BUILDING DAMAGE LEVEL CLASSIFICATION USING DEEP LEARNING: A CNN-BASED APPROACH FOR POST-EARTHQUAKE STRUCTURAL ASSESSMENT

Simone Saquella[1], Michele Scarpiniti[2], Livio Pedone[3], Giulia Angelucci[3], Mattia Francioli[3], Michele Matteoni[3], and Stefano Pampanin[3]

[1] Department of Astronautical, Electrical and Energy Engineering (DIAEE),
Sapienza University Rome, Italy
e-mail: simone.saquella@uniroma1.it

[2] Department of Information Engineering, Electronics and Telecommunications (DIET),
Sapienza University Rome, Italy
e-mail: michele.scarpiniti@uniroma1.it

[3] Department of Structural and Geotechnical Engineering (DISG),
Sapienza University Rome, Italy
e-mail: {livio.pedone, giulia.angelucci, mattia.francioli, michele.matteoni,
stefano.pampanin}@uniroma1.it

Abstract

In the aftermath of major earthquakes, a rapid and accurate structural damage assessment is crucial for emergency response and recovery. Convolutional Neural Networks (CNNs) have emerged as effective tools for automating this process, offering standardized evaluations that complement traditional visual inspections. This study explores the use of the VGG16 architecture for post-earthquake damage classification, leveraging transfer learning and data augmentation techniques to enhance accuracy. The dataset comprises 5,000 RGB images sourced from the PHI-Net dataset and the INGV DFM database, categorized into four damage levels. Through extensive pre-processing and augmentation, VGG16 achieved a test accuracy of 89.33%, with high precision and recall for undamaged and severe damage classes. However, distinguishing minor damage remains still challenging. These findings highlight CNNs' potential in automating structural damage assessment, supporting more efficient post-disaster decision-making.

Keywords: Building damage assessment, Post-earthquake structural damage, Structural health monitoring, CNN, Computer Vision, Transfer learning.

1 INTRODUCTION

In the field of structural and seismic engineering, a prompt and precise evaluation of structural integrity following an earthquake is crucial. Assessing damage of buildings caused by seismic activity is fundamental for disaster response and recovery planning. After a major earthquake, immediate inspections are essential to determine the safety condition of affected structures. Traditionally, this process involves a "tagging" methodology, which relies on visual surveys and expert evaluations to classify the severity of damage [1, 2]. However, a thorough and quantitative estimation of the remaining structural capacity is necessary to guide decisions regarding re-occupancy and whether to repair or demolish [3]. Accurate and timely classification of damage levels can significantly improve emergency response efficiency, inform computational structural models for refinement, and optimize the distribution of resources for risk management. Despite their reliability, conventional approaches are often labor-intensive and dependent on expert judgment, limiting their effectiveness in large-scale post-disaster assessments.

Machine learning techniques, particularly Deep Learning (DL) and Convolutional Neural Networks (CNNs), offer a promising pathway for Structural Health Monitoring (SHM), facilitating automation and standardization of image-based evaluations while improving accuracy and efficiency.

Recent progress in artificial intelligence has enabled the development of automated damage recognition frameworks, such as the crack detection method introduced in [4]. Notably, CNNs have exhibited exceptional performance in image-based classification, as evidenced by Gao and Mosalam [5], who utilized DL techniques to estimate seismic damage. Similarly, the PEER (Pacific Earthquake Engineering Research Center) Hub ImageNet (PHI-Net or Φ-Net) Challenge has significantly advanced the field by providing benchmark datasets for training and validating structural damage identification algorithms. Nevertheless, existing models frequently suffer from limited and inconsistently labeled datasets, reducing their generalization capability in practical scenarios.

A comprehensive review by Azimi et al. [6] outlined the application of data-driven SHM techniques using DL-based methodologies. Their findings emphasize the high potential of AI-driven approaches in analyzing structural data with remarkable precision. The integration of these techniques enables continuous monitoring and near real-time evaluation, thereby minimizing dependence on manual inspections.

Gao and Mosalam [7] investigated the role of deep transfer learning in image-based damage recognition, demonstrating that utilizing pre-trained networks significantly enhances classification accuracy. This method mitigates challenges associated with limited datasets by leveraging large-scale models, fine-tuned for specific structural assessment tasks. Their research highlights how transfer learning improves the reliability of automated damage evaluation systems [7, 8], while the Φ-Net dataset provides an extensive benchmark resource for structural image analysis, enabling rigorous evaluation of DL-based classification techniques [9], essential for advancing the field.

Building upon these foundational studies, Ogunjinmi et al. [10] explored the effectiveness of transfer learning in CNNs for rapid post-seismic structural assessment. Their study employed CNNs to classify damage in structural imagery, successfully identifying and categorizing severity levels. Transfer learning facilitated model adaptation by utilizing knowledge from large pre-trained datasets, thereby improving performance in real-world conditions.

To ensure the robustness of this study, a curated and meticulously annotated dataset has been

compiled, consisting of:

1. An optimized version of the Φ-Net dataset, refined for consistency and accuracy;

2. A newly constructed dataset derived from the Macroseismic Photographic Database (DFM) by INGV (Italian National Institute of Geophysics and Volcanology) [11], which includes labeled samples from past Italian earthquakes.

This dataset is used to train and validate different CNN-based architectures for the classification of the damage levels of buildings. Among the twelve implemented architectures, the VGG16 has provided the best performance with an overall accuracy of 89.33%, hence establishing as a good approach for real-world scenarios.

The remainder of this paper is structured as follows. Section 2 provides an overview of damage evaluation methodologies. Section 3 details the proposed approach, including dataset preparation, preprocessing steps, and training strategies. The results are presented in Section 4, while Section 5 describes a possible framework for a vision-based assessment of buildings' seismic residual capacity in post-earthquake scenarios. Finally, Section 6 discusses conclusions and potential directions for future research.

2 BACKGROUND AND DAMAGE LEVEL CRITERIA

A comprehensive analysis of methodologies employed at both national and international levels, with a particular focus on seismically active regions, has been conducted to establish appropriate classification criteria for assessing earthquake-induced damage and the residual structural capacity of existing buildings.

During the late 1990s, the United States saw the introduction of a quantitative framework by the Federal Emergency Management Agency (FEMA) for evaluating the residual seismic performance of earthquake-affected reinforced concrete (RC) and masonry structures (FEMA 306, [12]). This methodology relies on capacity reduction factors applied to damaged structural elements, modifying the plastic hinges' response in terms of stiffness, strength, and ductility. These reductions are associated to observed damage in post-earthquake visual inspections, requiring experts to categorize damage severity into four distinct levels: Insignificant, Slight, Moderate, and Extreme. Additionally, to support the selection of the capacity reduction factors, FEMA 306 provides both descriptive guidelines and graphical representations of typical crack patterns corresponding to various damage levels and structural components.

A conceptually comparable approach has been adopted in Japan, where the Japan Building Disaster Prevention Association (JBDPA) developed a classification system outlined in its official guidelines (summarized in English in [13]). This system categorizes structural damage into five classes, from "I" to "V", with the latter indicating the most severe deterioration.

Following the 1999 Kocaeli and Duzce earthquakes, Turkey introduced a mandatory seismic insurance system, leading to the development of a rapid post-earthquake damage assessment protocol by the Turkish Catastrophe Insurance Pool (TCIP) [14]. This framework classifies damage at both the building and component levels. The categorization of overall building damage includes six levels: (i) No Damage, (ii) Slight Damage, (iii) Moderate Damage, (iv) Severe Damage, (v) Urgent Demolition Required, and (vi) Collapse. Similarly, damage to structural components—both vertical and horizontal—is classified into five categories, ranging from "Type O" (minimal damage) to "Type D" (extensive structural failure).

In Europe, a significant milestone was the development of the European Macroseismic Scale (EMS-98) [15], which has become a widely accepted reference within the scientific commu-

nity. This classification delineates five progressive damage levels: D1 (negligible to slight), D2 (moderate), D3 (substantial to heavy), D4 (very heavy), and D5 (collapse). The EMS-98 framework also includes detailed illustrations and reference images to facilitate the damage classification of various structural typologies.

Italy has adopted a specialized approach through the AeDES ("Agibilità e Danno nell'Emergenza Sismica") form, developed by the Italian National Seismic Protection Group (GNDT) [2]. Designed as an essential tool for post-earthquake damage assessment, the AeDES form enables certified engineers to systematically evaluate building conditions, propose immediate safety measures, and determine usability following seismic events [12]. Figure 1 illustrates the damage classification system employed in the AeDES form, which is derived from the EMS-98 scale but simplifies the framework by merging damage levels D2 with D3 and D4 with D5, as summarized in Table 1.

Structural component Pre-existing damage	DAMAGE [1]									
	D4-D5 Very Heavy			D2-D3 Medium-Severe			D1 Light			Null
	$> 2/3$	$1/3 - 2/3$	$< 1/3$	$> 2/3$	$1/3 - 2/3$	$< 1/3$	$> 2/3$	$1/3 - 2/3$	$< 1/3$	
	A	B	C	D	E	F	G	H	I	L
1 Vertical structures	□	□	□	□	□	□	□	□	□	○
2 Floors	□	□	□	□	□	□	□	□	□	○
3 Stairs	□	□	□	□	□	□	□	□	□	○
4 Roof	□	□	□	□	□	□	□	□	□	○
5 Infills and partitions	□	□	□	□	□	□	□	□	□	○
6 Pre-existing damage	□	□	□	□	□	□	□	□	□	○

Figure 1: The (Italian) AeDES form: damage levels for each structural element.

Damage level	Description
D1 Negligible to slight damage	Beams: crack width $\leq$ 1 mm (not vertical) Columns or walls: crack width $\leq$ 0.5 mm (not vertical) Infill walls: diagonal cracks $\leq$ 1 mm (up to 2 mm if at the frame interface)
D2–D3 Moderate/Substantial to heavy damage	Beams: cracks = 4–5 mm Columns: cracks = 2–3 mm Imperceptible leaning Incipient buckling of reinforcing bars Concrete cover spalling Infill walls: diagonal cracks up to a few mm, evident crushing at the corners in contact
D4–D5 Very heavy/Collapse damage	Collapse or inclination more than 1% Beams: cracks $>$ 5 mm Columns: cracks $>$ 3 mm Buckling of reinforcing bars

Table 1: AeDES damage level classification and description of the observed damage to structural and non-structural components.

In this research, the AeDES classification is employed for the development of an automated damage evaluation algorithm. Aligning the proposed approach with AeDES criteria guaran-

tees that the damage categorization adheres to established Italian standards, ensuring a region-specific and precise assessment.

3 METHODOLOGY

3.1 Dataset

The dataset employed in this research consists of a diverse collection of open-access RGB images, resized to a uniform resolution of 224×224 pixels. These images originate from multiple high-resolution repositories, ensuring a comprehensive representation of structural damage. The primary sources include the PEER Hub ImageNet (Φ-Net) [5, 16], which aggregates data from platforms such as NISEE (National Information Service for Earthquake Engineering), NEEShub (Network for Earthquake Engineering Simulation), EERI (Earthquake Engineering Research Institute), and Google Images. Additional contributions come from the INGV Macroseismic Photographic Database (DFM) [11], reports from ReLUISS (Network of Seismic and Structural Engineering Laboratories), as well as images obtained from academic literature, scientific studies, and publicly available post-earthquake surveys, including documentation from the 2023 Turkey Earthquake.

Database source	Number of images
PEER Image Net Challenge dataset (NISEE, NEEShub, EERI, Google images)	2625
INGV DFM - Database Fotografico Macrosismico	1537
RELUISS reports	245
Turkey Earthquake 2023 survey	194
Scientific related papers and articles	104
Total	4705

Table 2: Number of images in the considered data sources.

As detailed in Table 2, a total of 4,705 RGB images were gathered, categorized into four damage levels: "Heavy" (H), "Moderate" (M), "Slight" (S), and "Undamaged" (U). The dataset construction involved the identification of critical structural components displaying seismic damage indicators such as cracks, fractures, and deformations. Structural elements, including beams, columns, and walls. Images were then resized and cropped to maintain a clear focus on the damaged sections, enhancing clarity for model training. Table 3 provides an overview of the number of images corresponding to different structural components and building types included in the dataset.

To ensure label accuracy, all images underwent a rigorous relabeling process conducted by expert structural engineers from our institution. Given their expertise in post-earthquake damage assessment, the engineers carefully analyzed each image and assigned labels reflecting the severity of structural damage. This procedure aligned the dataset with the damage classification criteria defined in the AeDES manual, ensuring consistency with established post-earthquake assessment standards. To this aim, an image labeling script was developed in a Google Colab environment and deployed on a shared Google Drive workspace through the Graphical User Interface (GUI) shown in Figure 2. This approach enabled all members of the research team to collaboratively label the dataset remotely, streamlining the annotation process while maintaining uniform classification standards.

Element type	Count	Building type	Count
Wall	1792	Reinforced Concrete	2606
Column	1180	Masonry	1852
Beam	394	Steel	19
Non structural	1178	Other	228
Stairs	22		
Other	139		

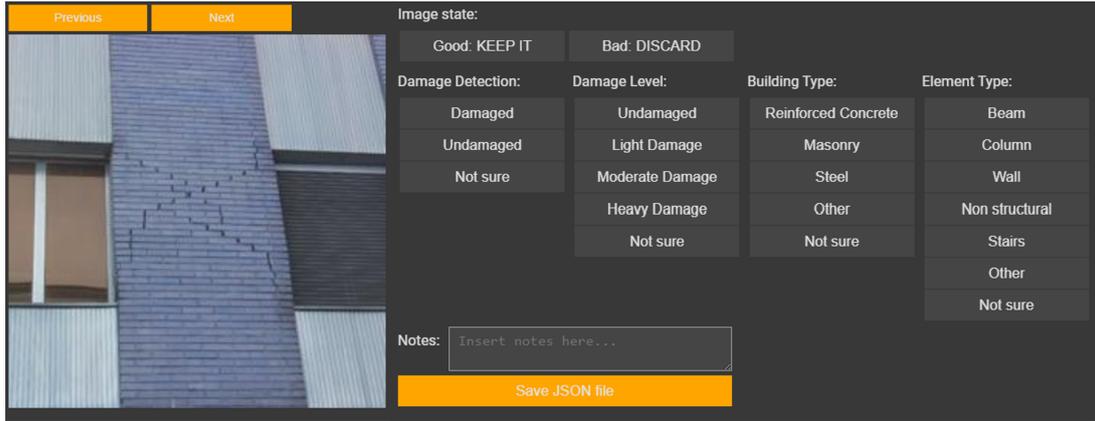Table 3: Count of element and building types in the dataset.

Figure 2: The labeling dataset GUI.

The dataset was split into training and test subsets, with an initial allocation of 3,505 images for training and 1,200 for testing, as presented in Table 4. The test set is well balanced, consisting of 300 images per damage category, while the training set retains an unbalanced distribution reflective of real-world scenarios.

	Damage level				
Set	**None**	**Slight**	**Moderate**	**Heavy**	**Total**
Training	552	732	1014	1207	3505
Test	300	300	300	300	1200
Total	852	1032	1314	1507	4705

Table 4: Distribution of damage levels in the dataset and its split into training and test sets.

3.2 Data Augmentation

To improve model generalization and increase data diversity, extensive augmentation techniques were applied. These included horizontal flipping, slight rotational transformations (up to 10 degrees), and Gaussian noise injection.

Several training sets were generated using different data augmentation strategies. The primary goal was to create a more balanced dataset across damage classes while increasing the total number of images and their variability.

Initially, the model was trained on the original dataset, which contained 3,505 images and

exhibited class imbalance. To address this, oversampling was applied in Training set 2, where underrepresented classes were randomly duplicated to improve balance, leading to a total of 4,828 images.

Further augmentation was introduced in subsequent training sets by incorporating geometric transformations to enhance data diversity. Training set 3 included horizontal flips, doubling the dataset size, while Training set 4 expanded this approach with random rotation transformations up to $10°$ by increasing the number of augmented images, reaching 14,484 samples. Training set 5 further improved variability by adding Gaussian noise with zero mean and standard deviation set to 0.2, simulating real-world image degradation, and producing a total of 19,312 images. Finally, Training set 6 explored perspective transformations to introduce additional distortions that could enhance model robustness.

These augmentations were implemented in Python using NumPy and Albumentations libraries, ensuring an efficient and systematic approach to dataset expansion. The final dataset compositions after augmentation are detailed in Table 5.

Dataset	Oversampling	Transformations	Total
Original training set	—	—	3505
Training set 2	Yes	—	4828
Training set 3	Yes	Horizontal Flip	9656
Training set 4	Yes	Horizontal Flip and $10°$ Rotation	14484
Training set 5	Yes	Horizontal Flip, $10°$ Rotation, Adding Noise	19312
Training set 6	Yes	Horizontal Flip, $10°$ Rotation, Perspective transf.	19312

Table 5: Different dataset compositions after augmentation.

Overall, the data augmentation expanded the training set size yielding a final dataset of 19,312 augmented training images. The test set remained unchanged to preserve evaluation consistency. This augmentation strategy was initially validated through a baseline CNN model [4], demonstrating optimal performance when trained on the expanded dataset. This extensive and well-structured collection provided a solid foundation for training and testing the CNNs used in this study. The integration of multiple data sources, coupled with augmentation and resampling techniques, enhances the model's capacity to generalize across diverse post-earthquake structural damage scenarios, ultimately improving the accuracy and reliability of automated seismic damage assessment.

3.3 CNN Architectures and Training Strategy

In this paper, we implemented twelve different and well-known CNN models. Among all these architectures, the best performing was the VGG16 model [17]. This is a CNN architecture consisting of 16 layers, including 13 convolutional layers and 3 fully connected layers. VGG16 is renowned for its simplicity and effectiveness: in fact, despite its simplicity compared to more recent architectures, it usually provides excellent performance.

To improve the effectiveness and reliability of the considered models, we exploited the transfer learning approach by using the pre-trained version of the VGG16 network [18]. The transfer learning is a technique in deep learning where a pre-trained (usually on the ImageNet) model is used as a starting point for a new related task, hence allowing to leverage the knowledge already gained by the pre-trained model.

The fine-tuning training procedure integrates supervised learning strategies with a structured validation approach. To effectively monitor the performance, a 10% of the training dataset

was set aside for its validation. The model was fine-tuned using mini-batches of 16 samples per iteration, optimizing the categorical cross-entropy loss function. The Adam optimizer was employed, with a learning rate scheduler ensuring efficient convergence. Specifically, the ReduceLROnPlateau strategy was applied, decreasing the learning rate by a factor of 2 whenever validation loss failed to improve over five consecutive epochs. The initial learning rate was configured at 10^{-5}, while all other hyperparameters remained at their default values.

Simulations have been performed on a standard office PC equipped with an Intel 14-th generation i9-14900KF CPU and 32 GB of RAM memory, and a Geforce RTX 3050 8 GB GPU. The data augmentation and classification pipeline are implemented in Python.

4 RESULTS AND DISCUSSION

Numerical results are evaluated by using the overall accuracy and the well known per-class precision, recall, and F1-score metrics, and their weighted averages evaluated on the predictions of the same test set (1200 images, see table 4).

The application of data augmentation techniques significantly improved the performance of the VGG16 model in classifying earthquake-induced structural damage. Table 6 shows the classification metrics obtained on the different training sets used in this work and described in Table 5. Initially, the model was trained on the original dataset, which exhibited an imbalance across damage classes. To address this, oversampling was employed by randomly duplicating images from underrepresented classes, effectively balancing the dataset (Training set 2). This enhancement strategy led to noticeable improvements in classification accuracy and generalization, as detailed in the results table 6. Further enhancements were achieved by incorporating geometric transformations into the training process. By progressively augmenting the dataset with modified copies of the original images (Training set 3 and Training set 4), the model's robustness to variations in real-world damage scenarios was increased. The most substantial performance gain was observed when the training set was expanded using a combination of horizontal flipping, 10-degree rotation, and Gaussian noise addition (Training set 5). These transformations introduced additional diversity while preserving structural features relevant to damage classification. The last line in Table 6 shows that the use of perspective transformations to introduce additional distortions (Training set 6) is not effective for the classification accuracy.

From a careful comparison of the rows in Table 6, we can argue that the best augmented training set is that exploiting horizontal flips, random rotation transformations up to $10°$, and an injection of Gaussian noise (Training set 5).

Overall, the VGG16 is able to obtain an overall accuracy of 89.33% on the Training set 5, showing an improvement of more than 3% with respect to the original training set. This clearly demonstrates the effectiveness of the implemented data augmentation approach. The VGG16 model also provides similar values for precision, recall, and F1-score, once again confirming its effectiveness.

The confusion matrices obtained by the VGG16 architecture on the original training set, the Training set 2, and the Training set 5 (the best performing), are shown in the three pictures in Figure 3. This figure clearly shows the advantages of the implemented data augmentation techniques, since the number of true positive entries increase by passing from the original training set to the Training set 2, and to Training set 5. Interestingly enough, the use of training set 5 decreases the number of images that are incorrectly classified as belonging to the "Moderate" class. In fact, by using the original training set 44 images with no damages and 65 images with a slight damage have been classified as a "Moderate" level, while using the Training set 5, these numbers reduce to 18 and 24, respectively. This fact is very important for the implementation

Training set	Accuracy	Precision	Recall	F1-score
Original training set	86.25	88.33	86.25	86.45
Training set 2	87.75	88.59	87.75	87.78
Training set 3	87.92	89.05	87.92	87.97
Training set 4	88.75	89.66	88.75	88.78
Training set 5	**89.33**	**89.53**	**89.33**	**89.37**
Training set 6	87.58	88.99	87.58	87.77

Table 6: Comparisons of VGG16 test set classification metrics (in percentage) using different data augmentation approaches.

of a robust and reliable tool for automatic classification of the damage level of buildings. However, we have to underline that some difficulties in distinguish between "Slight" and "Moderate" damage levels still persists and will be addressed in future works. The only drawback of the proposed approach is that, despite the general improvement of performance, the true positive entries for the "Moderate" class slightly decreases by using the Training set 5 with a consequent increase of the "Slight" class.
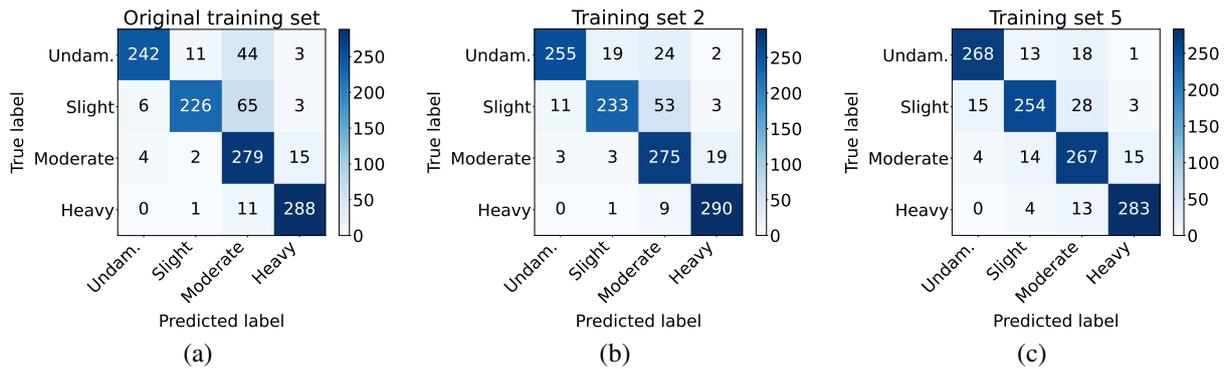

Figure 3: Test set confusion matrices related to VGG16 model trained using different datasets: the original training set (a), Training set 2 (b), and Training set 5 (c).

Comparisons with respect to the other eleven state-of-the-art approaches considered in this work, and evaluated on the Training set 5, are shown in Table 7. Specifically, we compare the results of the VGG16 model with well-known and famous architectures, such as AlexNet, GoogLeNet DenseNet201, MobilNet-V2, EfficientNet-B0, ResNet, XceptionNet, InceptionV3, and the CNN network introduced in [4] considered as a baseline. These architectures have been widely used in literature about crack detection and SHM tasks [6, 9, 18, 19].

5 SEISMIC RESIDUAL CAPACITY ASSESSMENT FRAMEWORK

The proposed CNN-based tool for visual-based damage classification can be used to support a detailed assessment of buildings' seismic residual capacity in post-earthquake scenarios. The information on the observed earthquake-related damage to structural components can be linked to capacity reduction factors for their plastic hinges' response, following state-of-the-art approaches in literature (e.g., FEMA 306, 1998). This way, it is possible to assess the seismic performance of the structure in its post-earthquake (i.e., damaged) configuration and evaluate its seismic residual capacity. The potential flowchart for vision-based seismic residual capacity

Architecture	Accuracy	Precision	Recall	F1-score
Baseline CNN	82.33	83.80	82.30	83.10
AlexNet	82.75	83.96	82.75	82.35
MobileNet-V2	83.25	84.80	83.30	84.00
EfficientNet-B0	84.42	86.80	84.40	84.60
DenseNet201	84.66	86.20	84.70	85.50
GoogLeNet	85.50	86.50	85.50	85.40
ResNet50	87.50	89.14	87.50	87.69
ResNet101	87.67	89.40	87.67	87.79
XceptionNet	87.50	88.82	87.50	87.58
InceptionV3	87.83	88.61	87.83	87.89
VGG19	88.75	89.57	88.75	88.90
VGG16	**89.33**	**89.53**	**89.33**	**89.37**

Table 7: Comparative results (in percentage) with respect to other state-of-the-art architectures evaluated on the same test set after training the models on Training set 5.

assessment of buildings is schematically illustrated in Figure 4, in line also with recent research in literature [20, 21]. Each step is briefly discussed below referring to an RC frame structure.
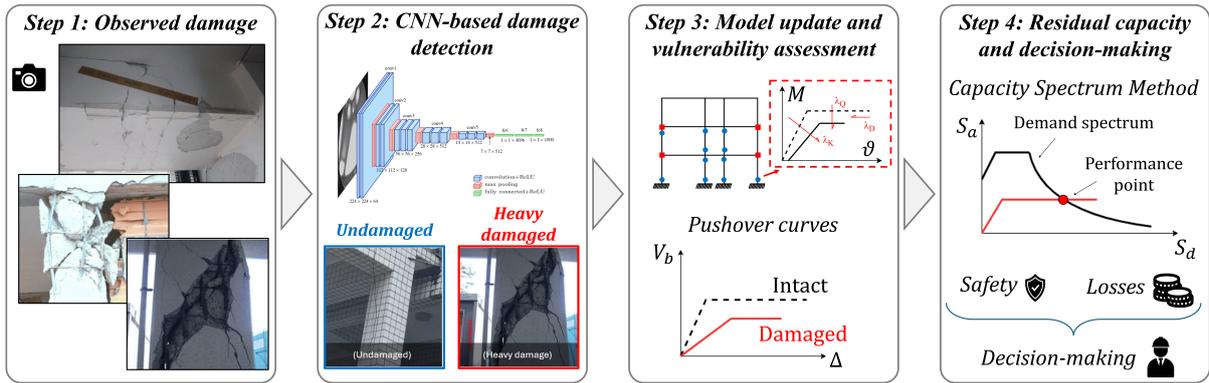

Figure 4: Flowchart for vision-based residual capacity assessment of earthquake-damaged buildings.

Firstly (Step 1), the information on the observed post-earthquake damage is collected (photographic surveys). In the emergency phase, this data can be obtained through both rapid post-earthquake surveys using drones (mainly for the damage observable from the outside) and in-situ inspection (outside and, in the absence of safety issues, inside the damaged building). These data are then processed using the proposed CNN-based algorithm (Step 2), thus obtaining as output information on the member typology (e.g., beams, columns, walls, non-structural components) and the level of damage (from "insignificant" to "heavy damage"). For the correct implementation of the framework, for each analyzed building, it is also fundamental to evaluate the number of damaged structural components and their location with respect to the structural skeleton (e.g., damage to a base column can be more critical than the same damage to a column in the last story). Results provided by the CNN-based tool are then used to update the structural model, in order to define its damaged configuration (Step 3).

According to state-of-the-art procedures, the seismic residual capacity of earthquake-damaged buildings can be assessed through simplified pushover-based methods employing capacity reduction factors for plastic hinges' response of damaged components. Among others, in the

FEMA 306 procedure, reduction factors are used to modify the response in terms of stiffness (λ_K), strength (λ_Q), and ductility (λ_D). These λ-factors can be defined as a function of the observed damage during visual inspections; to this end, the FEMA 306 report includes a schematic illustration of the crack patterns for different components, behavior modes and damage levels.

In line with this approach, research effort has been devoted in the past years to deriving plastic hinges' modification factors for damaged RC components through either experimental data or numerical simulations [22, 23, 24, 25, 26, 27]. Past research also investigated the possibility of linking suitable modification factors to observed damage [22]. Moreover, several frameworks to assess seismic residual capacity for damaged structures through a nonlinear static approach have been proposed, either employing numerical simulations [28, 29, 30, 31, 32] or simplified analytical-mechanical procedures [33, 34, 35].

If the FEMA 306 approach is adopted, information on the typology of components and the damage level can be used to select suitable reduction factors to update the plastic hinge response of damaged structural components. This information allows evaluating the structure's force-displacement capacity curve (pushover curve) in its damaged configuration. The force-displacement capacity curve of the structure in its "damaged" configuration is expected to show a lower (or at least the same) seismic performance than the "intact" (i.e., as-built) configuration. Specifically, a reduction in terms of stiffness, strength, and ductility can be obtained also for the building-level response. Finally (Step 4), the pushover curve can be used to perform seismic response analysis via spectrum-based approache, e.g., the Capacity Spectrum Method (CSM) [36] or the N-2 method [37]. Moreover, safety evaluation and loss assessment can be carried out through simplified pushover-based procedures, e.g., the approach described in the Italian "Seismic-bonus" guidelines [38]. The results can be used to support the decision-making in re-occupancy as well as repair versus demolition and reconstruction. Furthermore, if a simplified analytical/mechanical procedure – rather than numerical software-based simulations – is employed [34, 35], a rapid assessment tool for the emergency phase can be developed, allowing for a visual-based (yet mechanically informed) safety evaluation of earthquake-damaged buildings since from the early emergency phases.

6 CONCLUSIONS

In conclusion, this study demonstrates the significant potential of CNN-based deep learning techniques for automating post-earthquake structural damage assessments. Our implementation of the VGG16 model—enhanced through extensive data augmentation, transfer learning, and dataset balancing—yields an impressive overall accuracy exceeding 89% and robust F1-scores, particularly for damage classes well represented in the dataset. These results not only validate the use of deep learning as a reliable alternative to traditional visual inspections but also highlight its capacity for rapid, consistent, and objective evaluation in critical post-disaster scenarios.

Despite these promising outcomes, our experiments have also revealed persistent challenges, notably in differentiating between slight and moderate damage levels. This limitation suggests that further refinement is needed, potentially through the integration of more sophisticated architectures or additional data modalities that capture subtle variations in damage characteristics.

Future research will focus on several key areas: expanding the dataset with images from recent seismic events to enhance model generalization; exploring ensemble techniques and multi-scale feature extraction to better capture nuanced damage patterns; and integrating the CNN-based damage classification with seismic residual capacity assessment frameworks to provide a more comprehensive tool for post-earthquake evaluation and decision-making. Ultimately, the

advancement of such automated systems holds the promise of significantly improving emergency response and long-term structural safety in seismically active regions.

REFERENCES

[1] Applied Technology Council, "Procedures for post-earthquake safety evaluation of buildings," ATC 20, Redwood City, California, US, Tech. Rep., 1989.

[2] C. Baggio, A. Bernardini, R. Colozza, L. Corazza, M. Della Bella, G. Di Pasquale, and G. Zuccaro, "Field manual for post-earthquake damage and safety assessment and short term countermeasures (AeDES)," European Commission–Joint Research Centre–Institute for the Protection and Security of the Citizen, Tech. Rep. EUR, 22868, 2007.

[3] S. Pampanin, "Simplified analytical/mechanical procedure for post-earthquake safety evaluation and loss assessment of buildings," in *Advances in Assessment and Modeling of Earthquake Loss*, S. Akkar, A. Ilki, C. Goksu, and M. Erdik, Eds. Springer International Publishing, 2021, pp. 3–25.

[4] Y.-J. Cha, W. Choi, and O. Büyüköztürk, "Deep learning-based crack damage detection using convolutional neural networks," *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 5, pp. 361–378, 2017.

[5] Y. Gao and K. M. Mosalam, "PEER Hub ImageNet: A large-scale multiattribute benchmark data set of structural images," *Journal of Structural Engineering*, vol. 146, no. 10, p. 04020198, 2020.

[6] M. Azimi, A. Eslamlou, and G. Pekcan, "Data-driven structural health monitoring and damage detection through deep learning: State-of-the-art review," *Sensors*, vol. 20, p. 2778, 2020.

[7] Y. Gao and K. M. Mosalam, "Deep transfer learning for image-based structural damage recognition," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, pp. 748–768, 2018.

[8] K. Gopalakrishnan, S. K. Khaitan, A. Choudhary, and A. Agrawal, "Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection," *Construction and Building Materials*, vol. 157, pp. 322–330, 2017.

[9] S. Saquella, M. Scarpiniti, G. Laneve, and A. Uncini, "Benchmarking of CNN architectures for post-earthquake damage assessment," in *32nd Italian Workshop on Neural Networks (WIRN 2024)*, Vietri sul Mare, Italy, 2024.

[10] P. D. Ogunjinmi, S.-S. Park, B. Kim, and D.-E. Lee, "Rapid post-earthquake structural damage assessment using convolutional neural networks and transfer learning," *Sensors*, vol. 22, p. 3471, 2022.

[11] Istituto Nazionale di Geofisica e Vulcanologia (INGV), "Database fotografico macrosismico (DFM)," 2025, accessed: 2025-01-20. [Online]. Available: https://macrosismico.ingv.it

[12] FEMA, "Evaluation of earthquake damaged concrete and masonry wall buildings – Basic procedures manual," Federal Emergency Management Agency, Washington DC, US, Tech. Rep., 1998.

[13] M. Maeda, H. Al-Washali, and K. Matsukawa, "An overview of post-earthquake damage and residual capacity evaluation for reinforced concrete buildings in Japan," in *7th International Conference on Computational Methods in Structural Dynamics and Earthquake Engineering Methods in Structural Dynamics and Earthquake Engineering (COMPDYN 2019)*, Crete, Greece, 2019, pp. 930–943.

[14] A. Ilki, O. F. Halici, M. Comert, and C. Demir, "The modified post-earthquake damage assessment methodology for TCIP (TCIP-DAM-2020)," in *Advances in Assessment and Modeling of Earthquake Loss*, S. Akkar, A. Ilki, C. Goksu, and M. Erdik, Eds. Springer International Publishing, 2021, pp. 85–107.

[15] G. Grunthal, *European Macroseismic Scale*. European Seismological Commission. Conseil de l'Europe, Cahiers du Centre Européen de Géodynamique et de Séismologie: Luxembourg, 1998, vol. 15.

[16] Y. Gao and K. M. Mosalam, "PEER Hub ImageNet (Φ–Net): A large-scale multi-attribute benchmark dataset of structural images," Pacific Earthquake Engineering Research Center (PEER), Tech. Rep. 2019/07, 2019.

[17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, 2015.

[18] A. Shamila Ebenezer, S. Deepa Kanmani, V. Sheela, K. Ramalakshmi, V. Chandran, M. G. Sumithra, B. Elakkiya, and B. Murugesan, "Identification of civil infrastructure damage using ensemble transfer learning model," *Advances in Civil Engineering*, vol. 2021, no. 1, p. 5589688, 2021.

[19] Z. Hong, H. Zhong, H. Pan, J. Liu, R. Zhou, Y. Zhang, Y. Han, J. Wang, S. Yang, and C. Zhong, "Classification of building damage using a novel convolutional neural network based on post-disaster aerial images," *Sensors*, vol. 22, no. 15, p. 5920, 2022.

[20] X. Ji, Y. Zhuang, Z. Miao, and Y. Cheng, "Vision-based seismic damage detection and residual capacity assessment for an rc shaking table test structure," *Earthquake Engineering & Structural Dynamics*, vol. 52, no. 3, pp. 806–827, March 2023.

[21] Z. Miao, X. Ji, M. Wu, and X. Gao, "Deep learning-based evaluation for mechanical property degradation of seismically damaged rc columns," *Earthquake Engineering & Structural Dynamics*, vol. 52, no. 8, pp. 2498–2519, July 2023, special Issue: AI and data-driven methods in earthquake engineering – (Part 1).

[22] M. Di Ludovico, M. Polese, M. G. d'Aragona, A. Prota, and G. Manfredi, "A proposal for plastic hinges modification factors for damaged RC columns," *Engineering Structures*, vol. 51, pp. 99–112, 2013.

[23] K. Marder, K. J. Elwood, C. J. Motter, and G. C. Clifton, "Post-earthquake assessment of moderately damaged reinforced concrete plastic hinges," *Earthquake Spectra*, vol. 36, no. 1, pp. 299–321, 2020.

[24] C.-K. Chiu, H.-F. Sung, and T.-C. Chiou, "Quantification of the reduction factors of seismic capacity for damaged RC column members using the experiment database," *Earthquake Engineering & Structural Dynamics*, vol. 50, no. 3, pp. 756–776, 2021.

[25] C. Ceccarelli, S. Bianchi, L. Pedone, and S. Pampanin, "Numerical investigations on the residual capacity and economic losses of earthquake-damaged reinforced concrete wall structures," in *Proceedings of 8th International Conference on Computational Methods in Structural Dynamics and Earthquake Engineering (COMPDYN 2021)*, 2021, pp. 1242–1260.

[26] A. Rossi, C. Del Vecchio, and S. Pampanin, "Influence of earthquake damage and repair interventions on expected annual losses of reinforced concrete wall buildings," *Earthquake Spectra*, vol. 38, no. 3, pp. 2026–2060, 2022.

[27] G. Nicolò, L. Pedone, M. Matteoni, and S. Pampanin, "Numerical investigation on the seismic residual capacity of reinforced concrete beams' plastic hinges," in *2023 Symposium on Concrete and Concrete Structures*, 2023.

[28] P. Bazzurro, C. A. Cornell, C. Menun, and M. Motahari, "Guidelines for seismic assessment of damaged buildings," in *Proceedings of the 13th World Conference on Earthquake Engineering*, Vancouver, B.C., Canada, August 2004.

[29] M. Polese, M. Di Ludovico, A. Prota, and G. Manfredi, "Residual capacity of earthquake damaged buildings," in *Proceedings of the 15th World Conference on Earthquake Engineering*, Lisboa, Portugal, 2012.

[30] A. Cuevas and S. Pampanin, "Accounting for residual capacity of reinforced concrete plastic hinges: Current practice and proposed framework," in *Proceedings of the New Zealand Society of Earthquake Engineering Conference*, Auckland, New Zealand, March 2014.

[31] S. Pampanin, A. Cuevas, M. Kral, G. Loporcaro, A. Scott, and A. Malek, "Residual capacity and repairing options for reinforced concrete buildings," Natural Hazard Research Platform, Tech. Rep. Contract 2012-UOC-02-NHRP, 2015.

[32] L. Pedone, R. Gentile, C. Galasso, and S. Pampanin, "Nonlinear static procedures for state-dependent seismic fragility analysis of reinforced concrete buildings," in *fib Symposium*, vol. 56. fib: The International Federation for Structural Concrete, November 2021, pp. 47–54.

[33] M. Polese, M. Marcolini, G. Zuccaro, and F. Cacace, "Mechanism based assessment of damage-dependent fragility curves for rc building classes," *Bulletin of Earthquake Engineering*, vol. 13, pp. 1323–1345, 2015.

[34] S. Pampanin, "Simplified analytical/mechanical procedure for post-earthquake safety evaluation and loss assessment of buildings," in *Springer Tracts in Civil Engineering*. New York: Springer, 2021.

[35] M. Matteoni, L. Pedone, S. D'Amore, and S. Pampanin, "Simplified analytical/mechanical procedure for the residual capacity assessment of earthquake-damaged reinforced concrete

frames," in *9th International Conference on Computational Methods in Structural Dynamics and Earthquake Engineering (COMPDYN 2023)*, 2023.

[36] ATC, "Seismic evaluation and retrofit of concrete buildings," Applied Technology Council, Redwood City, CA, USA, Tech. Rep. ATC-40, 1996.

[37] P. Fajfar, "A nonlinear analysis method for performance-based seismic design," *Earthquake Spectra*, vol. 16, pp. 573–592, 2000.

[38] E. Cosenza, C. Del Vecchio, M. Di Ludovico, M. Dolce, C. Moroni, A. Prota, and E. Renzi, "The italian guidelines for seismic risk classification of constructions: Technical principles and validation," *Bulletin of Earthquake Engineering*, vol. 16, pp. 5905–5935, 2018.