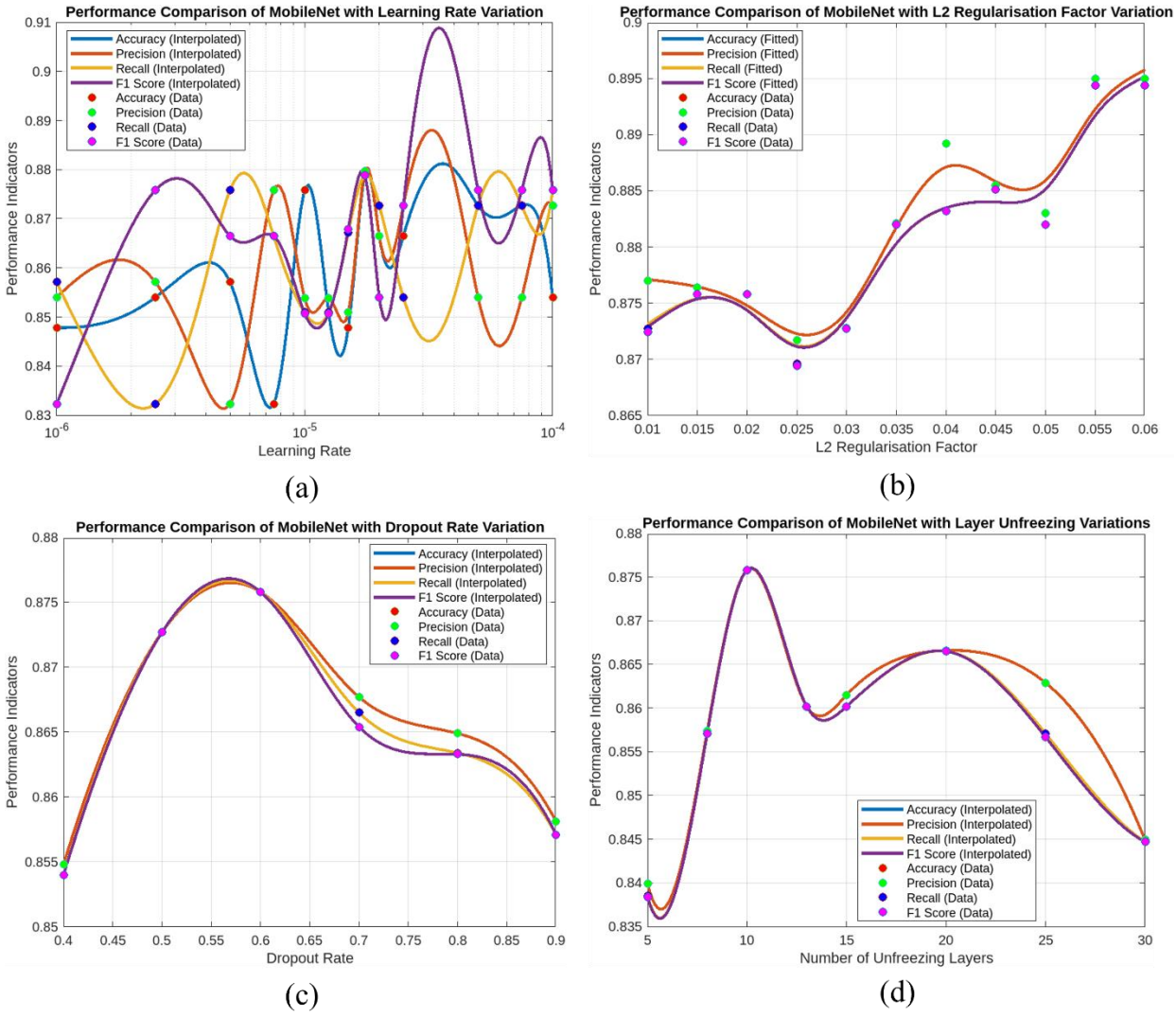

Figure 10: Ablation experiments for model hyperparameters, a) Learning rate, b) Model regularisation factor, c) Dropout rate, d) Number of unfreezing layers

5.3 Model Validation

Five-fold cross-validation was performed on the training dataset, which comprised 80% of the full set of 6,000 images (refer Table 3), using a dataset split that preserved class proportions in every fold. The pool was partitioned into five equal subsets. In each round the model was trained on four subsets and validated on the remaining subset, with the role of the validation subset rotated so that every sample served once for validation and four times for training. Within each fold the same model architecture with the same finalised model hyperparameters was used to ensure consistency. Model was reinitialised at the start of each fold to avoid information leakage. Fold-wise confusion matrices were recorded on the corresponding validation splits and then aggregated to summarise cross-validated performance, while the held-out 20% test set remained untouched until final evaluation.

This procedure works very well when dealing with a small amount of data. It helps in two important ways. First, it maximises the use of scarce labelled data for model fitting. This means it uses every piece of the limited labelled data as much as possible to train the model. Second, it provides a low variance estimate of generalisation through repeated, stratified resampling. This means it tests the model many times using different combinations of the data in a balanced way, which gives a more reliable prediction of how well the model will work on new, unseen data. The model cross validation performance indicator results are shown in Table 4.

Table 4: Model validation performance indicator results

Cross Validation Results	**Mean Accuracy**	**Mean Precision**	**Mean Recall**	**Mean F1 Score**
(mean +/- standard deviation)	0.919 ± 0.013	0.921 ± 0.011	0.919 ± 0.013	0.919 ± 0.013

Based on statistical judgement, the small standard deviations, which are about 1%, indicate stable performance across the folds and suggest that the training procedure is not sensitive to the split (Hastie, Tibshirani and Friedman, 2009), which is consistent with the cross-validation design described above for small datasets. The precision and recall values are very similar to each other, which means there is a balanced trade-off between false positives and false negatives. Additionally, the F1 score closely matches the overall accuracy, which suggests that no single class performs much worse than the average performance across all classes. This is a clear indication that the model architecture developed in this project works efficiently with small datasets, while ensuring the model can generalise well without overfitting.

5.4 Model Performance Evaluation

To assess out-of-sample generalisation of the model, the final model is evaluated on the untouched 20% hold-out test set. As this split was never used for training or model selection, it provides an unbiased benchmark against the cross-validation results and the fold-level validation metrics. Figure 11 presents the resulting confusion matrix for the test evaluation.

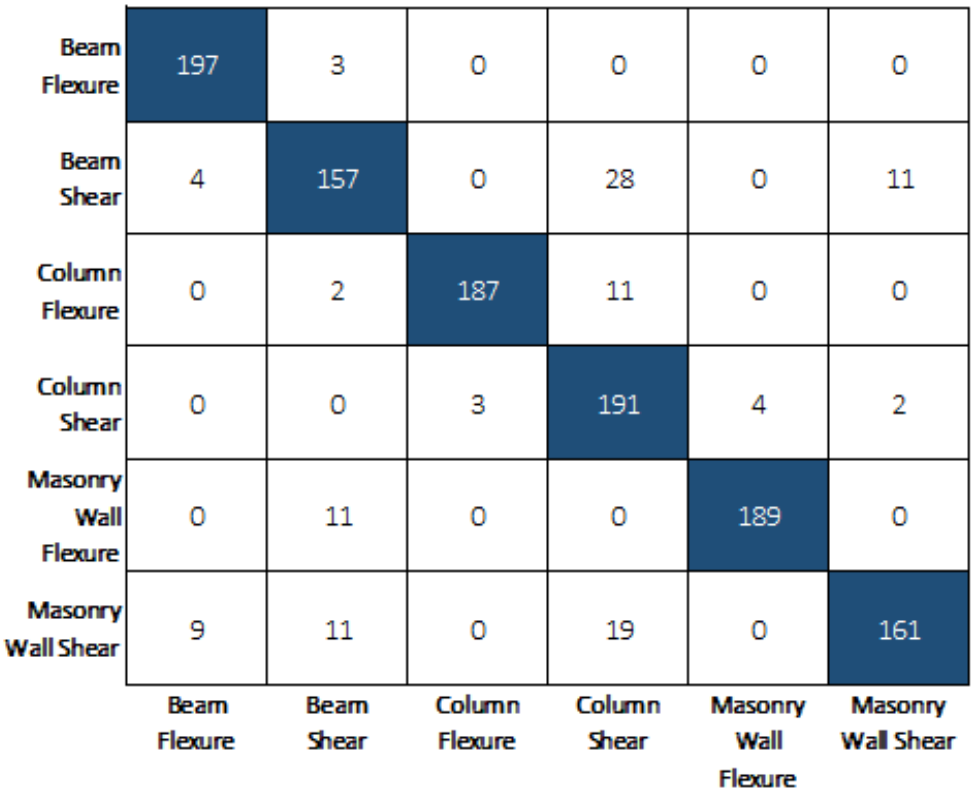

Figure 11: Confusion matrix of the damage assessment model testing

Table 5 summarises the model's performance on the test set that was held-out, along with class-wise and weighted indicators. Thus, the mean test accuracy of 0.902 closely matches the cross-validated mean accuracy of 0.919, indicating good generalisation from training to unseen test data. The held-out test results indicate that the multi-modal classifier generalises well, with a weighted F1 score of 0.902, weighted precision of 0.908 and strong ranking quality reflected by a weighted area under the curve (AUC) value of 1.00 and average precision (AP) of 0.980. Flexural damage modes are recognised most reliably: beam flexure attains an F1 score of 0.961 with recall 0.985 and precision 0.938, masonry wall flexure reaches an F1 score of 0.962 with precision 0.979 and recall 0.945, and column flexure achieves an F1 score of 0.959 with precision 0.984 and recall 0.935. Performance is slightly lower for shear classes. Beam shear records an F1 score of 0.818 with recall 0.785 and 43 false negatives, while masonry wall shear yields an F1 score of 0.861 with recall 0.805 and 39 false negatives. Column shear shows good

sensitivity (recall) of 0.955, but slightly reduced precision of 0.767, and the highest false-positive count at 58, indicating a tendency to marginally over-predict this class. These patterns in performance indicators suggest that the model sometimes gets confused between shear and flexural failure modes when they occur in the same type of structural element. However, the probabilistic separability of classes remains high, meaning the model can still distinguish between different failure modes in most cases. To fix this confusion problem, there are several approaches that could be explored. First, targeted dataset augmentation and synthetic data generation techniques or additional labelled examples for shear failure would give a better enhanced dataset, so that the model will have more examples to learn from. Second, class-aware loss tuning would help the model pay more attention to the categories it struggles with. These improvements would likely increase the model's sensitivity in these categories, and eventually improve its capabilities in classifying shear failure classes. At the same time, these changes would not sacrifice the strong overall precision and accuracy that the model already achieves.

Table 5: Model testing performance indicator results

Performance Indicators	**Beam Flexure**	**Beam Shear**	**Column Flexure**	**Column Shear**	**Masonry Wall Flexure**	**Masonry Wall Shear**	**Weighted Mean**
True Positives	197	157	187	191	189	161	180
True Negatives	987	973	997	942	996	987	980
False Positives	13	27	3	58	4	13	20
False Negatives	3	43	13	9	11	39	20
Mean Accuracy	-	-	-	-	-	-	0.902
Class-wise Accuracy	0.987	0.942	0.987	0.944	0.988	0.957	-
Precision	0.938	0.853	0.984	0.767	0.979	0.925	0.908
Recall	0.985	0.785	0.935	0.955	0.945	0.805	0.902
F1 Score	0.961	0.818	0.959	0.851	0.962	0.861	0.902
Area under Curve (AUC)	1.000	0.990	1.000	0.990	1.000	0.990	1.000
Average Precision (AP)	1.000	0.950	1.000	0.970	0.990	0.950	0.980

The receiver operating characteristic (ROC) analysis shown in Figure 12 demonstrates that the model can almost perfectly separate different failure mode classes. The ROC curve is plotted by showing the true positive rate (TPR) on the y-axis (Equation 5) against the false positive rate (FPR) on the x- axis (Equation 5) at different classification thresholds. The AUC value

represents the area under this curve, with values for all the failure mode classes closer to 1.000 indicating strong generalisation ability of the model. Each individual class achieved AUC values between 0.990 and 1.000, while the overall weighted mean AUC reached 1.000. These results show extremely low false positive rates across all classes, meaning the model rarely incorrectly classifies among the classes.

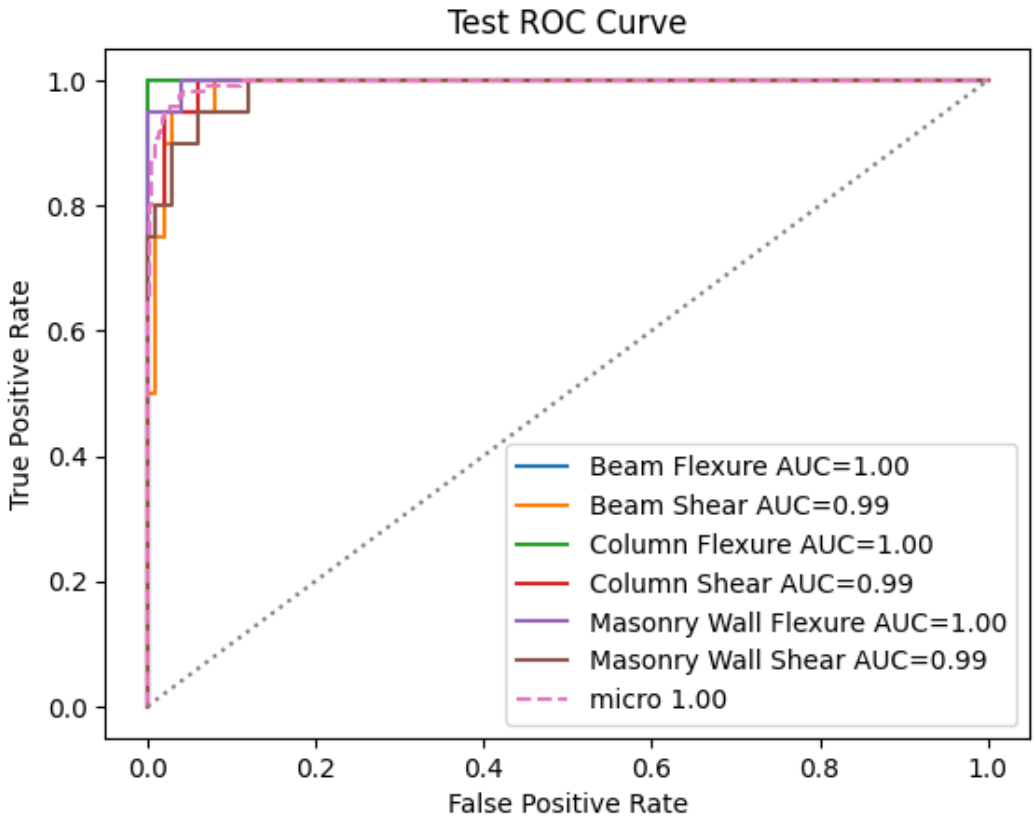

Figure 12: Receiver operating characteristic (ROC) curve for model testing

The precision–recall curves in Figure 13 show similarly strong performance. The model maintains high precision over a wide recall range, which means it stays accurate even when trying to classify among classes that have similar features or persist class imbalances. Each class achieved an Average Precision (AP) between 0.950 and 1.000 (Equation 6), with an overall weighted mean AP of 0.980. These results verify the robust performance even under class imbalanced conditions, meaning the model works well even when some damage types are much more common than others in the dataset.

Overall, these results validate the success of the model architecture developed in this project for classifying failure modes of RC framed residential buildings with URM infill walls. The

model has proven to be a useful tool for the damage assessment process in seismic prone regions, as demonstrated in this particular case study of Türkiye.

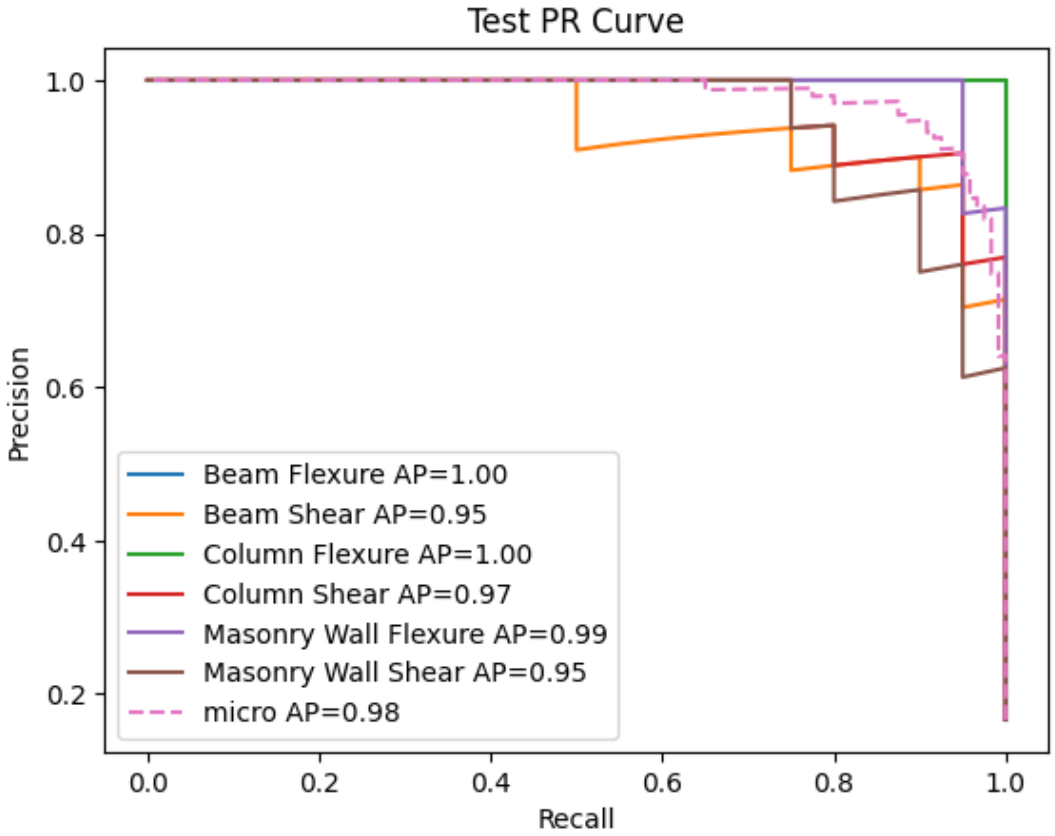

Figure 13: Precision-recall (PR) curve for model testing

6 Conclusion

This research project introduced a novel approach to assessing seismic structural damages in local elements of RC framed residential buildings with URM infill walls by developing a multi-modal DL-based classification model. Based on a case study on the RC framed residential building stock in Türkiye, the classification criteria for this DL model were specifically developed to ensure identification of the governing failure mode of RC elements, distinguishing between flexural failure and shear failure. This distinction is important because these failure modes exhibit fundamentally different characteristics. Flexural failure happens slowly and gives warning signs, representing ductile behaviour that allows for controlled structural response. In contrast, shear failure happens suddenly without much warning, representing brittle behaviour that poses immediate safety concerns. Understanding which

failure mode occurred in the local element is essential because it directly informs engineers about the element's residual strength, repairability potential and safety status for continued use.

The DL model leverages image-based data combined with textual descriptions of failure modes visible from the images to train the model. Once trained, the model can effectively classify new and unseen images of damaged elements accurately, identifying their governing failure mode. This capability provides engineers with a reliable tool for rapid damage assessment and informed decision-making in post-seismic assessments.

The model achieved superior performance in model validation, with a mean accuracy of 0.919 ± 0.013 and maintained strong generalisation during model testing, reaching a mean accuracy of 0.902, a weighted mean area under the curve (AUC) of the receiver-operating characteristic (ROC) curve of 1.000 and a weighted mean average precision (AP) of 0.980. These results prove that this model can successfully generalise to new and unseen data without overfitting, hence a useful tool to be deployed in field assessments in earthquake prone regions, as demonstrated in this particular case study of Türkiye.

While the model demonstrates strong performance, several limitations should be acknowledged to guide future improvements. The dataset, although carefully curated and audited, is relatively small in size and shows slight class imbalance even when augmented, where some damage types have fewer examples than others. Moreover, the current model analyses individual two-dimensional photographs along with brief text descriptions, but it lacks explicit geometric context and multi-view consistency that could provide richer understanding of the damage patterns. To further improve the model performance, several future directions are proposed. First, expanding the dataset diversity would provide the model with more varied examples of damage features in image data, helping it perform better across different building types and failure modes that have not been covered in this project. Second, exploring further data augmentation and synthetic data generation techniques could artificially enhance the training dataset.

Most importantly, the next crucial step involves pursuing field deployment studies to test the model in real-world conditions. This would include developing mobile inference using edge computing techniques, allowing engineers to use the model directly on smartphones or tablets at damage sites. Additionally, creating human-in-the-loop interfaces aligned with relevant code provisions would ensure that engineers can interact with the model's predictions while following established building codes and safety standards.

The findings of this project create transferable knowledge by successfully combining advanced AI techniques with traditional seismic structural damage assessment practices. Combined with the proposed future research directions, these results provide practical pathways to enhance urban resilience and strengthen global disaster preparedness.

Acknowledgement

The financial contribution of the 2024 Earthquake Engineering Field Investigation Team (EEFIT) Research Grant Scheme, the Institution of Structural Engineers, United Kingdom, is greatly appreciated.

References

Abeysuriya, K.G., Ciupala, M.A., Ghorashi, S.A. and Ilki, A. (2025) 'Enhancing Seismic Structural Damage Assessment of Low-to-medium Rise Reinforced Concrete Framed Buildings Using Artificial Intelligence', in *WIT Transactions on The Built Environment*. WIT Press, pp. 29–40. Available at: https://doi.org/10.2495/ERES250031.

Alberto, Y., Otsubo' M, Kyokawa, H., Kiyota, T. and Towhata, I. (2018) 'Reconnaissance of the 2017 Puebla, Mexico earthquake.', *Soils Found*, 58(5), pp. 1073–1092.

Alexander, D.E. (2010) 'The L'Aquila earthquake of 6 April 2009 and Italian Government policy on disaster response', *J Nat Resour Policy Res*, 2(4), pp. 325–342.

American Society of Civil Engineers (2017) *Seismic Evaluation and Retrofit of Existing Buildings*. American Society of Civil Engineers. Available at: https://doi.org/10.1061/9780784414859.

Disaster and Emergency Management Authority (2019) *Türkiye Earthquake Hazard Map*. Available at: https://www.afad.gov.tr/turkiye-deprem-tehlike-haritasi (Accessed: 14 August 2025).

Disaster and Emergency Management Authority (2023) *PRESS BULLETIN-36 about the Earthquake in Kahramanmaraş, Ministry of Interior, Republic of Turkey*. Available at: https://en.afad.gov.tr/press-bulletin-36-about-the-earthquake-in-kahramanmaras?utm_source=chatgpt.com (Accessed: 14 August 2025).

European Committee for Standardisation (2005) *Eurocode 8: Design of structures for earthquake resistance – Part 3: Assessment and retrofitting of buildings* . June.

Hastie, T., Tibshirani, R. and Friedman, J. (2009) *The Elements of Statistical Learning*. New York, NY: Springer New York. Available at: https://doi.org/10.1007/978-0-387-84858-7.

Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H. (2017) 'MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications', *ArXiv*, abs/1704.04861. Available at: https://api.semanticscholar.org/CorpusID:12670695.

Ishizawa Escudero, O.A., Daniell, J.E., Pomonis, A., Macabuag, J.L.D.C., Brand, J., Schaefer, A., Romero, R., Otálora, S.G., Khazai, B., Esper, S., Cox, K.D. and Gunasekera, R. (2023) *Global Rapid Post-Disaster Damage Estimation (GRADE) Report : Mw 7.8 Türkiye-Syria Earthquake - Assessment of the Impact on Syria : Results as of February 20, 2023 (English)*. Washington DC. Available at: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099084502282328299 (Accessed: 13 August 2025).

Kazama, M. and Noda, T. (2012) 'Damage statistics (Summary of the 2011 off the Pacific Coast of Tohoku Earthquake damage)', *Soils Found*, 52(5), pp. 780–792.

Lev Manovich (2017) *Instagram and Contemporary Image*.

Morales-Beltran, M. (2025) 'Understanding 60 years of soft storey in Türkiye: an interdisciplinary perspective', *Natural Hazards*, 121(10), pp. 11297–11336. Available at: https://doi.org/10.1007/s11069-025-07258-4.

Oktavianus, A., Chen, P.H. and Lin, J.J. (2024) 'Intelligent post-earthquake building recovery system: A framework combining BIM and deep learning', *Journal of Building Engineering*, 98, p. 111366. Available at: https://doi.org/10.1016/J.JOBE.2024.111366.

PDNA (2017) *Employment, Livelihood and Social Protection, PDNA Guidelines Volume B*. Washington, DC. Available at: https://documents1.worldbank.org/curated/en/950451493103222234/pdf/114523-WP-PUBLIC-ADD-SERIES-pdna-guidelines-vol-b-employment-livelihood-social-protection.pdf (Accessed: 13 August 2025).

Powers, D. (2008) 'Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation', *Mach. Learn. Technol.*, 2.

Reconstruction Agency, G. of J. (2025) *Great East Japan Earthquake (GEJE): Damage to buildings, https://www.reconstruction.go.jp/english/topics/GEJE/index.html*.

Rossetto, T. and Elnashai, A. (2003) 'Derivation of vulnerability functions for European-type RC structures based on observational data', *Engineering Structures*, 25(10), pp. 1241–1263. Available at: https://doi.org/10.1016/S0141-0296(03)00060-9.

Shatford, S. (1986) 'Analyzing the Subject of a Picture: A Theoretical Approach', *Cataloging & Classification Quarterly*, 6(3), pp. 39–62. Available at: https://doi.org/10.1300/J104v06n03_04.

Sokolova, M. and Lapalme, G. (2009) 'A systematic analysis of performance measures for classification tasks', *Information Processing & Management*, 45(4), pp. 427–437. Available at: https://doi.org/10.1016/j.ipm.2009.03.002.

Suzuki, K. (2002) 'Report on Damage to Industrial Facilities in the 1999 Kocaeli Earthquake, Turkey', *Journal of Earthquake Engineering*, 6(2), pp. 275–296. Available at: https://doi.org/https://doi.org/10.1080/13632460209350417.

Terry Barrett (2005) *Criticizing Photographs: An Introduction to Understanding Images*. 4th ed. McGraw-Hill Education.

Turkish Building Earthquake Code Committee (2018) *Turkish Building Earthquake Code 2018*.

Wang, S., Wang, W., Hou, H. and Yang, Z. (2025) 'Computer vision-aided damage assessment of corrugated steel plate shear links under earthquake excitations', *Engineering Structures*, 332, p. 120023. Available at: https://doi.org/10.1016/j.engstruct.2025.120023.

Wen, W., Xu, T., Hu, J., Ji, D., Yue, Y. and Zhai, C. (2025) 'Seismic damage recognition of structural and non-structural components based on convolutional neural networks', *Journal of Building Engineering*, 102, p. 112012. Available at: https://doi.org/10.1016/j.jobe.2025.112012.

Zou, D., Zhang, M., Bai, Z., Liu, T., Zhou, A., Wang, X., Cui, W. and Zhang, S. (2022) 'Multicategory damage detection and safety assessment of post-earthquake reinforced concrete structures using deep learning', *Computer-Aided Civil and Infrastructure Engineering*, 37(9), pp. 1188–1204. Available at: https://doi.org/10.1111/mice.12815.

Zou, Z., Yang, S., Wang, M. and Song, B. (2025) 'A damage assessment method for masonry structures based on multi scale channel shuffle dilated convolution and ReZero-Transformer', *Journal of Building Engineering*, 103, p. 112002. Available at: https://doi.org/10.1016/j.jobe.2025.112002.