



SAPIENZA
UNIVERSITÀ DI ROMA

Vision-based Deep Learning Approaches for Post-Earthquake Building Damage Assessment

Scuola di Dottorato in Scienze e Tecnologie per l'Innovazione Industriale
Facoltà di Ingegneria Civile e Industriale
Department of Astronautical, Electrical and Energy Engineering - DIAEE
PhD Program in ENERGY and ENVIRONMENT (XXXVIII cycle)

Simone Saquella

ID number 1474476

Advisor

Prof. Giovanni Laneve

Co-Advisor

Prof. Michele Scarpiniti

Academic Year 2024/2025

Vision-based Deep Learning Approaches for Post-Earthquake Building Damage Assessment
PhD thesis. Sapienza University of Rome

© 2025 Simone Saquella. All rights reserved

This thesis has been typeset by $\text{\LaTeX}$ and the Sapthesis class.

Author's email: simone.saquella@uniroma1.it

*Dedicated to my wife,
Claudia.*

Abstract

This doctoral thesis tackles a challenge in earthquake engineering: the timely and reliable assessment of damage to buildings after an earthquake. Traditional evaluations often rely on visual inspections conducted by structural engineers using standardized forms, such as the Italian *AeDES* protocol. While these methods are necessary, they tend to be time-consuming, subjective, and hard to apply during large seismic events. Driven by these challenges, this thesis examines how vision-based Artificial Intelligence (AI), specifically Convolutional Neural Networks (CNNs), can improve the post-earthquake damage assessment process. The main goals of this research are twofold: (i) to evaluate the performance of various modern CNN architectures for seismic damage classification in buildings, with a focus on the VGG16 model, and (ii) to create and validate a new dataset designed for earthquake damage classification, enhancing existing collections.

The contribution of this new proposed dataset is a key part of this work. Previous datasets, like PEER Φ -Net, ReLUIS, and the INGV photographic database, provided useful starting points but faced issues like class imbalances, inconsistent labeling criteria, and discrepancies across different sources. To address these problems, a new dataset was created. Structural engineers re-annotated it following the *AeDES* guidelines and expanded it through a systematic process that included geometric transformations and oversampling. The final balanced dataset contains nearly 20,000 labeled samples, covering four damage states (None, Slight, Moderate, Heavy) across various building types and structural elements. This dataset not only supports the experiments in this thesis but also serves as a valuable resource for future research in automated post-disaster assessments.

On the methodological side, the thesis has first gone through the problem of CNN efficacy in a comparative study of different well-known architectures like AlexNet, DenseNet, ResNet, EfficientNet-B0, and particularly VGG16. Results have corroborated the strength of CNN-based methods in the identification of earthquake-caused damage to be stable, with accuracies always exceeding 80% and the best-performing setups getting close to 90% on the curated dataset. VGG16, even though it is one of the earlier deep architectures, turns out to be the most powerful model when the correct data augmentation and transfer learning strategies are applied. What these outcomes emphasize is that in the case of relatively small and highly specialized datasets as seismic imagery, the choice of a less complex deep learning model along with the development of a carefully curated dataset can give better results than the use of more complex and deeper networks.

Moreover, the thesis goes beyond the CNN benchmarking to consider the implementation of more sophisticated data-fusion strategies for the enrichment of the classification process. To facilitate the extraction of high-frequency coefficients representing edges and the local texture details of the targeted images the Discrete Wavelet Transform (DWT) was introduced. These coefficients were then mixed with RGB images through early and intermediate fusion schemes. This approach proved to be the most promising in terms of accuracy with performance improved almost to 87% as well as better recognition of faint damage levels such as "Moderate" that usually are barely differentiable. This proves to be the perfect example of how spatial and spectral representations can be successfully combined for damage recognition in structures using computer-vision.

In addition, it was explored the potential of Transformer-based architectures: in particular, four variants of the Vision Transformer (ViT) were implemented and benchmarked, reaching

good performance with both accuracy and F1-score peaking at 88%, thereby surpassing most CNN-based approaches on the curated dataset.

Another consistent portion of this thesis was marked by the incorporation of AI-based classification alongside seismic capacity assessment. The conceptual framework was only briefly mentioned and later explained with a small case study: damage factors derived from CNN were converted into mechanical reduction factors of the stiffness, strength and ductility properties of the structure according to the guidelines of FEMA 306, and were thus usable for nonlinear structural analyses. As a result, the updated pushover curves for buildings with different damage levels were obtained, which enabled the residual seismic capacity to be estimated. Experimental situations involving earthquakes were used to test the performance of the method. The results indicate that automated visual classification can provide valuable input for structural performance evaluation. Hence, this study moves on to the next level of decision making in the field of structural engineering, which is the evaluation of safety after the quake by employing a quantitative method.

Overall, the findings of this research underline that the combination of AI-based vision techniques and structural mechanics can significantly accelerate, standardize, and objectify post-earthquake assessment workflows. The results also show that the construction of a tailored and rigorously annotated dataset is at least as important as model selection, confirming that data quality is a primary driver of performance in AI-based earthquake engineering. This thesis, therefore, contributes both a methodological framework and a novel dataset, laying the groundwork for future developments in AI-augmented seismic risk management and digital-twin integration.

Contents

1	Objectives and Background of the Research	1
1.1	Post-earthquake damage assessment: Introduction and motivation of the research	1
1.2	Innovations Proposed in This Research: CNN Architectures and Tailored Approaches for RC and Masonry Buildings	2
1.3	Context of the research: The CN1–Spoke 5 project, Environmental and Natural Disasters	4
1.4	Specific details of the CN1-Spoke 5 framework and methodology	6
1.5	The Role of Damage Assessment and Structural Capacity Recalibration of Buildings within the Digital Twin Framework	8
2	Determination of Post-Earthquake Damage Levels	11
2.1	Seismic Hazards and Damage Mechanisms in Structures	11
2.2	Overview of Post-Earthquake Damage Classification Standards	12
2.2.1	FEMA 306 (USA)	13
2.2.2	EMS-98 (Europe)	13
2.2.3	JBDPA (Japan)	14
2.2.4	TCIP-DAM (Turkey)	15
2.2.5	NZSEE (New Zealand)	16
2.3	The AeDES Form and National Italian Guidelines	16
2.3.1	AeDES Form Damage Classes	17
2.3.2	Potential Applications of AeDES Data for Global Building Damage Classification via Machine Learning	18
2.4	Selection of Damage Levels for This Research	20
2.5	Final Considerations on the Adopted Damage Levels for This Research . . .	21
3	Computer Vision-Based Damage Classification: An Overview	24
3.1	Introduction: Image-based Damage Classification in Earthquake Engineering	24
3.2	Deep Learning Methods for Image Classification	26
3.3	Deep Neural Networks (DNNs)	28
3.4	Convolutional Neural Networks (CNNs)	30
3.5	Transfer Learning and Data Augmentation	31
3.6	Review of Damage Classification Studies	32
3.7	CNN Architectures Used in Damage Assessment and Concluding Insights .	33

4	Image Datasets for Post-Earthquake Building Damage	36
4.1	PEER Hub ImageNet (Φ -Net) Dataset	37
4.1.1	Dataset Scope and Size	37
4.1.2	Image Sources and Collection	39
4.1.3	Use and Limitations	40
4.2	INGV Macroseismic Photographic Database (DFM)	42
4.3	IDEA: Image Database for Earthquake Damage Annotation	46
4.4	Additional and Ancillary Data Sources	48
4.4.1	ReLUIIS Post-Earthquake Damage Database: L'Aquila Case Study.	48
4.4.2	Data from the Turkey 2023 Earthquake Damage Survey (Pujol et al.).	48
4.4.3	DaDO (Database of Observed Damage)	49
4.4.4	IRMA (Italian Risk MAPs).	49
4.4.5	Mendeley DB Concrete Crack Images for Classification Dataset.	49
4.5	New Dataset Compiled for This Research	50
4.5.1	Motivations and Limitations of Φ -Net	50
4.5.2	Data Sources and Composition of the New Dataset	51
4.5.3	Key Advantages of the Newly Compiled Dataset and Concluding Insights	54
5	Experimental Setup and Benchmarking of Deep Learning Models for Post-Earthquake Damage Classification	56
5.1	Benchmarking of CNN architectures for post-earthquake damage assessment using the Φ -Net dataset	57
5.1.1	Dataset	58
5.1.2	Training Strategy and CNN Models	59
5.1.3	Discussion and Results	60
5.2	Benchmarking of CNN architectures for post-earthquake damage assessment using the proposed dataset	63
5.2.1	Dataset and Data Augmentation	64
5.2.2	Taining strategy And CNN Models	64
5.2.3	Discussion and Results - Selection of the optimal augmented dataset for training	65
5.3	Light wheight CNN and DWT Data-Fusion	68
5.3.1	Efficientnet-B0 architecture	68
5.3.2	Data-fusion techniques using Wavelet transforms	70
5.3.3	Training strategy: Gradual unfreezing	72
5.3.4	Discussion and results: Selection of the best model	73
5.4	A Vision Transformer based Application	76
5.4.1	B-16 and B-32 Models	76
5.4.2	L-16 and L-32 Models	77
5.4.3	Training Strategy, Discussion and Results	77
5.5	Other Approaches	80
5.5.1	Hierarchical CNN-based Classification	80
5.5.2	CAE-based Feature Extraction with Traditional Classifiers	81
5.5.3	Quaternion CNNs (QCNN) and Quaternion-based Architectures	83
5.5.4	Summary of Comparative Findings	83

6	Post-Earthquake Residual Capacity Assessment	84
6.1	Framework Concept and Workflow	85
6.2	Review of previous studies	87
6.3	Perspectives for integration with CNN-based classification	88
6.4	Illustrative Application	90
6.4.1	Case-Study RC Building: Geometry, Detailing, and Materials . . .	90
6.4.2	Simulated Post-Earthquake Scenario via NLTHA	90
6.4.3	CNN-Based Damage Classification and Explicit Uncertainty Modelling	92
6.4.4	Nonlinear Static Analyses: SLAMA vs. Refined Numerical Model .	92
6.4.5	Spectrum-Based Safety Evaluation (IS-V/%NBS)	94
6.4.6	Discussion	94
6.4.7	Limitations and Future Work	94
6.4.8	Implementation Notes (Practical Replicability)	95
6.5	Integration of Damage Classification with Repair Cost Estimation	95
6.6	Conclusive Insights	97
	Conclusions and Future Efforts	99
	Publications Related to This Thesis	102
	Acknowledgements and Thanks	103
	Bibliography	104

Acronyms

%NBS Percent New Building Standard.

Adam Adaptive Moment Estimation.

AE Autoencoder.

AeDES Agibilità e Danno nell’Emergenza Sismica.

AUC Area Under the Curve.

BGI Bagnoli Irpino Seismic Station (Irpinia 1980 record).

CAE Convolutional Autoencoder.

CN1 National Center 1 - Environmental and Natural Disasters.

CNN Convolutional Neural Network.

CSM Capacity Spectrum Method.

DA Data Augmentation.

DaDO Database of Observed Damage.

DFM Database Fotografico Macrosismico.

DL Deep Learning.

DNN Deep Neural Network.

DS Damage State.

DTM Digital Terrain Model.

DWT Discrete Wavelet Transform.

EERI Earthquake Engineering Research Institute.

EMS-98 European Macroseismic Scale 1998.

EUCENTRE European Centre for Training and Research in Earthquake Engineering.

FE Finite Element.

FEMA Federal Emergency Management Agency.

GNDT Gruppo Nazionale per la Difesa dai Terremoti (National Group for the Defence against Earthquakes).

GUI Graphical User Interface.

HPC High-Performance Computing.

IDEA Image Database for Earthquake Damage Annotation.

INGV Istituto Nazionale di Geofisica e Vulcanologia.

IRMA Italian Risk Map Archive.

IS-V Seismic Safety Index (% of New Building Standard).

JBDPA Japan Building Disaster Prevention Association.

MC Monte Carlo.

ML Machine Learning.

MLP Multi-Layer Perceptron.

MRF Moment-Resisting Frame.

Mw Moment magnitude.

NEEShub Network for Earthquake Engineering Simulation Hub.

NISEE National Information Service for Earthquake Engineering.

NLTHA Nonlinear Time History Analysis.

NZSEE New Zealand Society for Earthquake Engineering.

PEDA Post-Earthquake Damage Assessment.

PEER Pacific Earthquake Engineering Research Center.

PGA Peak Ground Acceleration.

PHI-Net PEER Hub ImageNet.

PNRR Piano Nazionale di Ripresa e Resilienza
(National Recovery and Resilience Plan).

PoC Proof of Concept.

R-CNN Recurrent Convolutional Neural Network.

RC Reinforced Concrete.

ReLU Rectified Linear Unit.

ReLUIS Rete dei Laboratori Universitari di Ingegneria Sismica.

RF Random Forest.

SGD Stochastic Gradient Descent.

SHM Structural Health Monitoring.

SLaMA Simple Lateral Mechanism Analysis.

SVM Support Vector Machine.

TCIP Turkish Catastrophe Insurance Pool.

TL Transfer Learning.

ViT Vision Transformer.

VTB Virtual Test Bed.

YOLO You Only Look Once.

Chapter 1

Objectives and Background of the Research

1.1 Post-earthquake damage assessment: Introduction and motivation of the research

Evaluating the destruction of the infrastructure after a seismic event is of primary importance during the emergency management phases, following an earthquake, in which the authorities and the technical staff stand before an overwhelming and very important work to be done in the shortest time possible: to what extent and in which way the buildings, infrastructures, and the components of the urbanized area are damaged. This stage of the crisis management cycle involves not only the emergency actions (intervention, rescue, etc.) directly affected by it, but also it provides the basis for institutional decisions of the structural retrofitting, resilience urban planning, and socio-economic recovery strategies. Hence, post-earthquake damage assessment (PEDA) is at the heart of the structural engineering, civil protection, and data-driven decision-making communities [17]. The traditional method of PEDA, by and large, has always relied heavily on primary field surveys done by teams of local civil and structural engineers, who employ structured evaluation checklists in their inspection works. Among others, the most prominent national and international documents that the technical procedures of such human-based in-field activities refer to as their standards are the AeDES form, used as a reference in Italy, the FEMA 306 guidelines, which have been developed in the United States, and the European Macroseismic Scale (EMS-98), which determines the observed damage levels in terms of structural impairment. All these normative systems provide a wide range of descriptions for damaged features and, at the same time, give some design guidelines for both short-term safety and usability. All these traditional methods, even though technically accurate and well validated, are, however, very labor-demanding that leave room for human error and inconsistency, as well as a critical limitation caused by the number of qualified inspectors and the time that it takes to get to every building affected by the disaster, especially when the earthquake is of a large magnitude and impacts numerous buildings, like in the case of large-scale disasters. One of the main reasons for this change is the growing availability of high-resolution RGB imagery. Besides that, the recent advances in Artificial Intelligence (AI) have the potential to radically change the way the automation is implemented, as well as the scaling of post-earthquake inspections. In particular, deep learning methods and especially Convolutional Neural

Networks (CNNs) [4, 25, 24], have not only been successful but also very convincing in terms of the extent to which they can efficiently automate the classification of structural damage levels using the photographic input. These models get trained on different types of damage (cracking, spalling, buckling, diagonal failure patterns) that are already labeled and skip the process of manual feature engineering, and, in addition, the inspection time can be scaled down from hours to a few seconds per building. In addition, the positions of CNNs make it possible that the integration of data from different sources such as drone imagery, street-level photographs, and archival documentation is feasible. Consequently, the damage scenario can be assessed more quickly, more comprehensively, and in a scalable fashion. Besides the benefits, the application of vision-based AI to PEDAs is still plagued by quite a few problems that remain unresolved. Out of these are the needs for: (i) large, diverse, and accurately labeled datasets of damaged structures under different lighting, angle, and occlusion conditions; (ii) the difficulty in distinguishing between visually similar damage levels (e.g., slight versus moderate cracking); (iii) the transferability of models trained in one geographical or construction context to another (e.g., from RC to masonry structures); and (iv) the integration of damage classification outputs with engineering models for residual capacity assessment and retrofitting decisions. In this regard, this thesis work embraces a pioneering role in pushing the envelope of AI-powered structural damage classification being more robust and reliable through combining CNN-based classification and domain-specific knowledge derived from AeDES methodology. To sum up, the camera-based inspection of damage after an earthquake appears as a key element in the renovation of the seismic risk management system. Such an approach provides the means for quick, uniform, and unbiased assessments, even those performed on a large scale and in real-time. Additionally, the combination of a sensor with Digital Twin systems or HPC-assisted simulation settings could facilitate the creation of advanced decision-support systems, which are considered to be the future of disaster mitigation and sustainable urban planning.

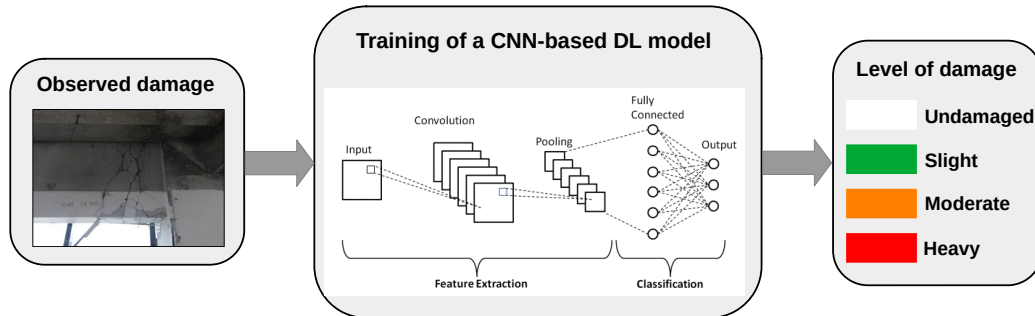


Figure 1.1. Generic workflow for CNN-based damage classification from RGB images. Image from [93].

1.2 Innovations Proposed in This Research: CNN Architectures and Tailored Approaches for RC and Masonry Buildings

The major contribution of this doctorate research will be the generation and corroboration of groundbreaking methods for the automated categorization of the severity of the post-earthquake building damage by RGB pictures. The implementation of the idea mainly

depends on a thorough examination, appraisal, and the generation of photo databases suitable for feeding the deep learning models that are capable of differentiating the structural and non-structural damage features according to national (e.g. AeDES) and international guidelines.

Initially, a thorough investigation was conducted to identify and assess the adequacy of existing publicly available datasets. Because none of the existing repositories are found to meet the optimal criteria in terms of resolution, damage labeling accuracy, building typologies, or architectural diversity, a new dataset was developed in this research.

This was a collaborative work led by DISG (Department of Structural and Geotechnical Engineering) and involving DIET (Department of Information Engineering, Electronics, and Telecommunications), DIAEE (Department of Astronautical, Electrical and Energy Engineering), and the Department of Earth Sciences (Geology) from the Sapienza University of Rome.

The labelling process of building damages will undergo a rigorous methodology that incorporates the expertise of structural engineers, AI image annotator tools, and geological-geotechnical contextualization. Through this joint operation, an original, top-tier RGB image dataset will be generated, with labels newly drawn to represent the pattern and progression of damage in reinforced concrete (RC) and unreinforced masonry (URM) structures. This will allow the creation of model architectures and training pipelines that are explicitly adjusted for different structural materials and building types, thus enabling the identification of damage signatures unique to each one of them.

From the viewpoint of methodology, the work will explore various convolutional neural networks (CNNs) methods. It includes research that spans the whole range from CNN base architectures to contemporary lightweight models. Some main innovations and technologies that were taken into account in this research are as follows:

- **Standard CNN architectures:**
 - **AlexNet** [53], a pioneering deep CNN for large-scale image classification.
 - **VGG16 and VGG19** [102], which use a deep stack of 3×3 convolution layers.
 - **GoogLeNet / Inception** [105], using inception modules for multi-scale feature representation.
 - **Xception** [13], an architecture based on depthwise separable convolutions.
 - **ResNet** [36], introducing residual connections to allow for very deep networks.
- **Lightweight CNN architectures for mobile and real-time applications:**
 - **MobileNetV2** [92], designed for low-resource environments.
 - **EfficientNet-B0/B1** [107], using compound scaling to balance performance and efficiency.
- **Advanced training strategies:**
 - **Gradual unfreezing** [44], where pretrained layers are progressively fine-tuned.
 - **Data augmentation and class balancing**, to mitigate overfitting and handle dataset imbalance.
- **Data fusion with wavelet transforms:**

1.3 Context of the research: The CN1–Spoke 5 project, Environmental and Natural Disasters

- Fusion of RGB image features with Discrete Wavelet Transform (DWT) coefficients [63], to capture both spatial and frequency information.
- **Transformer-based and generative models:**
 - **Vision Transformer (ViT)** [22], which replaces convolution with self-attention mechanisms for global context modeling.
 - **Deep Autoencoder**, trained to learn a compressed latent representation of the input images for damage classification.

Such techniques will serve as a basis for performance optimization experiments in order to achieve the best results in classification of damage severity levels, distinguishing between the methods of the RC and masonry buildings, and also in enhancing their real-world adaptability under the post-disaster scenario. This will ultimately enable the designed framework to be a reliable AI-assisted pipeline for post-earthquake structural assessment that eases tasks related to emergency response and digital twin recalibration.

Table 1.1. Overview of the techniques and models considered in this research.

Method / Architecture	Description
AlexNet [53]	Early deep CNN architecture for image classification.
VGG16 / VGG19 [102]	Very deep networks using 3×3 convolutions.
GoogLeNet / Inception [105]	Inception modules enabling multi-scale feature extraction.
Xception [13]	Optimized Inception with depth-wise separable convolutions.
ResNet [36]	Deep networks with residual (skip) connections to enable training of very deep models.
MobileNetV2 [92]	Lightweight architecture optimized for mobile / embedded deployment.
EfficientNet [107]	Compound scaling of depth, width, and resolution for accuracy-efficiency trade-off.
Gradual Unfreezing [44]	Transfer-learning strategy: progressively unfreeze layers during fine-tuning.
Wavelet Fusion [63]	Combines spatial and frequency domain features via wavelet transforms.
Vision Transformer [22]	Self-attention based model originally from transformer architectures for images.
Deep Autoencoder	Learns compressed latent representation and reconstruction; used here for dimensionality reduction / feature extraction.

1.3 Context of the research: The CN1–Spoke 5 project, Environmental and Natural Disasters

This study is a component of the scientific and technological framework of the CN1 National Research Center on High-Performance Computing, Big Data and Quantum Computing, a

1.3 Context of the research: The CN1–Spoke 5 project, Environmental and Natural Disasters

core project of the Italian National Recovery and Resilience Plan (PNRR), supported by the Next Generation EU program of the European Union. Among the various thematic directions delineated within CN1, Spoke 5 is particularly focused on the study of "Environment and Natural Disasters" with the strategic objective of introducing new methodologies and operational platforms for hazard modeling, risk quantification, impact forecasting, and mitigation planning. Spoke 5's goal is to develop a comprehensive, multi-scale, and multi-disciplinary paradigm that could provide a better understanding of the phenomena cascade triggered by natural hazards—earthquakes, landslides, floods, and extreme meteorological events, particularly those that are interwoven with the built environment. The Italian scientific community has largely documented risk assessment through the production of various literatures and practical frameworks during the last few decades. For instance, the authors Pampanin et al. (2025) [78] recently published papers, have also shown the need for integrated platforms capable of combining data of different types and scales in the geological, geophysical, structural, and socio-economic fields as a main challenge. In response to this challenge, Spoke 5 offers a high-fidelity, scenario-based approach that enables the detailed simulation of the spatial and temporal evolution of complex multi-hazard sequences, starting from seismic rupture modeling and further propagation of site amplification, cascading effects (e.g., landslides), and systemic urban impact. The "domino-like" interaction is the new feature: the influence of each occurrence is spread through the interconnected urban layers which are for instance, the building stock, water supply systems, and transportation networks that consequently lead off to measurable risk indicators which are of help in the area of civil protection and public policy. A fundamental element of Spoke 5 is the creation of a proof-of-concept (PoC) demonstrator, which is based on the idea of a Virtual Test Bed (VTB). This virtual urban environment, albeit not directly correlated to a real city, aims to depict the complexity and heterogeneity of standard Italian urban fabrics in areas affected by seismic activities. The PoC showcases the diverse structural typologies (e.g., reinforced concrete, unreinforced masonry, steel, timber) and the critical infrastructure, which are situated in a geo-referenced terrain with an accurate topography and stratigraphy sourced from the Campotosto area in the Central Apennines. In this setting, the high-resolution numerical simulations are achieved by employing the deterministic earthquake scenarios which involve 3D fault rupture models, site-response amplification through finite element analysis, and landslide hazard estimation via the PARSIFAL toolchain. Moreover, the AI-driven structural damage classifiers such as the CNN models developed in this thesis, which are installed on top of these geophysical outputs, make the easiest link between the damage assessment and the post-earthquake structural Capacity recalibration.

The integrated workflow of Spoke 5 through the computational infrastructure of CN1, particularly its HPC and big data capabilities, has risk assessment versatility that ranges from the coarsest to the most detailed work. The rapid typological assessments with census data are one end of the spectrum, while prototyping the Digital Twin for the buildings of strategic nature with the actual design, inspection, and damage occurrence are at the other end. Within this framework, AI-based image classification tools are there to provide damage assessment both as separate modules and as post-event model calibration facilitators: by converting visual representation into probability estimates of the residual structural capacity, they become the device that allows the structural behavior of the digital twin to be changed instantly. Eventually, the sensible interpretation of Spoke 5 is beyond just the academic or theoretical aspects. In fact, the whole integrated modeling proposed is the very tool to act as the operational dashboard of high level decision makers i.e., public authorities and

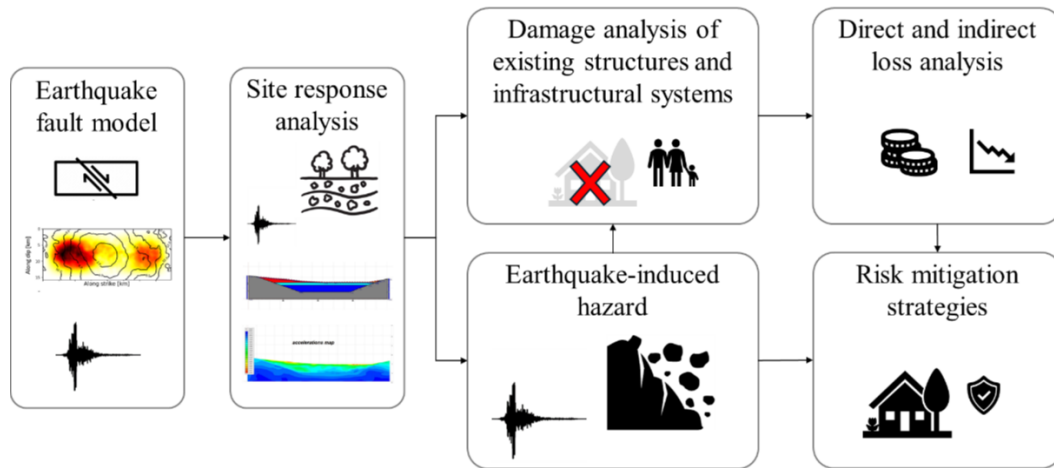


Figure 1.2. CN1 Sapienza Spoke5 - Integrated multidisciplinary workflow, from simulating earthquake sources to conducting risk assessments and implementing mitigation strategies. Image from [78].

emergency planners. By forecasting realistic situations that may occur in the future and by testing the proficiency of the mitigation strategies (for example retrofitting, emergency logistics, and zoning policies), it becomes a co-creator of resilient cities capable of surviving multi-hazard crises and recovering. Image-based damage classification is a vital constituent of this modular system as its presence guarantees that any damage-related ground-truth information can be collected, made available in a processed form, and included almost instantly, thus allowing the risk governance paradigm to be dynamic and responsive.

1.4 Specific details of the CN1-Spoke 5 framework and methodology

The CN1-Spoke 5 project, a part of the Italian National Recovery and Resilience Plan (PNRR), is conceived to define a methodology that is integrated, scalable, and interdisciplinary for the estimation and treatment of the natural hazards impact, especially the seismic events, on the built environment. A major breakthrough of this research is a comprehensive scenario-based multi-risk workflow that bridges the gap between the functions of geophysical simulation, site response analysis, urban vulnerability modeling, and mitigation strategy planning into a continuous pipeline. The basic design for the method is the idea of "domino-like" chain of causality where the first step is the simulation of multiple kinematic earthquake rupture scenarios based on fault systems such as the Gorzano fault in Central Italy. The simulations also feature heterogeneously distributed slips which are produced by using von Karman correlation functions, and the time-dependent rupture dynamics are illustrated by Regularized-Yoffe source-time functions [109]. The optimization of the set of model parameters - slip, rise time, and nucleation location - is done to give a seismic input spectrum that is realistic for the next part of the engineering work [58]. Next, these ground motions are transmitted through different geological layers on the locations by numerical models that are precise. They use the Digital Terrain Models (DTMs) and morphological classification tools like Geomorphons [49] for measuring the topographic amplification. The sources

for stratigraphic amplification are five detailed geological cross-sections in 2D which are discretized by finite element simulations with viscoelastic rheology and variable soil stiffness. Such simulations provide the location-specific ground motions - the accelerations, velocities, and spectra that are used for input in structural as well as infrastructural vulnerability analysis. One of the principal examples treated by the PARSIFAL framework is that of cascading disasters such as earthquake-induced landslides, which allows the local geomorphology-based distribution of the hazard and the triggering seismic conditions. In this way, the authors are able to generate the first comprehensive scenarios of the hazard that include not only the first-time failures but also those reactivations of the already unstable states [78]. The multi-scale modeling of urban systems as layered, interdependent networks is one of the key features of the CN1 methodology that sets it apart from other works. Buildings, lifelines, and infrastructure systems (e.g., water, transportation) are considered as coupled components whose interactions are regulated by the “master-slave” logic: for instance, the damaged buildings may make the roads unusable due to fallen debris [67]. By this, the loss of each urban layer can be evaluated both in terms of direct losses (e.g., repair costs) and indirect losses (e.g., service disruption), therefore, it is possible to estimate the values of several performance indices such as fatalities, the amount of economic loss, and the time of restoration. On the basis of available data and importance, different depths of analytical refinement can be chosen for each structural component. Typological methods to advanced nonlinear dynamic simulations (NLTHA) can be used for all the components, while, in some particular cases, the authors even go as far as turning digital twin representations of the structures into reality. These digital twins meld all the specifics of the geometry and the material with real-time monitoring in their subsequent editions [78]. The demonstrator PoC (Proof-of-Concept) is a Virtual Test Bed (VTB) implementation that represents the Campotosto area, a seismically active basin in Central Italy with complex geological and tectonic features. The urban fabric comprises 177 building archetypes, which represent different typologies (RC, masonry, steel, timber), and have been modeled using OpenSees and 3Muri software for non-linear seismic analysis [78]. Roads and water networks are also part of the infrastructure layers, but more attention has been paid to their interactions with the local topography and hazard scenarios. In particular, this design is a step towards integrating, in the near future, CNN-based vision systems for automating the classification of post-earthquake damage to building components from RGB images. These systems learn through labeled datasets that follow the AeDES classification scale [78], making damage assessments quick, unbiased, and scalable. Such findings do not only facilitate the work of emergency operators on the ground but are also used as dynamic inputs for updating the structural models in the digital twin architecture. To perform the complete job the HPC environment of the CN1 National Centre [78] was used, which allowed the demonstration of the complex hazard chains and their probabilistic impact scenarios over extensive datasets. The merging of AI, earth sciences, and structural engineering through this system, provides a decisive modulating platform of the civil protection forces, technical staff, and the political decision-makers engaged in the reduction of disaster risk and urban resilience planning.

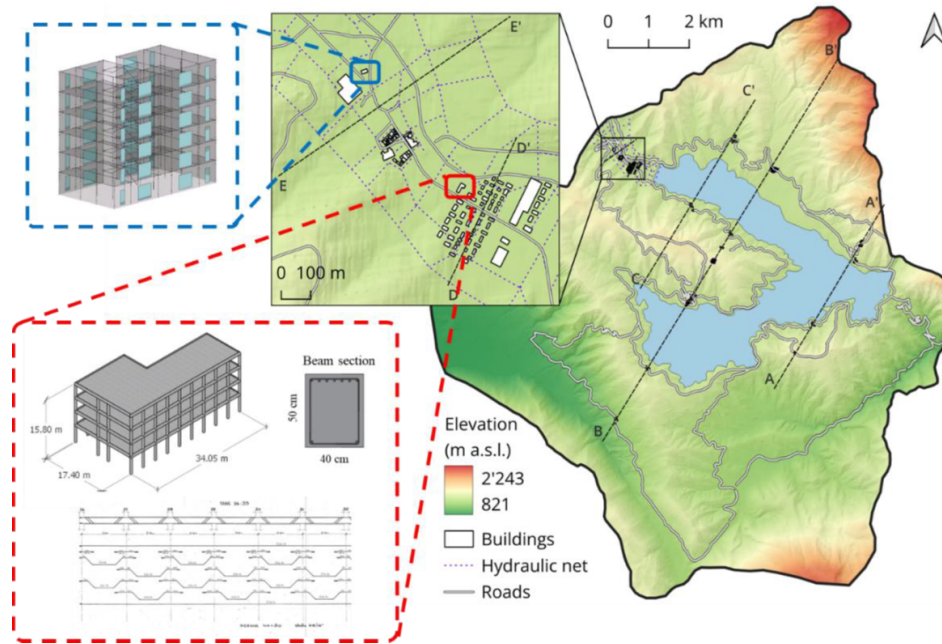


Figure 1.3. Visual overview of the PoC demonstrator, integrating geological setting, infrastructure elements (building stock, roads, water systems), and sampled structural numerical models. Image adapted from [78].

1.5 The Role of Damage Assessment and Structural Capacity Recalibration of Buildings within the Digital Twin Framework

One of the key technological advancements that is gaining traction as a major performer in the post-earthquake scenario is the use of digital twin environments integrated with vision-based damage assessment techniques. The traditional ways of evaluation of structures after an event do not entirely depend on the visual inspections alone but also on the analytical back-calculations. These are, in essence, very reliable but still, they take a lot of time and in large-scale emergencies, they are usually not feasible. To overcome the above constraints, deep learning tools have been advocated, mainly convolutional neural networks, to pave the way from image-based damage recognition to the quantitative evaluation of residual structural capacity. The idea of Ji et al. [2023], linking the digital twin concept as a unifying methodology for going from the visual identification of seismic damage to the numerical safety assessment of the structure, is the very heart of the paper [50].

As shown in Figure 1.4, the conceptual workflow starts with capturing RGB images of a reinforced concrete (RC) structure after a seismic event. Such images are inputs of a CNN model, which is pre-trained and fine-tuned on a labeled dataset derived from the shaking table experiments to classify and localize damage features e.g., spalling of concrete, crack propagation, and degradation of joints. After the visual damage is classified, the framework's next step is to employ these class labels in updating the finite element (FE) model of the structure, thus redefining the stiffness and strength parameters of the individual elements (e.g., plastic hinge rotation capacity, post-yield stiffness degradation). With this, the simulation of residual load-bearing capacity under seismic or gravitational demands can be done depending on the specific damage scenario detected by the vision system.

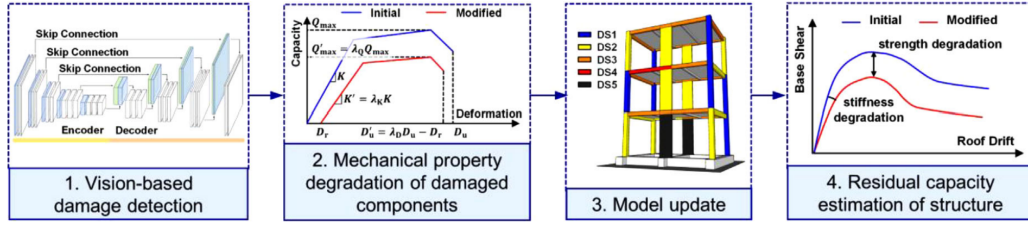


Figure 1.4. Workflow diagram for vision-driven post-earthquake residual capacity evaluation of reinforced concrete structures. Image adapted from [50, 94].

The Visual Damage State Identification and Residual Performance Assessment framework by Ji et al. marks a significant step towards the implementation of machine-vision-based automated and quantitative residual performance assessment systems. The experiment setup of their research involved a five-storey RC frame, which was gradually shaken in incremental shaking table tests. This not only allowed seismic damage evolution to be captured under controlled excitation but also created the basis for a high-fidelity dataset. The images-based CNN model achieved high accuracy in the detection and separation of different levels of degradation of a structure. More importantly, the damage-informed FE model updates enabled the prediction of the residual inter-storey drift capacity and the energy dissipation potential that are the two key indicators, in the safety assessment of structures after an earthquake, with great reliability.

Using deep learning (convolutional neural networks, CNNs) for damage detection, classification, and severity estimation represents a paradigm shift in the way first responders and structural engineers might typically work after a seismic event. Scenario-based research is presented to demonstrate how state-of-the-art AI methodologies can be realistically embedded as a crucial part of DT-enabled urban risk resiliency strategies, which involve continuous performance assessment of critical infrastructure through rapid in situ visual inspection post-earthquake. These AI-enabled workflows not only bring the potential for near real-time operational intelligence but also allow for epistemic uncertainty quantification during structural reliability recalibration via physics-based FE simulations coupled with Bayesian updating.

Such implications become even more prominent if we look through the lens of the CN1-Spoke 5 project and this doctoral research. The very conduct of integrating damage classification models with numerical structural simulators is reshaping the digital twin framework for urban-scale risk evaluation. This method allows for on-the-fly or almost on-the-fly updating of the structural state of the most vital assets from the visual data collected in the field, thus engineers and decision-makers can move on from just tagging the damage to dynamic recalibration of structural reliability. Here, not only is the time of emergency response (e.g., prioritization of inspection or repair) decreased, but also the scenario-based risk modeling in the presence of epistemic uncertainties caused by damage becomes much more accurate.

Ultimately, the use of AI-driven vision systems for residual capacity estimation enables a transformative shift in how post-earthquake safety assessments are conducted. By embedding this approach into the DT-based urban modeling environment described in Section 1.4, the output of CNN models, whether trained on experimental or real-world imagery, can serve as

a direct input for the reassessment of performance-based safety levels of individual buildings or clusters. This lays the groundwork for adaptive, data-driven, and scalable methodologies for post-event structural diagnostics and decision-making.

Chapter 2

Determination of Post-Earthquake Damage Levels

2.1 Seismic Hazards and Damage Mechanisms in Structures

Earthquakes are among the most devastating natural disasters, capable of causing widespread damage to the built environment. The seismic vibrations can lead to the generation of inertial forces within structures that can be of such a level that they overwhelm the buildings' design capacity, hence causing different kinds of damage. These damages can extend to those **structural elements** like reinforced concrete frames, masonry walls, columns, beams, and slabs, and **non-structural components** such as infill panels, internal partitions, facades, and mechanical systems. Certainly, if the earthquakes are minor, they might not cause anything more than some superficial cracks in plaster or infill panels, but in case the seismic events become more intense, they can lead to the most severe damage of the load-bearing elements, which is isolation of the elements by cracking, spalling, and at the fracturing of concrete in which the rebar is initially exposed to be followed by the reinforcements being unbent and in the case of partial or complete collapse. After an earthquake has hit, the condition of the structural elements should be checked as soon as possible and in the most correct way for the sake of guiding the emergency response, allocating the areas with the most rescuing operations, and making decisions regarding the evacuation of buildings, the use of temporary shoring, and long-term rehabilitation. Therefore, the classification of damage resulted from the earthquake is the most basic phase in managing the risk of seismic occurrences [78, 94]. Damage can be categorized in general as follows:

- **Structural damage:** Affecting load-bearing components, including beams, columns, walls, and slabs. Typical indicators include diagonal cracks, crushing of concrete, detachment of the protective cover, bar buckling, excessive deformations, and in severe cases, partial or total collapse.
- **Non-structural damage:** Non-structural damage means those changes that do not contribute to seismic resistance of the building, for instance, masonry infills, partition walls, windows, fixtures, cladding, chimneys, and installations. They can show small plaster cracks, detachment of panels, and falling hazards. Even though they do not reduce structural stability, they can lead to people getting injured and economic losses.

For making sure that damage assessment procedures are standardized and that they can support decision-making, various international frameworks use multi-level damage severity classification. Usually, three levels are taken into consideration:

- **Slight damage:** Slight damage is referred to as superficial or localized damage that mainly impacts non-structural elements or finishes (e.g., hairline cracks in plaster or minor damage to infill walls). The first and foremost point that should be noted is that structural integrity is not affected.
- **Moderate damage:** The description of moderate damage can involve the expansion of cracks both in the range of deeper or wider ones in the structural components, the occurrence of localized concrete crushing, as well as partial detachment of masonry walls. The load-bearing capacity decreases, but the structure is still standing.
- **Heavy (severe) damage:** By heavy damage, the loss of structural integrity that is concerning major aspects such as, significant cracking, widespread spalling, surface damage, and buckling of elements, and partial collapses is meant. The first stage of help including taking down of the building can be allowed [50].

This chapter reviews the main international and national post-earthquake damage classification frameworks with an emphasis on their terminology, criteria, and application. The last section is devoted to the reasons for using three damage classes, i.e., *slight*, *moderate*, and *heavy*, in agreement with the Italian AeDES form.

2.2 Overview of Post-Earthquake Damage Classification Standards

All over the world, different organizations and professional communities have come up with various damage classification systems to assess buildings after earthquakes in a uniform manner. These frameworks are a way to ensure that there is a certain level of consistency in the field inspection and also, they serve as a tool for an easy, fair, and reliable decision-making under crisis situations. Some of the most significant are:

- **FEMA 306 (USA)** — A classification system in the U.S. that steers evaluators to define the damage levels through both structural and non-structural components.
- **EMS-98 (Europe)** — The European Macroseismic Scale is an extensively used model that determines the degree of structural damage and the corresponding potential use of the affected property within the limits of the five categories of damage, thus it leads to the standardization of the reporting system of the macroseismic studies.
- **JBDPA (Japan)** — A fault classification framework established by the Japanese Building Disaster Prevention Association that is closely related to local construction types and seismic resilience.
- **TCIP-DAM (Turkey)** — A damage evaluation methodology designed for on-site quick assessment of reinforced concrete and masonry buildings for the Turkish Catastrophe Insurance Pool to use after severe earthquakes in Turkey.

- **NZSEE (New Zealand)** — A guideline for post-disaster assessment released by the New Zealand Society for Earthquake Engineering, which describes the rapid safety placarding system to classify building usability following seismic emergencies.
- **AeDES (Italy)** — The Italian national standard for damage and usability assessment in emergencies, which is the principal reference in post-earthquake inspections carried out by civil protection authorities.

This brief introduction outlines the major globally recognized rules, the whole set of principles being the basis for planning an immediate and typical damage evaluation process after an earthquake.

2.2.1 FEMA 306 (USA)

In the US, post-earthquake on-site assessment teams refer to FEMA 306, *Evaluation of Earthquake Damaged Concrete and Masonry Wall Buildings – Basic Procedures Manual*, as their main guide published in 1998 [23]. The latter manual is basically an instructional material which explains in detail the methods of how the residual capacity of buildings is determined by the extent of the reduction applied to the stiffness, strength, and ductility of the concerned structural elements. In the first place the document, that is, FEMA 306 identifies the damage categories visually and descriptively with four gradations:

- **Negligible (no significant damage)** An instance of this category is such damage pattern that no capacity change can be linked to the unit.
- **Slight:** The kinds of damage included in this category are those characterized by the appearance of a few cracks, minor structural impairment.
- **Moderate:** Enlarged cracks, some decrease in load-bearing capacity are the features of damaged structures ranked under this category.
- **Severe:** Such damage as penetration of the cover, the breakdown of the system load capacity, the transition to instability is the group of extensive damage nearing the collapse of the structure.

On each damage class, representative diagrams and parameters indicating conservation of capacity are presented. The latter examples are simply an engineer's help, or a base, for updating the structural model used for nonlinear analyses. The pictures of damage (e.g., concrete crushing, cracking, and reinforcement exposure) are visually compared with descriptive matrices, thus enabling a semi-quantitative estimation of structural performance degradation [23] to be carried out.

2.2.2 EMS-98 (Europe)

The European Macroseismic Scale (EMS-98) [34] is an essential tool in both seismology and civil engineering, as it allows building damage and seismic intensity to be quantified in a standard way. The concept of damage to the structural part of the building is one of the main issues addressed in the EMS-98 and it has been approached by dividing structural damage into five categories or grades:

- **Grade 1 (Negligible to slight damage):** The most common defect is the appearance of very fine cracks in the plaster of internal walls and no damage to the structure.
- **Grade 2 (Moderate damage):** Small cracks are visible in the load-bearing elements of the structural frame of the building (e.g. concrete beams and columns).
- **Grade 3 (Substantial to heavy damage):** The description here is of such types of damage as the breaking of the main structural elements into large pieces, the detachment both of the structural and non-structural parts, etc.
- **Grade 4 (Very heavy damage):** Intense structural distress is present in the building and at least some parts of it have collapsed.
- **Grade 5 (Destruction):** Most probably, there is the near or complete structural failure of the building.

Besides, these levels are illustrated with characters depicted in the visual damage and the textual descriptions. The EMS-98 scale is considered as one of the reference scales for the scientific studies and the scales used in operations after the disaster in many countries in Europe are based on it. It also has a role in the evolution of national instruments like the AeDES form in Italy.

2.2.3 JBDPA (Japan)

The Japanese method predominantly steered by the Japan Building Disaster Prevention Association (JBDPA) encompassed a five-tier damage grouping system that was characteristically fitted for reinforced concrete structures [60]. The system depicts the changes in the damage features of the structure from the category of Class I (minor damage) to that of Class V (severe damage or near-collapse).

It is worth noting that every level in this system is linked with a numerical factor of performance reduction that allows the modification of structural models of capacity after an earthquake event. Accordingly, for their non-linear modeling, engineers here employ the weakened efficiency values as they take into account the phenomena of cracking, joint failure, or plastic hinge rotations which have been observed [60, 11].

Element-level damage:

JBDPA guidelines specify the damage characteristics of cracked concrete cover and steel corrosion as the main observed parameters for damage classification of reinforced concrete members [61, 60]:

- Class I – Only very fine surface cracks; the crack widths are less than 0.2 mm; no concrete spalling or exposure of reinforcement.
- Class II – The cracks with a width of 0.2 mm to 1 mm were described; flaking of concrete cover was minor, and initial signs of spalling were present.
- Class III – The main cracks of no more than 2 mm and more than 1 mm were described; alongside concrete spalling of a moderate extent, steel reinforcement is partially exposed.

- Class IV – The main cracks even more than 2 mm in width, considerable spalling were found; as a result of exposure of reinforcing bars, there may be an onset of rebar deformation besides being affected by corrosion.
- Class V – Buckling of reinforcement, crushing of core concrete, and vertical deformation constitute the severe damage symptoms; the damaged structure may demonstrate side-sway, partial collapse, or bar fracture.

In reality, Class I mostly means no visible damage; Class II is related to minor cracking or spalling; Class III describes cases of conspicuous cracking with some degree of cover loss; Class IV is mainly characterized by severe cracking, spalling, or reinforcement buckling; while Class V is typically associated with collapse or total loss of structural capacity. Experimentally, researchers have determined reduction factors related to different damage classes; for example, reinforced concrete columns belonging to Class I would be able to keep up to 95% of their strength, and columns from Class IV possibly only close to 10% of their strength, while those from Class V retain almost no strength [11]. The use of these parameters in nonlinear simulation allows the implementation of a method for evaluating post-earthquake performance of RC elements with the help of damage distribution maps as well as the high accuracy of damage estimating in structural assessment workflows.

2.2.4 TCIP-DAM (Turkey)

Following the 1999 great earthquakes, Turkey regulated the procedures for the damage assessment after the earthquake through the Turkish Catastrophe Insurance Pool (TCIP) and introduced the TCIP-DAM method in 2020 [47].

The new classification protocol of TCIP-DAM features two complementary categories: Building-level damage:

- No Damage
- Slight Damage
- Moderate Damage
- Severe Damage
- Urgent Demolition
- Collapse

Element-level damage: Individual structural elements such as beams and columns are categorized from Type O (undamaged) through to Type D (collapsed), based on observed degradation indicators.

- Type O: No visible damage.
- Type A–C: Increasing severity of damage - e.g., cover spalling, shear or flexural cracking, buckling of reinforcement, or core crushing.
- Type D: Represents catastrophic damage, including member collapse or total loss of load-bearing capacity.

Among other points of interest, TCIP-DAM-2020 very explicitly sets out the criteria for when a building should be judged to be “Heavily Damaged,” namely that at least three vertical RC elements must have concrete cover failure, or only one vertical element must have reinforcement buckling, crushing, or large residual drift. This stepwise, component-based method allows for a quick, objective damage rating, which is a great advantage for vast-scale post-disaster activities, and also provides assistance for making the right decisions in terms of repair, retrofit, or demolition.

This method is applicable in technical inspections and insurance compensation assessments, and it can be identically used in both residential and public buildings.

2.2.5 NZSEE (New Zealand)

Back in 2009, the New Zealand Society for Earthquake Engineering (NZSEE) put out the "Building Safety Evaluation During a State of Emergency – Guidelines for Territorial Authorities," that not only set a uniform rapid assessment procedure after any calamity but also provided detailed recommendations on the scope, categories, and roles of the assessment team [71]. Territorial Authorities were the target audience for these instructions, which mainly dealt with the initial on-site inspection of the building's safety condition. A three-class placarding system was proposed in the document, very much like the U.S. ATC-20 approach concept, but the terminology used was different:

- Green (Inspected): The building is safe for the activities it was designed for and maintenance programs will continue as expected.
- Yellow (Restricted Access): The structure can be accessed by authorized persons only, with safety limitations and under the supervision of a trained person, until further studies are conducted.
- Red (Unsafe): Because of the probable high existence of structural hazards due to which the building might collapse, causing abrupt, fatal accidents, extreme caution must be taken so very importantly no human should be allowed to enter the area.

Although NZSEE (New Zealand Society for Earthquake Engineering) guidelines have largely been revised since the first seismic events in Canterbury, NZSEE (New Zealand) protocols are still not there yet in terms of a structural element-level classification framework (columns, walls, beams) basis for gradation of damage in engineering evaluation of a post-earthquake scenario. The current methods implemented mainly concentrate on whole-building placarding along with usability categories that are grades of damage of separate components rather than evaluation of those components.

2.3 The AeDES Form and National Italian Guidelines

The AeDES form [5] used in Italy to identify the damaged structures after an earthquake is the format of the **AeDES** form, an abbreviation for *Agibilità e Danno nell’Emergenza Sismica*, which means "Usability and Damage in Seismic Emergency." GNDT (National Group for the Defence against Earthquakes) and the Italian Department of Civil Protection (Protezione Civile) worked together to create this standardized tool in the late 1990s, and it

was first used in the entire country after the seismic events of Umbria-Marche in 1997 and institutionalized after the 2002 Molise earthquake.

The purpose of the AeDES form was to lead the procedure of the quick but systematic survey of buildings damaged by earthquakes, mainly those related to the usability of structures and damage to both the structural and the non-structural parts of the buildings.

In fact, the form is also a comprehensive data collection and classification tool that supports decision-making on whether buildings should be evacuated, restricted in access, or can remain safely in use or not.

2.3.1 AeDES Form Damage Classes

The AeDES form is divided into nine thematic blocks with the first referring to the general metadata of the building and the second to the geographical location of the building. In the third section, construction typology is described, while structural damage is the main focus of the fourth section dealing with damage to beams, columns, walls, and slabs. Damage to the non-structural parts is explained in the fifth section. In the sixth and seventh parts, the presence of short-term hazards and the overall judgment of usability are given respectively. The peculiarity of the AeDES form is its linkage to the European Macroseismic Scale EMS-98, which is the source of a simplified damage severity classification presented in the form. The EMS-98 scale contains five damage grades, from D1 to D5. AeDES, however, merges these into three main severity levels to make the practical use in the field more straightforward [5]:

- **Slight damage** (corresponding to EMS-98 D1): Hairline cracks in plaster, non-structural damage without impairment of structural elements.
- **Moderate damage** (corresponding to EMS-98 D2–D3): Visible cracking in beams and columns, detachment of infill walls, moderate deformation of structural joints, partial spalling of concrete.
- **Heavy damage** (corresponding to EMS-98 D4–D5): Severe structural failure such as large cracks, instability, buckling of reinforcement, or partial collapse of components.

Damage level - extension 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	
1	Vertical structures	☐	☐	☐	☐	☐	☐	☐	☐	☐	○
2	Floors	☐	☐	☐	☐	☐	☐	☐	☐	☐	○
3	Stairs	☐	☐	☐	☐	☐	☐	☐	☐	☐	○
4	Roof	☐	☐	☐	☐	☐	☐	☐	☐	☐	○
5	Infills and partitions	☐	☐	☐	☐	☐	☐	☐	☐	☐	○
6	Pre-existing damage	☐	☐	☐	☐	☐	☐	☐	☐	☐	○

Figure 2.1. The (Italian) AeDES form: damage levels for each structural element [5].

Table 2.1. AeDES damage level classification and description of the observed damage to structural and non-structural components [5].

Damage level	Description
D1 Negligible to slight damage	Beams: crack width ≤ 1 mm (not vertical) Columns or walls: crack width ≤ 0.5 mm (not vertical) Infill walls: diagonal cracks ≤ 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

AeDES classification is independently applied to each and every structural and non-structural element of the building such as vertical bearing systems (walls, columns), horizontal systems (slabs, beams), staircases, roof elements, and secondary systems. The judgment of each element is based on direct observation and the damage class representing the corresponding damage are assigned, thus the building is made local damage profiles from the combination of the categories. This information is crucial for the correct drawing of the outline of activities, of the kinds of interventions, of the emergency planning, and of post-earthquake retrofitting programs.

Moreover, the AeDES form is an invaluable source for scientific research and AI applications. Its widespread use over decades has led to the creation of rich national damage databases, which have been exploited in probabilistic seismic risk models, fragility curve calibration, and now, increasingly, in machine learning frameworks for automatic damage detection.

The reason that our research uses the AeDES classification as the baseline for all re-labeled images in our dataset is to ensure that the data and the AI models trained on it are consistent with the official national standards for post-earthquake assessment [94].

2.3.2 Potential Applications of AeDES Data for Global Building Damage Classification via Machine Learning

One of the recent research works shows that the granular damage data captured with the help of the AeDES form- which was originally designed for per-element usability assessments- can not only be combined but also can be used to classify damage severity at the whole-building level by using machine learning (ML) techniques. For example, Aloisio et al. (2024) have shown how ML models developed on AeDES-derived field evaluations of the 2016–2017 Central Italy earthquakes can provide accurate predictions regarding a building's overall usability state, such as safe, restricted, or not usable [1]. The proposed approach is supported by two main reasons:

- **Element-to-Building Translation:** The ML model can aggregate damage / fault



Figure 2.2. Damage level on building scale: (D4/D5) - Collapse or inclination more than 1%. Image from [5]



Figure 2.3. Damage level to single structural element (vertical structures): (D4) - Damage to the beam-column joint with concrete spalling and Crushing of the infill panel. Image from [5]

indicators from the smallest unit of components (e.g., beams, columns, infill walls) and then it can figure out the degree of damage to the whole structure more than what conventional rule-based aggregation can do.

- **Automation for Rapid Response:** By employing data-driven algorithms (e.g., random forests, support vector machines, neural networks), one can carry out the assessments faster and on a larger scale, which is very important in the case of post-earthquake emergencies when rapid decisions are required.

By using AeDES data in ML workflows, researchers can locate global damage states of the houses with the same or even higher accuracy as traditional methods that allow systematic, scalable, and automated damage evaluation strategies. Such an approach is fully consistent with the current study objectives where the annotations made with the help of AeDES are used to train the vision-based, AI-driven models for damage classification following the Italian standards of structural assessment.

2.4 Selection of Damage Levels for This Research

					
CNN algorithm classification	AeDES	EMS-98	FEMA 306	JBDPA 2015 (Maeda et al. 2019)	TCIP (Ilki et al. 2021)
Light	Light	D1 – Negligible to slight damage	Insignificant	I	Type O
Moderate	Medium-Severe	D2 – Moderate damage	slight	II	Type A
		D3 – Substantial to heavy damage	moderate	III	Type B
Heavy	Very Heavy	D4 – Very heavy damage	heavy	IV	Type C
		D5 – Collapsed	extreme	V	Type D

Figure 2.4. Damage scale comparison.

Considering the different regulatory frameworks and damage classification scales, for this work it was decided to use a damage scale similar to the AeDES system, with the main characteristics of the damage divided into three levels: the first level is minor damage, the second - moderate damage, and the third - heavy damage. The scientific and practical considerations are the basis of this decision. One of these grounds is the correspondence of the AeDES criteria to the Italian standards, achieved through the Italian standard classification system, as it is illustrated in this paper [94].

As mentioned before, the AeDES form is instrumental in damage classification, the categories of *slight*, *moderate*, and *heavy* being those that it employs. The most straightforward way to make the results of our research comparable to field reports is to have the same levels. In addition, this allows for as much applicability as one can consider: the tools and methodologies that this research creates may get directly integrated into or be part of the

support of the activities of the post-earthquake on-site inspection, etc. Here the "language" is not different, the one used in the Italian seismic emergency system is the exact one [94].

Moreover, by summarizing the damage types to only the three main categories, the damage evaluation through the automated system can be facilitated as well. Many recent studies on the application of computer vision algorithms - especially the convolutional neural networks (CNNs) - for damage recognition have demonstrated that when there are too many categories of damages to discriminate, it is very likely to get ambiguous results, particularly in the cases of slight and moderate damage [93]. Labeling with three easily distinguishable classes not only enables the training of more reliable classifiers but also reduces the amount of uncertainty in those cases where the borders are crossed.

In particular, the dataset of images of the structural damage such as cracks, partial collapses, spalling, etc. was reclassified according to the AeDES criteria of the following categories: *Slight*, *Moderate*, and *Heavy*. This sets a clear and standardized data basis to automatically train our algorithms of classification and also to verify their performance [94, 93].

This decision to simplify the damage range also enables easier verification of the classification results. For example, if a CNN-based model categorizes a particular structural damage as the *heavy* class, then this corresponds without any doubt to an "H" level (heavy damage) in the AeDES form, thus comparison between automated and manual assessments is very easy.

On the ground, going three levels deep in the case of damage in different structural components (e.g., column, beam, wall) will give the possibility to tag an element as one of the following: intact or slightly damaged, moderately damaged, seriously compromised. Thereby, such resolution at the element level is of particular importance as in most cases, the final decision on the reusability of a building is heavily reliant on the component, which has suffered the most drastic damage (i.e., the structural "weakest link"). Hence a type of automatic system that is able to spot the few critical parts with heavy damage in a building that is quite lightly affected can be of great help in directing the evacuation of the inspecting team to the area and also prioritizing the interventions of the time-critical facilities.

In conclusion, it is evident from Table 2.2 that the classification of levels of damage into three categories, which was implemented in this study with the AeDES method as its basis, can be considered the most suitable choice that balances the criteria of accuracy, compliance with other standards, and functional value in facilitating the work of the technical staff in the aftermath of an earthquake crisis.

2.5 Final Considerations on the Adopted Damage Levels for This Research

This chapter has addressed the topic of post-earthquake damage level determination by reviewing major international frameworks (FEMA in the United States, EMS-98 in Europe, JBDPA in Japan, and the TCIP protocol in Turkey), as well as the Italian AeDES method. A comparative study of the abovementioned standards shows that all of them define the same basic damage levels albeit in different technical details. Hence, the main correspondence between the damage categories in the different systems (especially *slight*, *moderate*, and *heavy/severe*) is evident not only in terms of the regions addressed but also the types of organizations involved. Such matching is not at all accidental; on the contrary, it truly reflects

Table 2.2. International Damage Classification Systems and Their Mapping to AeDES Classes

Framework	Original Damage Levels	Mapped to AeDES Classes
FEMA 306 (USA)	None, Slight, Moderate, Severe	Slight → Slight; Moderate → Moderate; Severe → Heavy
EMS-98 (Europe)	D1 (Negligible), D2 (Moderate), D3 (Heavy), D4 (Very Heavy), D5 (Collapse)	D1 → Slight; D2–D3 → Moderate; D4–D5 → Heavy
JBDPA (Japan)	Damage Classes I through V based on increasing cracking and deterioration, with associated reduction factors	Classes I–II → Slight; III → Moderate; IV–V → Heavy
TCIP-DAM (Turkey)	No Damage, Slight, Moderate, Heavy, Urgent Demolition, Collapse	Slight → Slight; Moderate → Moderate; Heavy/Urgent Demolition/Collapse → Heavy
AeDES (Italy)	Slight, Moderate, Heavy (simplified from EMS-98)	Adopted directly in this research

a common requirement for damage assessments that are both easy-to-understand and apply and that also provide a sufficiently detailed description of the structures damaged by the earthquake. The emergency actions that will be taken rely heavily on the damage scale used; if scales are too simplistic (e.g., usable/unusable), vital information may not be captured, while too detailed (five or more levels) may hinder the work of the rater and the harmony of rater's conclusions. The approach embraced in this work to use three-level damage scale (*slight, moderate, heavy*) as per the AeDES form constitutes a very balanced and proper solution drawn from the Italian emergency experience yet it can be implemented worldwide. Also, by resorting to the AeDES scale, the AI-driven classification tools that will be used become compatible with the paper-based forms spoken "language" and the communication that is both traditional and digital is made easier and more consistent due to the coexistence of these two workflows. The decision guarantees:

- Observance of national emergency protocols
- Harmonization with the data of the field inspection
- Easier training goals for CNN models, as there are no more minor ambiguities
- Element-wise labeling facilitating the update of the digital twin and the prioritization of risk

This approach will form the backbone of the classification systems that rely on vision and that have been developed in the present research. It enables direct support to field engineers, real-time decision-making, and integration into national digital twin platforms for post-seismic structural evaluation. Moreover, having three different classes can give the model a stable performance even with a small number of images per class, and it can also

reduce the problem of classification uncertainty that is usually encountered when the class boundaries are close [94]. Hence, by linking AI-generated labels to AeDES-compliant terminology, this framework can disconnect computer vision-based automated systems from operational field workflows. To sum up, the step of adopting a unified three-level damage classification system for each structural element has been proved to be the most decisive in bridging traditional visual inspections with automatic methods that are being developed. It is this synergy that will be instrumental in speeding up and objectivizing post-earthquake assessments, thereby, at the same time, ensuring enhanced public safety and efficient seismic emergency management.

Chapter 3

Computer Vision-Based Damage Classification: An Overview

In this chapter, will be presented a comprehensive overview of computer vision-based methodologies for classifying post-earthquake structural damage from RGB images. The chapter is structured as follows: after a general introduction to image-based damage assessment in structural and earthquake engineering in Section 3.1, mentioning conventional approaches that rely on crack edge-detection and feature extraction combined with classical machine learning algorithms. Section 3.2 transitions into modern deep learning strategies, introducing the fundamentals of neural network-based approaches for damage classification. Section 3.3 then focuses on the foundation on Deep Neural Network (DNNs), while Section 3.4 on Convolutional Neural Networks (CNNs), mentioning the architectures most commonly employed for image classification tasks. Section 3.5 briefly describes Transfer Learning and Data Augmentation techniques used to enhance the training performance metrics of these models. Building upon that, Section 3.6 and Section 3.7 present a detailed literature review of studies that have applied CNNs, and more recently, variants and other state of the art approaches, to the task of structural damage classification from RGB images. This chapter lays the foundation for the subsequent methodology and experimental chapters focused on leveraging these models in an Italian context aligned with AeDES standards.

3.1 Introduction: Image-based Damage Classification in Earthquake Engineering

Post-earthquake damage assessment has been conventionally carried out by physical inspections performed by qualified engineers who gather both qualitative and quantitative information to judge the damage extent. Yet, these manual inspections are usually time-consuming, subjective, and thus, unscalable in the scenario of large seismic disasters. As a solution, the structural health monitoring (SHM) community has started to use the computer vision field to execute the damage evaluation process automatically. In detail, the procedure of **damage classification** from RGB images can be considered as a supervised image classification problem, where the target is to assign a picture showing a structural element or façade to one of the discrete damage levels such as *Undamaged*, *Slight*, *Moderate*, or *Heavy*. This problem does not call for object detection or semantic segmentation but is only a form

of global image classification. Such a classification is typically done in accordance with the likes of AeDES, FEMA306, or EMS-98. Recent works, including the comprehensive surveys by Azimi et al. [4] and Ferraris et al. [24], outline various traditional and AI-based methods. Classical computer vision workflows consist of the manual feature extraction (e.g., histogram of oriented gradients, SIFT, or wavelets) and the subsequent classification of the extracted features with SVMs or Random Forest, for instance. These techniques frequently call for domain expertise and are quite vulnerable to the condition of the images. On the contrary, deep learning (DL) frameworks, most notably Convolutional Neural Networks (CNNs), have shown outstanding capability in one go developing suitable features directly from the input images without the need for any user intervention. The models that are built on the CNN architecture have found a large number of successful applications in different SHM areas such as the detection of the crack, identification of corrosion, the inspection of the disaster after the event [25, 27, 41].

Before the Deep Learning Era, the issue of structural damage classification was mainly tackled using pipelines that merged image preprocessing and manually handcrafted feature extraction with statistical or machine learning classifiers. Common steps included:

- **Image pre-processing, enhancement and filtering:** The main techniques were grayscale conversion, histogram equalization, edge-preserving filtering, thresholding, each of which was used to increase visual contrast and reveal the damage pattern. Other classical techniques were edge and ridge detection filters to highlight linear damage features, and matched-filter approaches, in particular, aerial imagery contexts, to detect cracks with morphological post-processing [57].
- **Feature extraction:** To extract features the following descriptors were used: Gabor filters, Local Binary Patterns (LBP), morphological features and gradient-based representations from the interest areas [114].
- **Classification:** The available features were used as an input to classifiers such as Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), Decision Trees or Random Forests wherein the features were assigned a damage severity label.
- **Template matching:** The approach took a small reference image (template) of a representative damage pattern, such as a crack or a detached patch, and, as it moved the template across the target image, it calculated similarity metrics (e.g. normalized cross-correlation) between the two. The areas with the highest similarity to the template were considered the closest to damage occurrences. The method was efficient in detecting the distinctive damage pattern under controlled conditions.

Template matching, in particular, was appreciated for its interpretability and low computational overhead, but its performance suffered significantly under variations in damage orientation, scale, background complexity, and illumination. As such, while these traditional methods could be effective in constrained laboratory settings, they generally struggled to generalize to real-world post-earthquake imagery, where environmental variability and damage diversity are substantial. These limitations ultimately catalyzed the shift toward more robust deep learning-based approaches [24, 112].

Figure 3.1, taken from [4], depicts a comprehensive overview of the large number of machine learning algorithms applied in computer vision. The deep learning approaches, marked in blue, constitute the next chapter of the present work.

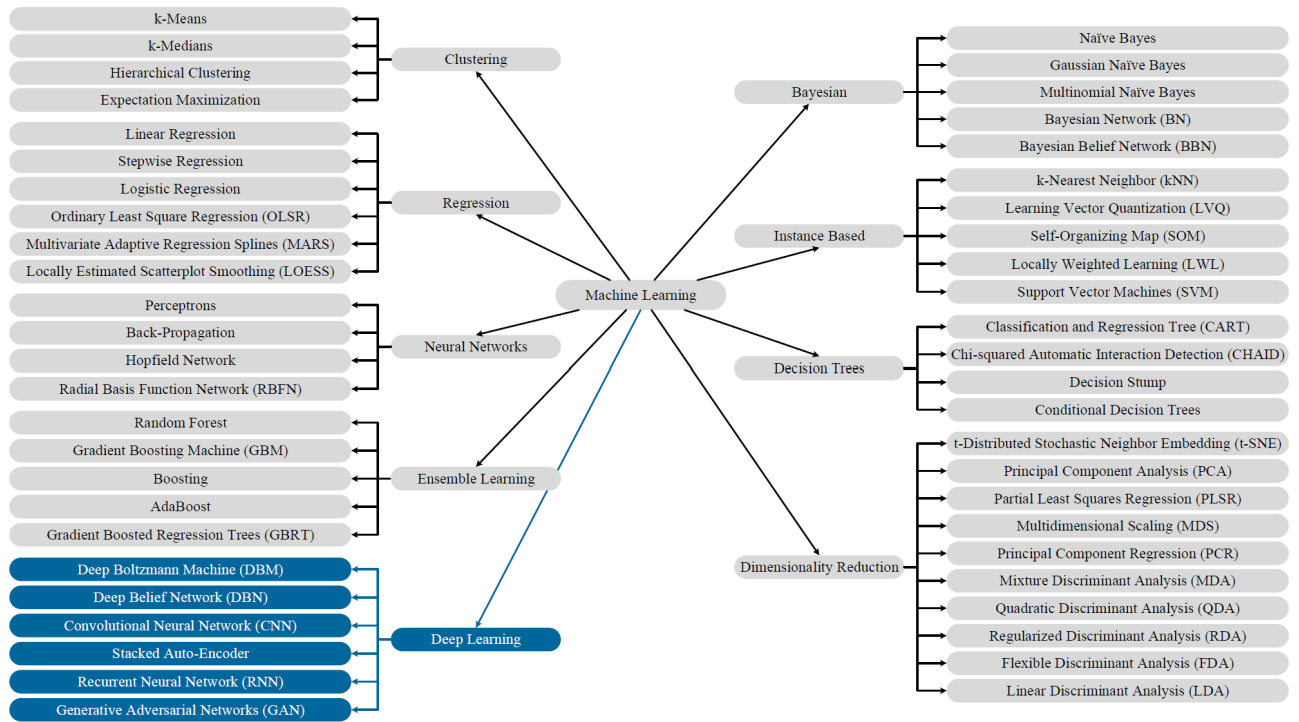


Figure 3.1. Overview map of ML algorithms used in CV applications. Image from [4].

3.2 Deep Learning Methods for Image Classification

In deep learning and, to be more specific, in a structural damage assessment application after an earthquake, several architectures are available for image classification. Below is a summary of major methods available and considerations:

- **Convolutional Neural Networks (CNNs):** CNN structures include as one of their important features the efficient hierarchical extraction of spatial patterns that appear in the images such as micro-cracks and spalling patterns [54, 35].
- **Autoencoders (AEs), along with Convolutional Autoencoders (CAEs) ones:** These are unsupervised learning models aimed at learning concise latent representations of images. The features which are encoded can be used as input features in combination with a classifier for downstream supervised classification [37, 66].
- **Vision Transformers (ViTs):** The idea behind the transformers is to break down the images into patches-tokens and use the self-attention mechanism to capture the global dependencies among the tokens. In this way, the transformers offer a potent alternative to the standard CNNs [21, 110]. However, ViTs generally require larger annotated datasets or strong pretraining to match or exceed CNNs, especially in domains with limited data.
- **Generative Adversarial Networks (GANs):** Though they were mainly proposed for generative tasks, the use of GANs extends also to semi-supervised learning, data augmentation, as well as to learning latent representations, which then help classification tasks [31, 87, 113].

Among CNNs there are different and possible approaches:

- **Residual and Dense Networks (ResNet, DenseNet):** ResNet uses skip-connections to make the training of very deep models easier [35], while DenseNet links each layer with the layers that follow to allow feature reuse and gradient propagation [45].
- **Lightweight Architectures (MobileNet, EfficientNet):** These models are designed to run in resource-constrained environments without losing much accuracy. MobileNet uses depthwise separable convolutions to reduce the number of parameters [42], while EfficientNet comes up with a compound scaling method for the depth, width, and resolution [106].
- **Inception and Xception Models:** Inception networks use multi-scale convolutions in the same layer to get more features [104], but Xception makes a better use of the resources by applying depthwise separable convolutions only [13].
- **Hybrid Architectures (CNN–Transformer):** The combination of these architectures allows the use of convolutional layers to grab local spatial features and transformers to get the global self-attention, combining in this way local feature extraction (via convolutions) with global context modeling (via self-attention). ViT [21] and the hybrid models like ConViT [16] present the integration as one that not only deepens the contextual understanding but also retains spatial awareness.
- **Neural Architecture Search (NAS) Models (e.g., NASNet):** Architectures that are found through neural architecture search (NAS) are one of the reasons to propose such a system to maximize the performance and efficiency of the algorithm as it is mentioned in NASNet [115].

In a general categorization of deep learning models, besides the most well-known convolutional architectures, other innovative deep learning families that appear in the literature include:

- **Recurrent Neural Networks (RNNs):** The main difference between RNNs and regular feedforward networks is the addition of feedback connections that make it possible for the network to have a memory of past events while it processes sequential data. It is due to this feature that RNNs are good candidates for time series analysis, speech recognition, and natural language processing tasks [89, 39].
- **Capsule Networks:** The main idea is to train CapsNets to segment and classify objects by building feature hierarchies while explicitly modeling the part–whole relationships thus providing the network with improved viewpoint adaptation and affine invariance properties [90, 38].
- **Graph Neural Networks (GNNs):** Highly flexible models capable of learning from graph-structured data, GNNs extend the concept of deep learning to inputs that can naturally be represented as a graph. They allow the acquisition of representations in domains such as social networks, molecular structures, or structural systems modeled as graphs (e.g. structural elements connected, meshes, relationships among components) [97, 52].

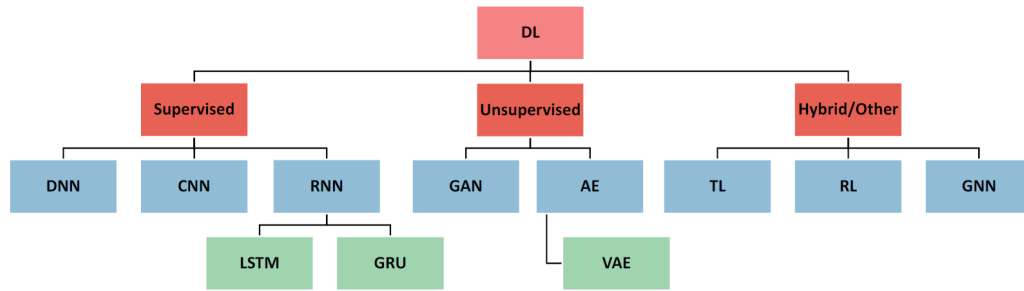


Figure 3.2. Overview of different deep learning methodologies, including DNNs, CNNs, autoencoders, GANs, transformers. Image adapted from [112]

To sum up, CNNs, autoencoders, and ViTs are particularly suitable for the problem of post-earthquake damage classification with RGB images.

Figures 3.1 and 3.2 are two diagrams that broadly show the existence of different machine learning and deep learning models. The figure in the context of image-based structural damage classification informs about the main families of models that are used under the supervision and control of learning. The power that CNNs have garnered over the years still positions them as the most referred to architecture for image classification in the scientific community; their success comes from their hierarchical extraction of features that stand for the spatial arrangement in RGB images of damage. In parallel, deep convolutional autoencoders (CAEs) are providing unsupervised methods which allow compression of image data and the encoding of the information into the latent space from which classification or anomaly detection are the possible subsequent uses. Even though the figure does not name transformer-based architectures, a small but clear and definite statement addressing Vision Transformers (ViT) is necessary since they have become one of the most challenging competitors of CNNs in setting new benchmark results through the implementation of self-attention mechanisms in modeling image patches as token sequences. Overall, while CNNs remain the backbone of many state-of-the-art systems in post-earthquake image-based damage classification due to their robustness, lower data requirements, and efficient learning of local spatial features, emerging architectures like ViTs and hybrid CNN-Transformer models are proving viable competitors, especially as larger annotated datasets become more available and computational resources grow. For this reason, benchmarked comparisons among architectures, considering accuracy, robustness to domain shift, interpretability, cost, and latency, will determine which architectures are most suitable for real-world deployment applications.

3.3 Deep Neural Networks (DNNs)

Before introducing Convolutional Neural Networks (CNNs), it is useful to briefly recall the principles of Deep Neural Networks (DNNs), and in particular the Multi-Layer Perceptron (MLP), since CNNs can be regarded as a specialized extension of this foundational architecture. Deep Neural Networks (DNNs) represent a broad class of architectures within deep learning, typically defined as networks composed of more than three hidden layers [32]. Most DNNs are based on the Multi-Layer Perceptron (MLP) structure, which is a supervised feedforward

model. In the MLP model, the information moves from the input layer to the output layer, passing the hidden layers in between. Every neuron applies a weighted sum to the output activations from the layer that came before it, and then it applies a non-linear activation function. This step can be written as:

$$a_j^{[k]} = f \left(\sum_{i=1}^n w_{ij} a_i^{[k-1]} + b_j \right), \quad (3.1)$$

where $a_i^{[k-1]}$ are the activations from the previous layer, $w_{ij}^{[k]}$ the weights, $b_j^{[k]}$ the bias term, and $f(\cdot)$ the nonlinear activation.

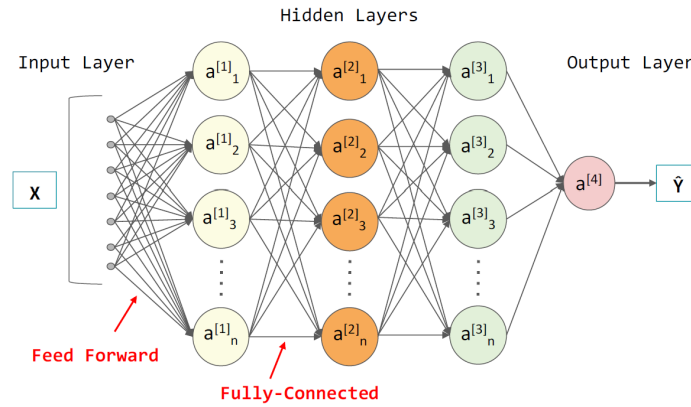


Figure 3.3. Multi-layer perceptron scheme. Image adapted from [112]

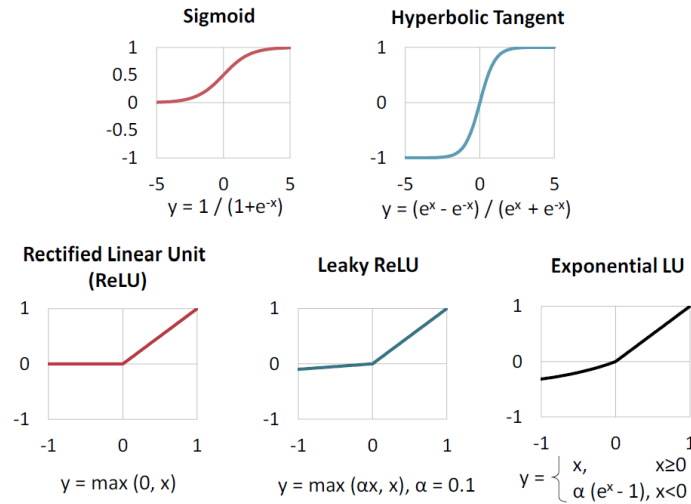


Figure 3.4. Different Activation Functions. Image adapted from [112]

The process of training is done by minimizing a loss function L using gradient descent. At the same time, backpropagation carries the error efficiently through the network to update

the parameters. Although a method like MLP-based DNNs is very useful for tabular and vectorized data, it fails when it comes to high-dimensional structured data like images. That is why architectures such as CNNs have been developed.

3.4 Convolutional Neural Networks (CNNs)

Unlike Multi-Layer Perceptrons (MLPs) that depend on the dense connectivity of fully connected layers, CNNs utilize local receptive fields via convolutional filters which significantly reduce the number of trainable parameters and enhance generalization capabilities [54]. The process of a simple 2D convolutional layer, with stride = 1 and no padding (valid convolution), for a single-channel input, is as follows:

$$z_{i,j}^{(k)} = \sum_{m=1}^M \sum_{n=1}^N w_{m,n}^{(k)} x_{i+m,j+n} + b^{(k)}, \quad (3.2)$$

In this equation, x stands for the input feature map, $w^{(k)}$ is the learnable kernel of the k -th filter with the size $M \times N$ and $b^{(k)}$ is the bias term. The output, $z^{(k)}$, now becomes a feature map that indicates the unique local features like cracks or spalling patterns of the damaged buildings. Moreover, for further invariance to local translations and lower computational cost, pooling layers are generally used which introduce such operations as max- or average-pooling.

These design choices (filter size, stride, padding, activation, pooling) must be carefully balanced: smaller filters and stride improve resolution of local detail (important for fine damage), while larger receptive fields, pooling, and deeper layers help capture more global context but risk losing fine structural damage cues. In practice, tuning these hyperparameters is the way to optimize performance in damage classification.

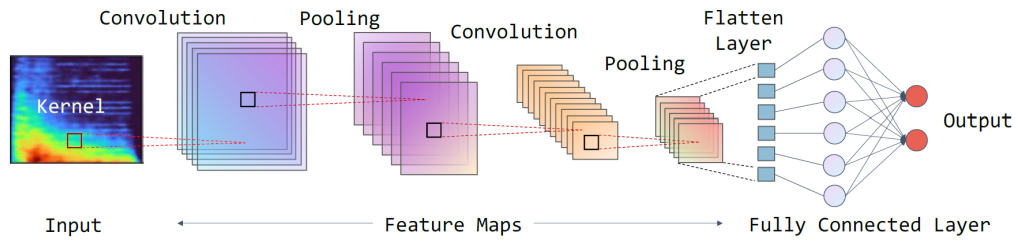


Figure 3.5. A typical scheme of a 2D Convolutional Neural Network. Image adapted from [112]

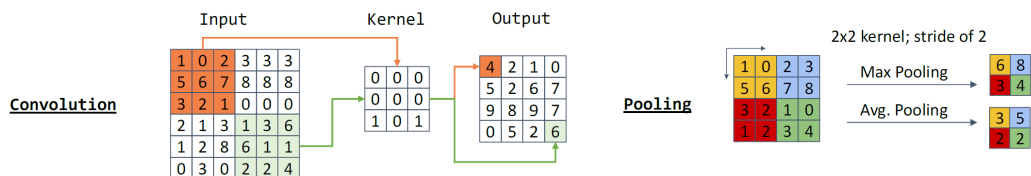


Figure 3.6. Example of a convolution operation, on the left. Example of and two forms of pooling, on the right. Image adapted from [112]

Convolutional neural networks (CNNs) are capable of hierarchically extracting more abstract features as they stack multiple convolutional and pooling layers. The process culminates in the fully connected layers which perform classification. Activation functions (e.g., ReLU) are used to introduce the nonlinear mapping between inputs and outputs, and the parameters are optimized using gradient descent with backpropagation. This feature learning hierarchy is the main reason why CNNs are so effective in the application of structural health monitoring, a field where the co-occurrence of global façade damage and local defects is typical [26]. To sum up, CNNs refer to deep neural networks designed and trained on data with a spatial structure, such as images. These networks include layers that execute local convolutions, nonlinear activation functions (ReLU), and pooling operations. Through this architecture, CNNs can progressively uncover visual features ranging from low-level characteristics, e.g., edges, to high-level complex structural damage patterns. Moreover, CNNs have been shown to be very effective in the structural damage severity classification task when the classification is in line with the official guidelines, such as the AeDES scale for the Italian scenario. This result was achieved by applying transfer learning and data augmentation to deal with the problem of the dataset size [25, 93, 94].

3.5 Transfer Learning and Data Augmentation

One of the most critical challenges in damage classification is the scarcity of labeled data. Transfer Learning (TL) mitigates this limitation by initializing models with weights pretrained on large datasets such as ImageNet. Fine-tuning is then performed on domain-specific datasets, leading to improved accuracy and faster convergence [25].

A major limitation in vision-based post-earthquake damage classification is the scarcity and imbalance of labeled data, which hinders the ability of deep networks to generalize across diverse structural scenarios. Transfer Learning (TL) provides an effective solution by exploiting models pretrained on large-scale datasets, such as ImageNet, where millions of images across thousands of categories are available [79]. The pretrained weights encode low-level and mid-level visual features (e.g., edges, textures, shapes), which are highly transferable to structural images. By fine-tuning the final layers on domain-specific datasets (e.g., Φ -Net or INGV-DFM), TL accelerates convergence, improves accuracy, and reduces the amount of labeled data required [25, 27]. Depending on the application, TL can be applied as feature extraction—keeping most of the pretrained network frozen—or as full fine-tuning, where all layers are adapted to the new task.

Deepening on the number of domain-specific datasets (e.g., Φ -Net or INGV-DFM), TL can achieve better accuracy, faster convergence, and reduced labeled data requirements by fine-tuning the last layers only [25, 27]. It is worth mentioning that based on the application, TL could be done as a mere feature extraction—most of the pretrained network is left untouched—or as full fine-tuning, where all the layers are adapted to the new task. Along with TL, Data Augmentation (DA) methods are trying to enhance the training set artificially and introduce larger variability into the training phase, thus reducing overfitting and increasing robustness. Traditionally, the augmentation operations may include horizontal and vertical flipping, random rotations, scaling, translation, and cropping. Variations in brightness, contrast, hue, and saturation, which are frequently used to reproduce changes in illumination and acquisition conditions, are the photometric transformations.

More advanced augmentation methods may inject Gaussian noise, apply blur or sharpen-

ing filters, or perform perspective distortions, which help simulate real-world imperfections such as motion blur, occlusions, or partial visibility of damaged elements.

In the classification of seismic damage, DA is very important to bring balance to the categories that have a small number of samples (e.g., moderate damage), thus, ensuring that CNNs obtain a sufficiently diversified set of examples. Recent research emphasizes that the integration of TL with a large-scale DA regime generates state-of-the-art results in building damage recognition, thus, representing a trustworthy plan for solving the twin problems of data scarcity and class imbalance [73, 94].

3.6 Review of Damage Classification Studies

Vision-based structural damage classification has rapidly evolved with the advent of deep learning, particularly convolutional neural networks (CNNs). The initial studies have already shown CNNs' potential in implementing the crack detection process, one of the most simple visual inspection tasks without human aid. For instance, Cha et al. (2017) showed a CNN-based concrete crack detection method with which they managed to separate the cracked from the non-cracked surface with high accuracy. [9]. At that time, Gopalakrishnan et al. (2017) trained CNNs by transferring the learning graphs for the detection of pavement distresses, thus, emphasizing the significance of the use of pre-trained models in the field of civil infrastructure [33]. These CNN-based methods harvested the same hierarchical features for the structural components as those obtained by human experts and indeed, could significantly enhance the performance of conventional image-processing or handcrafted feature methods on damage identification tasks. Following work not only confirmed the possibility of identifying cracks as a binary class but also opened the door to more detailed multiclass damage classifications and localizations. Spencer et al. (2019) recognized the limitations of CNN-based techniques and declared that as the main topic of discussion in their overview of computer vision applications in infrastructure technologies. Pan and Yang (2019) gave a structural element damaged severity quantification algorithm besides an object-detection algorithm that could even estimate the repair costs. Hoskere et al. (2018) [41] has crafted a multiscale CNN which brought together the benefits of a classifier and a segmenter (both based on the ResNet and VGG architectures) for automating post-earthquake building inspections [41]. In a similar vein, Hong et al. (2019) leveraged deep CNN techniques for processing of the post-disaster aerial survey, successfully accomplishing the task of damaged building identification from the drone or satellite images. Liang et al. (2019) [56] presented the failure classification at the system level utilizing a pre-trained VGG16 CNN, along with a Faster R-CNN for detection at the component level and a fully convolutional network (FCN) for pixel-wise segmentation. This multi-tiered approach demonstrated the potential of deep learning in simultaneously tackling different scales of the damage detection problem. A key challenge for training deep models in structural health monitoring (SHM) has been the limited availability of labeled damage images. Gao and Mosalam tackled this with two influential contributions. First, they showed that deep transfer learning could significantly improve image-based damage recognition when data are scarce [25]. After that, Gao and Mosalam (2020) developed the PEER Hub ImageNet (Φ -Net or Phi-Net), an annotated benchmark dataset of photographs of structures with multi-attribute labels to foster automated visual recognition research [26]. Φ -Net released images depicting the development of damage under various scenarios, materials used, and situations, thus

enabling the researchers to have access to a standardized data resource which is very vital for them. Azimi et al. (2020) carried out a comprehensive review of the application of deep learning in SHM, that involved both vibration-based methods and vision-based damage detection [4]. As far as images are concerned, it was found that most of the research works were based on the explanation of the surface of the materials with the help of small cracks. These works usually employ simple image processing techniques as a starting point and then move in the direction of the end-to-end CNN solutions. Azimi et al. observed that design of CNN-based methods were not limited to the three types of damage represented in this study: besides cracks in concrete and pavements, corrosion detection, spalling detection and bolt loosening detection have also joined the field of research. Several researchers have come up with new CNN architectures that are designed to match the features of the new infrastructure inspection research problems. For example, Fan et al. built a multi-label CNN for identifying and quantifying the physical properties of cracks, while Liu et al. created FPCNet for the on-the-spot crack detection. Efforts have been made to augment the traditional visible light-based detection system with the use of infrared. One way is the deployment of attention-based U-Net architectures for object extraction and another is the image fusion of visible and infrared. Ferraris et al. (2023) gathered a comprehensive overview of the recent applications of vision-based structural health monitoring (SHM) [24]. Their last table summarizes more than 200 publications, disclosing that a sizable portion of deep learning-based SHM studies considers relatively simple classification tasks. Correspondingly, there were significantly fewer research works that dealt with multi-class classification due to the increased complexity and data requirements. Ferraris et al. furthermore observed the predominance of accuracy as the main evaluation metric, as well as the general trend in the employment of transfer learning utilizing pre-trained models such as VGG, ResNet, and the like. Transformer-based models have just recently appeared on the stage. Gao et al. (2023) have come up with a Multiattribute Multitask Transformer (MAMT²) that integrates classification, localization, and segmentation [28], and on the Φ -Net dataset, it achieves better results than the rollover traditional CNN pipeline. To sum up, the damage classification employing CNN as a base has turned into one of the fundamental pillars of today's SHM, thus making inspections much quicker and less subjective. The very first studies were concentrated on the binary crack detection but now the field has progressed to the level of complex multi-task models even including transformer-based ones. The emergence of standardized datasets such as Φ -Net and the acknowledgment of this trend by the comprehensive reviews of Azimi et al. (2020) and Ferraris et al. (2023) are clear indications that the use of deep learning in vision-based SHM is the future [24, 4].

Indeed, these contributions underpin the growing consensus that CNNs are the first-choice tools for rapid and scalable image-based structural damage classification, especially when combined with high-quality datasets and domain-specific fine-tuning.

3.7 CNN Architectures Used in Damage Assessment and Concluding Insights

Below are some CNN architectures that are commonly adopted for classifying damage in after earthquake images.

- **AlexNet:** It was one of the first deep CNNs that made a breakthrough, it is compatible with small datasets though it has a depth and feature richness limitation.

- **GoogLeNet (Inception):** Has multi-scale convolutional filters that can be used in different layers. It is energy-efficient and lightweight, particularly for a mobile application.
- **VGG16, VGG19:** These are admirable due to the clarity and homogeneity of their architecture with usage of stacked 3×3 convolutions. Most of the time, they are engaged as the main source of features in transfer learning application.
- **ResNet50/101:** The residual connections concept is used here in order to make training of very deep networks smoother. It is on top of the performance charts in many classification tasks.
- **DenseNet201:** Each layer is connected with all the other layers to create a deeper network that ensures more efficient feature propagation and uses fewer parameters.
- **XceptionNet:** With separable convolution for each layer in the depth dimension as the core concept, this network is very capable yet it is computationally only a fraction of the full original network.
- **MobileNet (V1, V2, V3):** Designed for mobile and edge devices, MobileNet uses depthwise separable convolutions (and in later versions inverted residuals / linear bottlenecks) to reduce the number of parameters and computational cost, trading off a small amount of accuracy for large gains in efficiency.
- **EfficientNet (e.g. EfficientNet-B0 ... B7):** Introduces a compound scaling method, simultaneously scaling width, depth, and resolution of the network in a principled way, to reach state-of-the-art accuracy with fewer computational resources compared to many earlier architectures.

In summary, CNNs have become the foundation for modern approaches to damage classification from RGB imagery. While traditional ML methods remain useful for constrained problems, deep CNNs offer a flexible and powerful solution that can generalize across building types and image conditions. However, performance remains dependent on dataset quality, class balance, and the effective use of transfer learning.

Table 3.1. Deep Learning applications using CNNs for structural health monitoring problems.

Model	Application	Related works
VGG16, VGG19	– Crack detection – Corrosion detection – Steel damage condition assessment – Post-earthquake damage assessment	[25], [26], [41], [72]
AlexNet	– Concrete crack detection – Damage classification for masonry structures – Post-disaster inspection	[4], [51], [111]
GoogLeNet	– Crack detection – Post-disaster inspection – Inspection for bridges	[4], [111]
ResNet (20, 50, 101, 152)	– Crack detection – Structural inspection – Building damage classification	[25], [26], [41], [72], [80]
MobileNet	– Building damage classification – Road damage detection	[4],[72]
InceptionNet	– Crack detection – Masonry building damage classification	[4],[72]
DenseNet	– Post-earthquake aerial damage assessment	[40]
XceptionNet	– Damage classification	[72]

Chapter 4

Image Datasets for Post-Earthquake Building Damage

In post-earthquake scenarios, large collections of labeled images are useful for developing and testing computer vision models that can recognize structural damage. This chapter reviews several important image datasets specifically compiled for post-earthquake building damage assessment or similar tasks. Both international and Italian initiatives in this domain will be discussed. **Section 4.1** introduces the PEER Hub ImageNet (Φ -Net) dataset, a pioneering large-scale structural image database developed at UC Berkeley. **Section 4.2** presents the INGV Macroseismic Photographic Database (DFM), an Italian archive of earthquake damage photos. **Section 4.3** highlights the recent IDEA dataset (Image Database for Earthquake Damage Annotation) created by EUCENTRE in Italy, which provides a new repository of annotated damage images. **Section 4.4** highlights additional and ancillary data sources, e.g., the Italian Network of University Laboratories for Seismic Engineering (ReLUIs), the Italian DaDO and IRMA Platforms and others. In the end, **Section 4.5** describes the new dataset compiled for this research, which merges and refines data from some of the above sources to better suit our study's needs in the next chapter of the thesis. Each section will detail the dataset's origin, content, and relevance, laying the groundwork for the use of these images in training deep learning models.

Beyond these, other structural health monitoring (SHM) datasets, especially those oriented towards vibration-based damage detection, damage localization in bridges, or long-term infrastructure monitoring, also exist. For a comprehensive overview of such datasets and data-driven deep learning approaches in SHM, see the state-of-the-art review by Azimi, Dadras Eslamlou and Pekcan [4]. This review by Azimi catalogs many important datasets employed in vision-based and vibration-based structural monitoring, including crack detection datasets, time-series sensor logs, UAV-derived images, and digital twin testbeds.

However, in this thesis the focus is specifically on datasets that meet the following criteria:

- They are **publicly available or open source**, ensuring reproducibility and accessibility;
- They contain **post-earthquake building damage images**, rather than laboratory simulations or non-destructive test artifacts;
- They align with our objective of **damage severity classification**, particularly in the style of AeDES-like labeling (slight, moderate, heavy).

Therefore, although numerous SHM datasets exist in the broader literature, our discussion is restricted to the most relevant image repositories that satisfy the above conditions and are widely recognized in post-earthquake damage assessment research.

4.1 PEER Hub ImageNet (Φ -Net) Dataset

One of the first large-scale endeavors to create a structural damage image database is the PEER Hub ImageNet, commonly abbreviated as Φ -Net (Φ -Net), guided by the Pacific Earthquake Engineering Research Center (PEER) at the University of California, Berkeley. PEER is a leading earthquake engineering research consortium, and it recognized the need for big data in vision-based structural health monitoring.

The development of the Φ -Net dataset by Gao and Mosalam is grounded in a series of prior efforts in post-earthquake damage documentation, structural performance assessment, and computer vision-based inspection of civil infrastructure. In fact, their work builds upon field observations from major seismic events [99, 55], the integration of laser scanning in structural surveys [70], and recent advances in deep learning for structural damage recognition [25]. Actually, one of the major reasons for the creation of the Φ -net dataset, which goes beyond the objectives of the paper, is the need acknowledged in the whole research field represented by the references for dependable visually annotated benchmarks to serve as a basis for the classification of damages by automated methods. Such studies not only provided the inspiration for creating the Φ -Net dataset but also the extent and its purpose, which is mainly to offer a common and complete repository of images to facilitate the training of deep learning models and their benchmarking in the field of earthquake engineering applications [26].

Summarizing, before Φ -Net, progress in automated post-disaster damage recognition was hampered by two key issues:

- the lack of a general automated detection framework based on structural domain knowledge;
- the lack of benchmark datasets with large amounts of well-labeled structural images[25].

To cover these missing points, Gao and Mosalam (2019) [26] developed Φ -Net as an open-source repository of structural images that aim to facilitate the use of machine learning and deep learning techniques for damage assessment.

4.1.1 Dataset Scope and Size

The first edition of the PEER Hub ImageNet brought together about 36,413 images of civil infrastructure and buildings, each annotated with several attributes describing the earthquake damage and structural characteristics, as the dataset scope and size were concerned. It is, thus, an open dataset of structural images with the largest number of such to date, and the decision to make it publicly accessible was made to promote research collaboration. Initially, the dataset was divided into eight subsets (one for each classification task defined by the authors) and made available for non-commercial use only. Researchers can access these image sets (i.e., pre-packaged as tensors for convenience) and download them without any restrictions from the PEER Hub ImageNet website [29]. This accessibility makes Φ -Net

an important reference dataset for training and testing computer vision models in structural engineering. One of the characteristics that distinguish Φ -Net is its multi-attribute labeling framework that deals with damage recognition as a hierarchical tree of classification tasks.

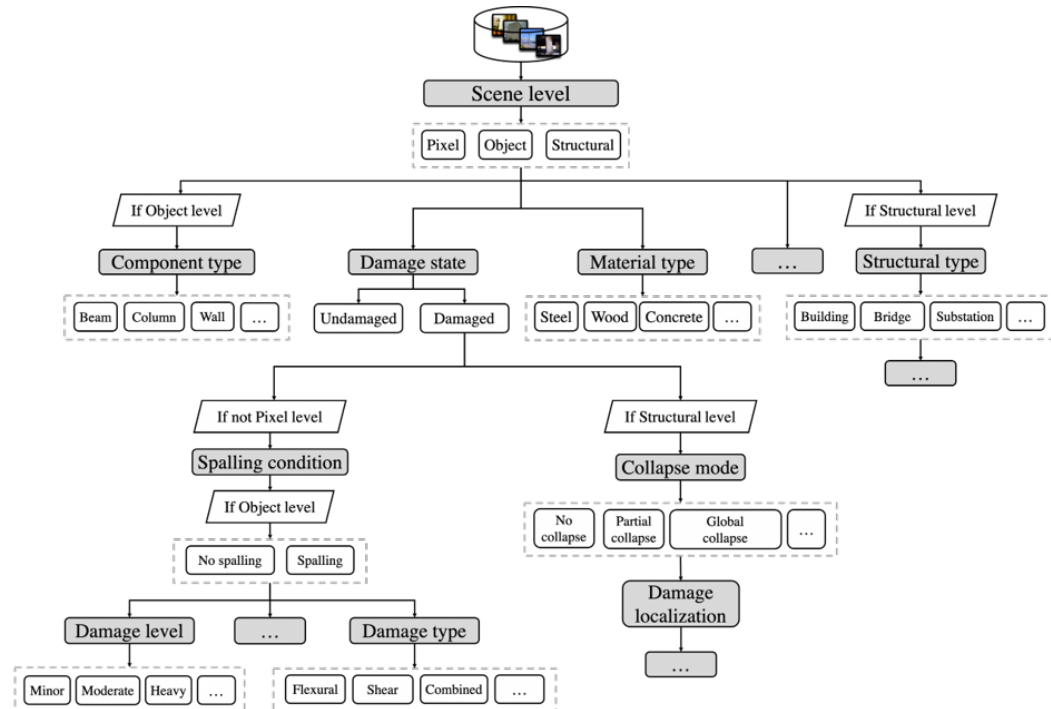


Figure 4.1. Scheme of classification task hierarchical tree for Φ -net dataset. Image from [29].

Drawing from past earthquake reconnaissance experience, Gao and Mosalam defined eight benchmark classification tasks that an image can be labeled with, covering various aspects of the scene, damage, and structural components. In summary, the Φ -Net annotations include: Scene Scale Classification: Whether an image is taken at the pixel level, object (component) level, or structural (building) level.

This is the main difference between close-up detail images and those showing the general structure of a building.

- **Damage Presence (Damage Check):** A binary label indicating if the structure or component in the image is damaged or undamaged.
- **Spalling Condition:** A binary label for material loss, distinguishing images with no spalling versus those showing spalling (concrete cover peeling).
- **Material Type:** A binary category indicating the primary material, typically steel versus other (e.g., concrete or masonry).
- **Collapse Level:** A three-class classification of collapse status – no collapse, partial collapse, or full collapse of the structure or component in the image.
- **Component Type:** A four-class classification identifying the structural component shown, e.g., beam, column, wall, or other structural element.

- **Damage Severity Level:** A four-class scale for the extent of damage – typically none, slight/minor, moderate, or heavy damage in the element/structure.
- **Damage Mechanism Type:** A four-class category describing the failure mode, e.g., no damage, flexural damage, shear damage, or combined (complex) damage.

Every Φ -Net image is linked to multiple textual descriptors (one label maybe present in each of the aforementioned categories wherever applicable) making it a multi-label dataset, with variety of classification tasks. An image, for instance, may have labels such as a "structural-level" picture, "damaged" concrete building, "partial collapse", and that the "wall" is the damaged component with "heavy" shear damage as the damage mechanism. The mechanism of hierarchical labeling that the automated algorithm can move through the tree of classifications to find out all the properties of a particular image. The company focused on a structured approach that took inspiration from the success of the ImageNet challenge hierarchy in computer vision, but specifically designed for structural engineering needs.



Figure 4.2. Example of images for Damage Severity Level classification task. Images extracted from the Φ -net dataset [93]

Notably, Φ -Net was the first open dataset in this domain to provide such rich, multi-attribute labels at scale, establishing a benchmark for vision-based structural health monitoring research.

4.1.2 Image Sources and Collection

Compiling over 36,000 images of earthquake damage required aggregating data from many sources. The Φ -Net collection is truly diverse in origin. A significant portion of images were gathered from online search engines and digital libraries.

To begin with, developers extracted photos available publicly online that depicted structural damage as the main source of data. For instance, they used Google Images and Baidu Images (the latter to ensure coverage of non-English, especially East Asian, sources) to get diverse post-earthquake pictures. On top of that, Φ -Net took advantage of image archives in the field of earthquake engineering that were already there like NISEE, the National Information Service for Earthquake Engineering e-library hosted at UC Berkeley, which has the earthquake reconnaissance photos collection from the past. Moreover, other such research repositories, for example, NEEShub (Network for Earthquake Engineering Simulation), and EERI's collections, were also merged. PEER has brought photos taken by its own teams during field surveys, online databases, and web searches together to have a PHI dataset that is representative of different kinds of buildings damaged due to various earthquakes to that extent. In addition, PEER implemented a community-driven strategy to

broaden the dataset. It arranged the PHI Challenge, which is a structural damage location competition using images, back in 2018. The training set of images from Φ -Net along with the labels were given to the contestants; they were also allowed to upload their images and create models for damage detection. Thus, the competition heightened the interest of algorithmic developers and at the same time, jars more image data out of the crowd. This not only spurred algorithmic development but also crowdsourced more image data. The “Call for uploading images” as part of the PHI Challenge invited engineers and researchers worldwide to contribute their post-earthquake photos to the growing repository. Through such efforts, Φ -Net became an extensive collaborative dataset, with images drawn from earthquakes around the globe (e.g., 1999 Kocaeli, 2008 Wenchuan, 2010 Chile, 2011 Japan, etc., as indicated by its references) and a variety of structural materials and damage patterns.

Table 4.1. Number of images by damage level.

Set	No Damage	Minor	Moderate	Heavy	Total
Training	1551	869	799	919	4138
Test	207	93	104	94	498
Total	1758	962	903	1013	4636

4.1.3 Use and Limitations

The Φ -Net dataset has been the basis of all the recent researches on the automation of the categorization of the destruction. These are the main results of the experiments conducted by Gao and Mosalam. They confirm the applicability of deep learning models (e.g., ResNet and VGG) to the kinds of problems in Φ -Net, thus bringing about the accuracy encouraging for such types of issues as crack detection and identifying the component. Nevertheless, there are some restrictions related to Φ -Net, which limited its adoption. To begin with, although it comprises a large number of images, the labeling was a colossal task mainly performed by the research assistants and volunteers who may have their own interpretations of the labeling criteria. For example, the damage severity labels may not be a stringent embodiment of engineering based post-earthquake assessment scales as they were defined vaguely (none/minor/moderate/heavy) with an implicitly defined FEMA scale standard. Secondly, the pictures of Φ -Net vary from the very small details of the cracks to the whole frame of a collapsed building, which is great for the dataset’s diversity but can be a problem when a model is being prepared only for few particular tasks (for example: a safety level evaluation of the building). Because of that, there could be many images that are not directly addressing that specific question. To put it simply, if someone uses the raw PHI data for a focused classification problem, then task ambiguity and class imbalance could be the issues. The last thing is that not all images in PHI are in high resolution or solely focused on structural damage (some include surrounding context or non-damage scenarios for completeness), so it means that a model may still require the careful selection of this data. In spite of the difficulties, Φ -Net is still a landmark dataset and a very valuable starting point for large training set creation in this domain.

Its creation has undoubtedly spurred further data collection efforts with newer datasets like INGV’s DFM and the IDEA dataset. Further examples and detailed discussions regarding

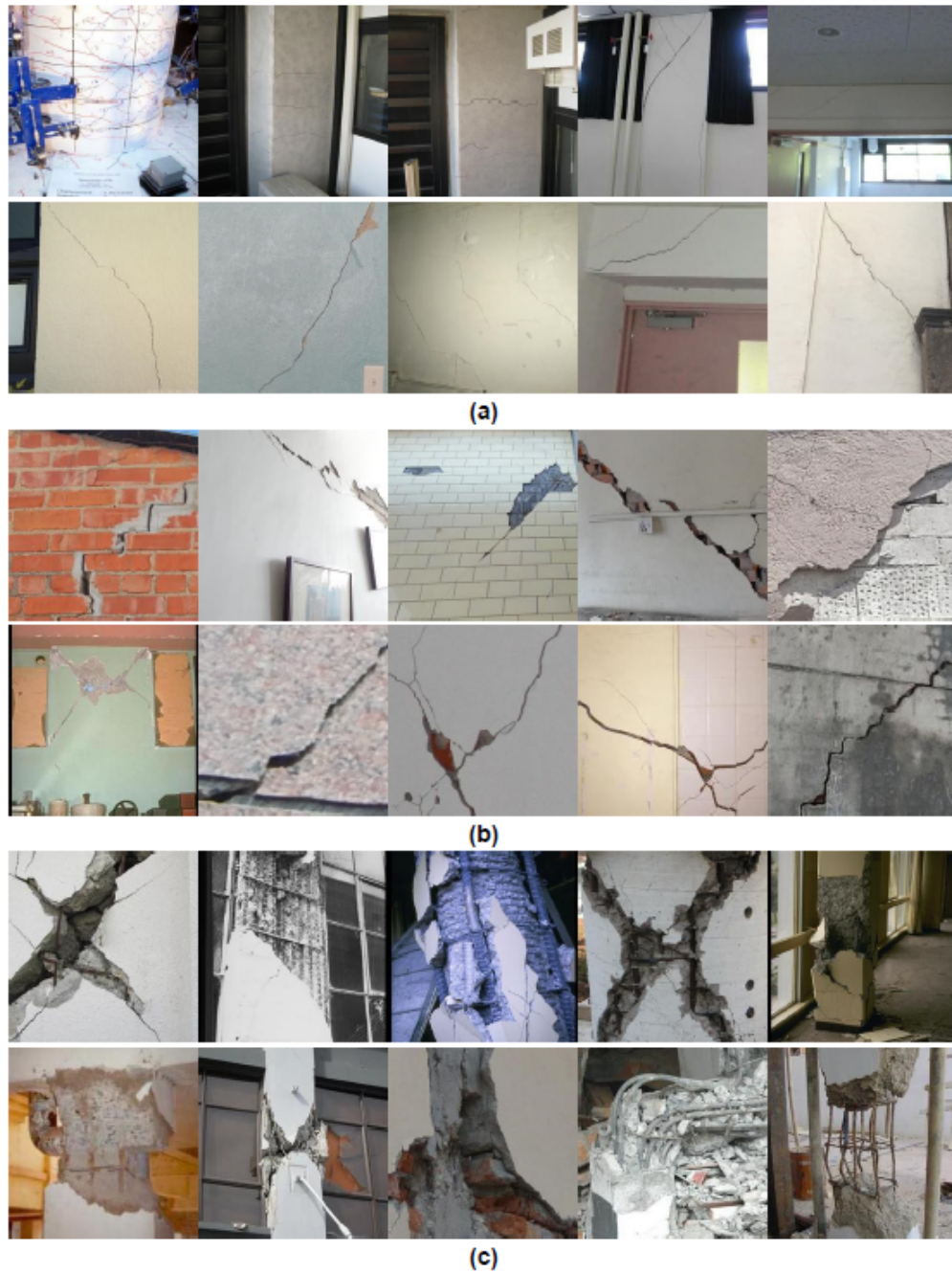


Figure 4.3. Example of images for Damage Severity Level classification task. Images extracted from the Φ -net dataset. (a) minor damage, (b) moderate damage, and (c) heavy damage. Image from [26]

Table 4.2. Summary of the Φ -Net Dataset Characteristics.

Aspect	Description
Origin	Developed by Gao and Mosalam at the University of California, Berkeley, as part of the PEER (Pacific Earthquake Engineering Research) initiative [26].
Content	Comprises over 2,600 RGB images of post-earthquake building damage, collected from sources such as NISEE, NEEShub, EERI, Google Images, and reconnaissance archives.
Annotation	Images are labeled using a multi-attribute framework including structural type (e.g., RC, masonry), damage severity, material, and function.
Relevance	First large-scale, multi-label, real-world dataset for training and benchmarking computer vision models in post-earthquake damage assessment.
Usage	Widely adopted as a benchmark for classification tasks using CNNs and other deep learning architectures due to its diversity and open accessibility.

the limitations of this dataset will be provided in the subsection Section 4.5.1, which serves as a foundational basis for the development of the new dataset introduced in this research. The new dataset was specifically designed to address and overcome several of the critical issues identified.

4.2 INGV Macroseismic Photographic Database (DFM)

Though Φ -Net furnishes an extensive, global array of pictures, Italy has assembled its own all-inclusive photographic database with the aim of documenting the disaster caused by the earthquake. The Istituto Nazionale di Geofisica e Vulcanologia (INGV), Italy's National Institute of Geophysics and Volcanology, has developed the Macroseismic Photographic Database (Database Fotografico Macrosismico, DFM) to set up a centralized repository of post-earthquake damage photographs. INGV preserves the DFM archive that is accessible through the DFM platform at the website dfm.ingv.it [48]. INGV is the first and foremost scientific research institute in Italy that keeps a watch on seismic activities and is in charge of the post-earthquake scientific studies. For a period of several years, in the process of repeatedly dealing with the consequences of earthquakes in Italy, INGV researchers have gathered an extensive number of field photos of damages to buildings and infrastructure. DFM was initiated as a method of both scientists and general public having an access to that visual encyclopedia of pictures. The first part of the text will explain the background, the need, and the purpose of this project. It was in late 2023 that the DFM was introduced as an open and accessible web platform for the seismic image heritage of Italy. The database is available on a dedicate website [48], which allows users to view the photographs according to the earthquake event, location, or the type of damage. The DFM project is a source of motivation for the scientific as well as for the educational aspects behind it; it is one of the very few records available of the earthquake impacts, while on the other hand, it is also an extremely valuable dataset for research in the area of damage analysis. INGV's QUEST team (Quick Earthquake Survey Team), a specialized group of seismologists and engineers who perform rapid post-earthquake surveys, is responsible for collecting most of these images

during their field missions. The database therefore represents the result of years of systematic post-earthquake surveys conducted by expert teams across Italy. In addition, there are also some pictures depicting the effects of earthquakes in Italy dating back to the late 1970s, which means that the visual recording of the earthquake effects by the Italian scientists has been going on for quite a time.

- **Contribution and Arrangement:** Currently, the INGV DFM is a multimodal image collection of over 1,000 plus pictures from various historical earthquakes in Italy and is constantly updated with new data from current events. These ground truth data include such historical earthquakes as the 1976 Friuli earthquake, 1980 Irpinia, 1997 Umbria-Marche, 2009 L'Aquila, 2016 Central Italy (Amatrice/Norcia), and a few others for which seismological teams of INGV have done damage surveys. The photographs show various damage scenarios: for example, masonry buildings that have collapsed or cracked, reinforced concrete frame structures in which the failure of the supporting columns occurs, historical monuments with cracks or partial collapses besides the environmental effects like landslides or ground cracks. The distinguishing feature of the DFM is the very detailed categorization conducted according to the macroseismic intensity. All the pictures stored in the database are identified with references to an intensity judgment in the framework of the European Macroseismic Scale (EMS-98). The EMS-98 scale is a European scale that is employed to rate the severity of an earthquake qualitatively (from I to XII) on the basis of the damage caused and people's perception. The DFM, by assigning each picture an intensity level of the EMS-98 scale (or range), gives a scientific context for the observed damage. For instance, a photo of the collapse of thick walls can be characterized by intensity VIII or IX (heavy damage), whereas images showing hairline cracks in walls can be related to intensity V or VI (light damage). This solution was intentionally picked for a purpose to keep the archive accessible not only to the specialists (who can assess the intensity levels) but also to non-experts who can receive a clear and standardized presentation of damage severity with each photo. Essentially, DFM links every picture to a quantitative damage metric (macroseismic intensity) that is very valuable when one wants to train machine learning models which goal is to estimate the intensity or damage grades just by looking at the pictures. Besides, DFM's adoption of EMS-98 ensures also compatibility with both Italian and European earthquake engineering practices as macroseismic intensity continues to be the most suitable way of communication in the field of seismic impacts. For example, scientists working with the DFM could select images by intensity level to find appropriate training data for a classifier that recognizes intensity from pictures. Furthermore, the DFM is not only organized event-wise but also geographically, thereby enabling studies that establish the relationship between damage patterns and seismic parameters or construction types. All the images have accompanying metadata like the date and the location of the earthquake, and usually, a description of the damage observed.
- **Access and Importance:** The DFM platform is publicly accessible and is meant as a long-term archive. INGV, by making these images openly available, hopes both to preserve the photographic heritage of Italy's seismic history and to encourage research in earthquake engineering and risk mitigation. The project is one of the many benefits of visual data in the training of new post-earthquake analysis and automated tools. We, on the other hand, can take advantage of the DFM as a source of numerous and diverse

experts' images illustrating building damage in the Italian context. The pictures most of the time highlight the affected parts of the structure (e.g. the cracked wall or the broken column), and these can be perfect for training convolutional neural networks to figure out the damage present. Also, since DFM combines data from multiple earthquakes, it creates a more diversified dataset that shows the different building types (unreinforced masonry, reinforced concrete, historical structures) and damage stages, all over the Italian building stock. In other words, this is a great set for training models that will be used in Italy or areas with similar construction styles.

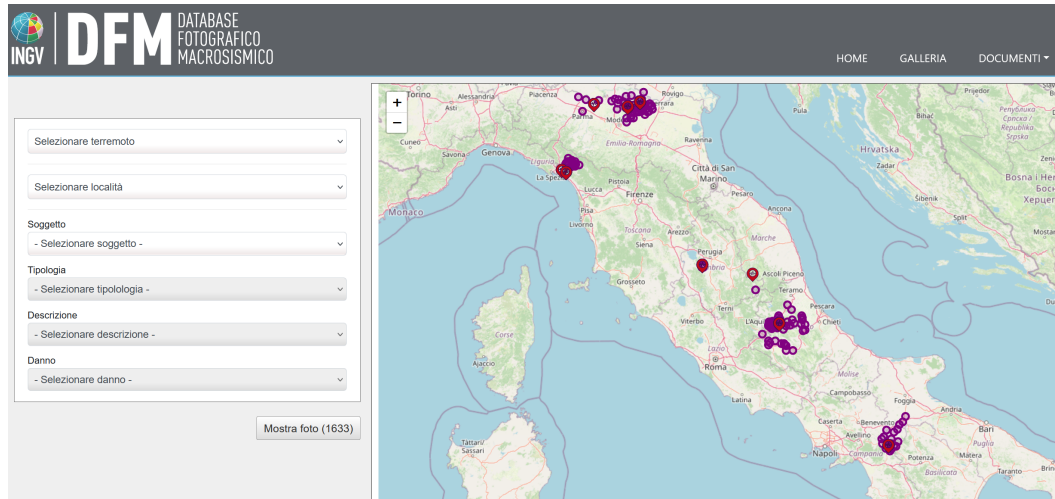


Figure 4.4. Example of images selection from the italian available earthquakes on the DFM platform [48]

The INGV DFM platform provides a structured and interactive interface for accessing post-earthquake photographic documentation in Italy. Through its online portal [48], users can query the archive by selecting a specific earthquake event, the affected locality, and the subject type, such as "building", to narrow down the visual records. Subsequently, it is possible to refine the selection by specifying the building typology (e.g., masonry, reinforced concrete, or other), as well as additional metadata such as the structural function and a qualitative damage description.

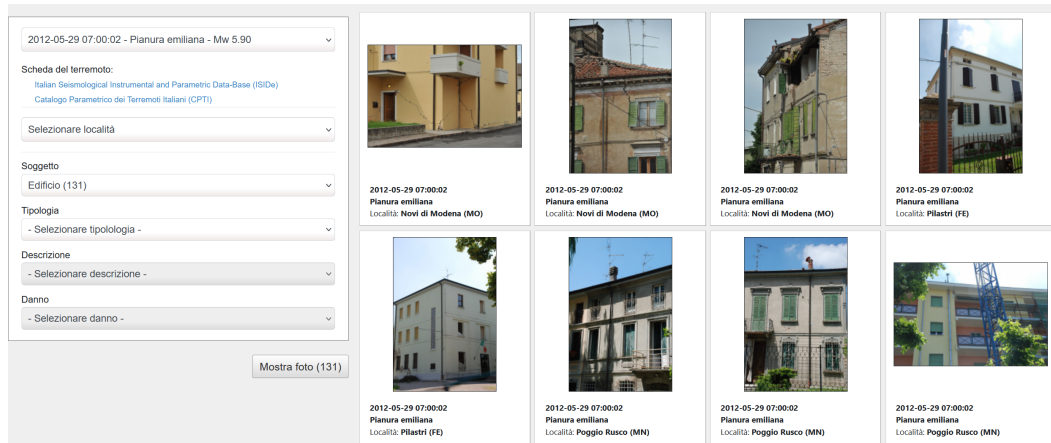


Figure 4.5. Images selection from the Pianura Emiliana earthquake in 2012 (Mw 5.90) on the DFM platform [48].

Table 4.3. Summary of the DFM (Macroseismic Photographic Database) by INGV

Aspect	Description
Origin	Created and maintained by INGV (Istituto Nazionale di Geofisica e Vulcanologia) as part of national seismic response initiatives. Available through the official portal [48].
Content	Contains thousands of geo-tagged post-earthquake damage photographs collected across Italian seismic events, from historical earthquakes to recent crises such as L'Aquila (2009), Amatrice (2016), and others.
Search Filters	Users can query by earthquake event, affected locality, subject type (e.g., building), construction typology (e.g., masonry, RC), structural description, and damage description.
Relevance	Provides detailed, expert-curated photographic documentation aligned with national seismic assessment practices, particularly the AeDES form.
Usage in this Work	Used as a supplementary image source to build the new dataset. Selected images were manually cropped to isolate the damage, resized, centered, and relabeled based on AeDES categories.

This structured flow facilitates targeted visual assessment of damage scenarios, supporting both post-event evaluation and research applications. Alternatively, the platform allows users to browse the full set of available events and all documented structures. This broader approach was adopted in our research to expand the visual dataset. A large number of relevant images were downloaded and manually curated. Each image was cropped to highlight only the damaged area of interest and the corresponding structural component. The resulting regions of interest were then centered, squared, and resampled to a consistent input size suitable for training convolutional neural networks. Finally, each image was re-labeled according to the official AeDES standard damage levels, i.e. Slight, Moderate, or Heavy, to ensure consistency with the supervised classification framework developed in this work. In summary, the INGV DFM is a specialized post-earthquake image dataset that complements broader datasets like Φ -Net. It brings a level of standardization (through intensity labeling) and regional specificity (focus on Italian earthquakes) that is extremely useful. In this research, the DFM will be leveraged as one of the key sources of training images, especially

when fine-tuning models to recognize damage severity in a manner consistent with Italian post-earthquake survey standards.

4.3 IDEA: Image Database for Earthquake Damage Annotation

A very recent contribution to the field of earthquake damage imagery is the IDEA dataset, "Image Database for Earthquake Damage Annotation". This dataset was introduced in 2025 by a team at the EUCENTRE Foundation (European Centre for Training and Research in Earthquake Engineering) in Pavia, Italy, led by Ilaria Senaldi et al. [98].

The IDEA dataset represents a modern, high-quality collection of earthquake damage images, distinguished by its detailed annotations geared towards training advanced computer vision models. Overview and Scale: IDEA is described as "an extended dataset of annotated real-world structural damage images". In total, it contains over 5,400 RGB images of damaged structures, which were collected during post-earthquake reconnaissance surveys as well as routine (ordinary) structural inspections. This means the dataset not only includes photos from major earthquake events (post-disaster field missions), but also images from general building inspections where various damage or deterioration might be observed. All images in IDEA have been expert-annotated by structural engineers for the presence and characteristics of damage.

Notably, the dataset spans different structural typologies and materials, for example, reinforced concrete buildings, masonry buildings, bridges, etc., and captures a variety of damage mechanisms. This breadth makes IDEA one of the most comprehensive image datasets for structural damage to date in terms of content diversity.

The "Annotation" in IDEA's name highlights that it is not just a collection of photos, but each image is paired with rich metadata or labels about the damage it depicts. The creators developed a refined ontology for structural damage to guide the annotation. According to the dataset description, the primary labeling identifies structural elements and specific damage types in the image. They report a "secondary, broader classification of damage" with categories including: Crack, Spalling, Corrosion/Exposed Rebar, Leaching, Debris, Other Damage/Mechanism, and Non-Structural Damage.

In reality, the annotators went through each image and decided which damage phenomena from the list were visible and probably localized them. So, a picture of a cracked concrete column with spalled cover could be annotated with "Crack" and "Spalling", and also terms like the structural element (column) and the material (reinforced concrete) could be there. The presence of such categories as "Non-structural damage" in the dataset means that the collection of damages to the likes of infill walls, partitions, chimneys, etc., is also considered, which are the most minor of all the damages that can still be helpful in post-earthquake loss assessment. These detailed annotations essentially make the IDEA dataset capable to support not only classification but also object detection or segmentation tasks. Actually, the existence of multiple instances (e.g., multiple cracks or spalls in one image) and the reference of an ontology gives a hint that there could be bounding boxes or masks for each damage instance. The intention of the dataset is made clear, that is, to allow the training of deep learning models for the automated damage detection process. In other words, IDEA was curated with state-of-the-art computer vision applications in mind, especially tasks like detecting cracks on a building or identifying regions of spalled concrete in images. This is a step beyond earlier datasets like Φ -Net which focused more on image-level labels. IDEA

moves towards instance-level annotations, which are crucial for developing detectors (e.g., using algorithms like Faster R-CNN or YOLO for object detection of damage).

The IDEA dataset is published as a Data Article (Data in Brief, 2025) and is hosted by the EUCENTRE Foundation, with a DOI provided for reference.



Figure 4.6. IDEA dataset: example of a typical building with instances of damage annotated using rectangular bounding boxes and polylines. On the right: content of the annotation file. On the left: original image (top) and annotated image (bottom). Image from [98]

The dataset is indexed on platforms like Zenodo (an open data repository), although at the time of writing the actual image files are under controlled access (likely to manage usage rights). It was developed with support from the Italian Department of Civil Protection and through various EUCENTRE research projects. The involvement of surveyors and structural engineers from EUCENTRE’s teams is gratefully acknowledged in the dataset release, indicating a robust, vetted annotation process. The significance of IDEA is that it provides a new, high-quality benchmark for the research community. Early studies have already started to use IDEA to train and evaluate models. For example, Dondi et al. (2025) report experiments where a damage detector trained on a combination of real images from IDEA and computer-generated “synthetic” damaged images outperformed a model trained on real images alone [20]. This suggests that IDEA’s data, when augmented properly, can lead to highly effective damage recognition models. The availability of around 5000 expertly annotated images is an important asset, given that the lack of annotated data was a long-standing hurdle in this field. However, because IDEA has only been released in 2025, it is still very new. It will likely become a key resource in the near future, but many current studies (including our own) were already underway using earlier datasets. In our research, we primarily utilize the Φ -Net data (which has been available since 2019) and other open sources to train our models. The IDEA dataset was not fully available or integrated at the

start of our project. Moreover, for reasons of continuity and comparability with prior works, we focused on Φ -Net as a baseline. Nonetheless, the emergence of IDEA is acknowledged as an important development. In future extensions of this work, incorporating the IDEA dataset could significantly improve model robustness, because it offers higher annotation quality and a well-defined taxonomy of damage types. For the current thesis, we use IDEA more as a point of reference (to validate the relevance of the problem and potential improvements) rather than as a primary training set. Ultimately, the limitations we encountered with Φ -Net (discussed below and observed in Chapter 5) were a driving factor for creating our own curated dataset. If Φ -Net alone proved insufficient for generalized damage classification, the existence of datasets like IDEA reinforces the notion that more targeted, well-labeled data is necessary, something we have addressed by constructing a new dataset with expert annotations, described next.

4.4 Additional and Ancillary Data Sources

This section extends an overview of various supplementary datasets that have influenced image collection improvement as part of this research. Such elaborations cover the minor but strategically beneficial data sources, like ReLUIS field findings and the 2023 Turkey earthquake survey by Pujol et al., along with non-RGB content, metadata, and platforms that enrich and contextualize our core imagery-based dataset without being the primary dataset (the DaDO and the IRMA platform).

4.4.1 ReLUIS Post-Earthquake Damage Database: L'Aquila Case Study.

The Italian Network of University Laboratories for Seismic Engineering (ReLUIS) contributed for the structural damage assessment after the 2009 L'Aquila earthquake, in particular, the role it played after the 2009 L'Aquila earthquake. ReLUIS managed the on-the-go field assessments and was the technical overseer of the damage surveys carried out in the different municipalities of the affected area. ReLUIS worked very closely with the local authorities, who gave the green light for the plans of reconstruction and structural interventions [30]. ReLUIS was not a repository of images as such but was a technical consultant in charge of auditing expert damage report and verifying the inspection data at the component and building level.

ReLUIS has also played a part in offering resources for the establishment of engineering databases and made available some of the photographic evidence included in larger platforms such as the DaDO web-GIS database [19]. Moreover, ReLUIS field missions brought not only the photographs but also the architectural details and the quality of the photographs, especially the focus on masonry and RC buildings, which are the most common types of buildings in Abruzzo. This information has been given through these data collections especially for Italian case study areas, to witness the cases of structural damage and the solution of repair work by the intervention.

4.4.2 Data from the Turkey 2023 Earthquake Damage Survey (Pujol et al.).

On February 6, 2023, a doublet of earthquakes with magnitudes of 7.8 and 7.7 respectively hit southeastern Turkey and northwestern Syria, with the first shock near Pazarcik (Kahramanmaraş Province) less than an hour before the second one. The area, where this disaster

has taken place, has been shaken hard (Mercalli Intensity XI) by the quakes that led to the failure of structures over a wide range of several provinces and, therefore, the people living there were affected to a great extent [108] [10].

In their research, Pujol et al. went through a field data-based survey that includes almost 250 reinforced concrete buildings reaching from Antakya to Malatya [86]. Their quantitative evaluation indicates that a lot of buildings have undergone large lateral movements as a result of insufficient structural durability, which has consequently led to the occurrence of brittle failure, p-delta effects, and the extensive damage of the nonstructural parts of the buildings. These findings tie in with the high peak ground motion that has been recorded all over the region.

Our main analysis of RGB imagery does not heavily rely on this dataset, yet we have included some images from the survey to supplement our dataset, as described in Section 4.5. The classification of building types by Pujol et al. helps to broaden the representational range of our dataset, thereby facilitating the academic studies of related topics and the development of mitigation strategies.

4.4.3 DaDO (Database of Observed Damage)

The DaDO (Database of Observed Damage) is a WebGIS platform representing a technical collaboration project between the Italian Civil Protection Department and EUCENTRE. It gathers data about the structural features and damage observed in ordinary buildings and churches after major seismic events in Italy starting from the 1976 Friuli earthquake up to the 2012 Emilia–Romagna sequence [18]. By converting various inspection datasets into one common schema, DaDO allows scientific users to access, compare and perform statistical analyses of the earthquake damage records over different events. This functionality assists in the tuning of fragility models as well as in the seismic vulnerability assessments at the local and regional levels, thus becoming a key reference both for vulnerability modeling and for damage statistics [19].

4.4.4 IRMA (Italian Risk MAPs).

IRMA (Italian Risk MAPs) is a top-notch online platform made by EUCENTRE and supported by the Italian Civil Protection Department. The main purpose of this platform, designed for the scientific community, is to integrate hazard, exposure, and vulnerability data through the OpenQuake engine along with the MPS04 hazard model implemented by INGV for generating damage scenarios both conditional and unconditional for residential buildings in Italy [6]. Its extremely adaptable design allows users to upload their own exposure and fragility data, and also to view and download the risk and damage maps which in turn gives them the possibility of performing national-scale risk modeling compatible with the National Risk Assessment of Italy [6]. Although IRMA is scenario-based modeling tool, the use of DaDO damage data by IRMA is an implementation of the observed damage metrics in the engineering and risk planning processes.

4.4.5 Mendeley DB Concrete Crack Images for Classification Dataset.

The DB Mendeley Crack Dataset is one of the first and most influential instances of a large-scale dataset of RGB images mined for crack detection purposes. The dataset "Concrete

Crack Images for Classification" is a large-scale collection with the goal of detecting the crack. It contains 40,000 RGB image patches (227×227 pixels) evenly split between those with and without cracks on concrete surfaces. The patches come from 458 high-resolution source images (4032×3024 pixels) taken on the METU campus using the method introduced by Zhang et al. (2016). The dataset is characterized by a wide range of natural variations in the surface texture and lighting, but no artificial data augmentation was applied. Although this dataset is not intended for building damage assessment after earthquakes, it is still a helpful educational resource for training neural networks in supervised image classification tasks. [74].

4.5 New Dataset Compiled for This Research

While datasets like PHI-Net provide advantages in terms of scale and accessibility, they have limitations that are quite significant for our particular task. To be more specific, PHI-Net is a dataset that has a large number of noisy and ambiguous images, an unbalanced distribution of classes with respect to the severity of damage, and a certain vagueness in the labels. All these factors are impediments to the model's ability to achieve very high accuracy. Besides that, none of the datasets that have been compared, even PHI-Net, are compatible with AeDES classification standards, which is very important for the damage assessment of Italian engineering. The two aspects, inconsistency of the dataset and lack of standards that the dataset adheres to, are the reasons for the new, custom dataset made for reliability and high-fidelity classification, which we also present in Chapter 5.

4.5.1 Motivations and Limitations of Φ -Net

The primary reason for creating a new dataset was the realization that Φ -Net, despite its large size, had limitations for our specific task of classifying damage severity in buildings. As we will see in Chapter 5, initial models trained purely on Φ -Net data did not generalize well to certain real-world cases, the accuracy dropped when tested on images outside the PHI distribution [93, 26].

In their foundational work, Gao and Mosalam (2018) acknowledged that large, noisy, and inconsistently labeled datasets like PHI-Net may undermine classification performance. To address this, they manually curated a smaller subset, roughly 2,000 images, from the full Structural ImageNet and carefully re-labeled them for four primary recognition tasks, including damage level evaluation [25]. By applying VGGNet-based transfer learning with both feature extraction and fine-tuning, they achieved markedly improved accuracy, particularly in the three-class damage-level task, compared to training on the unfiltered collection.

However, even these results were not flawless. Firstly, Gao and Mosalam combined two of the damage classes ("moderate" and "severe"), with the moderate one being the most challenging to classify, recognizing that moderate damage was often the most difficult to classify correctly, which potentially simplifies the classification landscape. Secondly, their evaluation relied on a very small and exceptionally well-curated test set that may not generalize well to more variable conditions. Finally, the study remained entirely focused on PHI-Net imagery, limiting its applicability to other datasets or real-world Italian Aedes standards.

This is partly because Φ -Net’s broad scope (multi-attribute, global data) means it includes many images that are either not directly relevant to our focus (e.g. extreme close-ups of cracks without context, or images of non-building structures) or are labeled in ways that don’t align with standard damage scales. Additionally, Φ -Net’s damage level labels, while conceptually similar to slight/moderate/heavy, were applied in a generic sense and not verified against a formal standard like the European macroseismic scale or the Italian AeDES criteria. This can introduce inconsistency. Our task requires distinguishing damage grades in a way that aligns with how engineers in Italy assess buildings (e.g., as in the AeDES damage form used by civil protection).

These limitations underscore the necessity for a new dataset: one that achieves both high accuracy and adheres to engineering-relevant classification frameworks like AeDES.

Therefore, we decided to construct a new dataset by filtering and relabeling images from Φ -Net and merging them with images from other sources, to achieve a more uniform, task-specific collection.

4.5.2 Data Sources and Composition of the New Dataset

We aggregated images from multiple high-quality sources to form the new dataset, as shown in the following table.

Table 4.4. Number of images in the considered data sources.

Database source	Number of images
PEER Image Net Challenge dataset (NISEE, NEEShub, EERI, Google images)	2,625
INGV DFM – Database Fotografico Macrosismico (Italy)	1,537
ReLUIS reports (post-earthquake field photographs)	245
Earthquake Damage Survey (Turkey 2023, Kahramanmaraş reconnaissance images)	194
Scientific publications and related articles (damage figures, data extracts)	104
Total compiled dataset	4,705

Table 4.4 provides an overview of the image sources contributing to the compiled dataset and the number of images drawn from each. Specifically, the largest contribution comes from the *Phi*-Net, which alone provides approximately 2,625 images. This subset includes imagery originally pulled from multiple repositories such as NISEE, NEEShub, the EERI archives, and web-scraped collections from Google and Baidu. The second-largest source is the INGV DFM, the Italian Macroseismic Photographic Database, which contributes about 1,537 curated images. These photos document earthquake damage across numerous Italian seismic events and are linked to macroseismic intensity values. Next, 245 images were obtained from ReLUIS field reports—photographs captured by the Italian Network of University Laboratories for Seismic Engineering in the aftermath of post-earthquake inspections. In addition, a dataset of 194 photographs was specifically selected from publicly available sources associated with the 6 February 2023 Kahramanmaraş (Turkey) earthquake reconnaissance. A further 104 images were extracted from scientific publications, data articles, and other academic studies or minor sources, representing building damage captured

in figures or supplemental materials. Altogether, these contributions sum to a total of 4,705 images in the final compiled dataset.

As shown, the largest portion of images (over 2,600) came from the Φ -Net dataset itself. Rather than use all around 4700 Φ -net task 7 images (damage level classification task), we selected a subset of around 2,600 images that were most relevant for building damage evaluation. This involved filtering out images that were too unrelated or not useful for our classification goal, for example, images that were exclusively of bridges or non-building structures, very close-up material tests, duplicates, or images with ambiguous damage. We retained PHI images that clearly showed building components (columns, beams, walls) with varying damage states. Each selected PHI image was then manually reviewed and re-labeled by our team of structural engineering experts to fit our unified labeling scheme (described below). The next major source is the INGV DFM, from which we incorporated about 1,537 images. These were extremely valuable because they come with context of Italian earthquakes and often had high-fidelity depictions of damage with an intensity label. We ensured to include images spanning different intensity levels from DFM, from slight damage (Intensity V–VI) up to very heavy damage (Intensity IX–X), to balance the dataset across damage levels. Using DFM images also adds regional realism to the dataset, as they reflect Italian construction practices and damage patterns. We also included a set of images from ReLUIS field reports (245 images). In fact, as stated before, ReLUIS is a consortium of academic labs in Italy that often participates in post-earthquake surveys: for example, after the L'Aquila 2009 earthquake, ReLUIS teams compiled detailed damage reports. We obtained photographs from such reports, when available publicly or through our academic network, to enrich the dataset, especially for certain types of structural details or retrofitting failures documented by these reports. Additionally, recognizing the importance of making our model robust to varied seismic contexts, we incorporated nearly 200 images from the 2023 Türkiye (Turkey) earthquake surveys. The twin earthquakes in February 2023 in Turkey and Syria produced a vast number of publicly shared images of collapsed buildings, many of which have been used in recent studies and even compiled into datasets from [86].

Every image in the dataset underwent a meticulous relabeling process carried out by our institution's expert structural engineers. Leveraging their deep expertise in post-earthquake damage evaluation, they critically assessed each image and assigned damage severity levels in strict accordance with the guidelines outlined in the AeDES manual. To facilitate consistent and efficient annotation, we developed a custom labeling tool hosted on Google Colab, linked to a shared Google Drive workspace and featuring a user-friendly GUI (Figure 4.7). This collaborative setup allowed multiple team members to label images remotely while ensuring adherence to the same classification standards across the dataset.

We curated a selection of these (focusing on buildings with different collapse modes, as well as some images of buildings that survived with minor damage) to test our model's generalization to non-Italian but still relevant scenarios (modern concrete buildings, etc.). Lastly, we scavenged academic literature and other data articles for any usable images (around 100). For instance, if a published paper on damage detection included example photos or if a data repository (like Mendeley Data) hosted some earthquake damage image collections, we took advantage of those. These often provided unique cases, such as infrequently seen failure modes or images from historical earthquakes not covered elsewhere. Bringing all these sources together, the final compiled dataset consists of 4,705 images. Each image in this dataset has been standardized and labeled anew for consistency.

To describe the labeling strategy, the primary label we assign to each image in our dataset

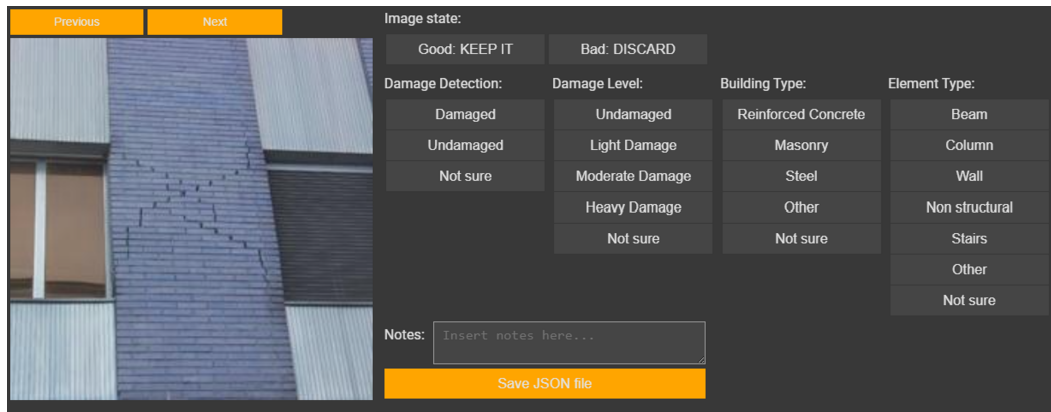


Figure 4.7. The labeling dataset GUI, customized in Google Colaboratory. Image adapted from [94]

is the damage severity level of the building or structural element depicted. We adopted a four-class damage level classification scheme as stated in Chapter 2: Undamaged (U), Slight damage (S), Moderate damage (M), and Heavy damage (H). This scheme is aligned with the typical post-earthquake damage classifications used in Italy and Europe. In fact, it parallels the damage grading of the AeDES form (the official Italian post-earthquake building safety assessment form) wherein damage levels D1, D2–D3, D4–D5 correspond to slight, moderate, and heavy/collapse damage respectively. By aligning our labels with AeDES criteria, we ensure that our dataset's categories have a clear engineering interpretation and adhere to established Italian standards for damage assessment.

During dataset construction, each image was carefully examined by at least two experienced structural engineers from our research team. They assigned one of the four damage level labels, referencing any available context (such as the known earthquake intensity for DFM images, or descriptions from reports) to validate the label. Where an image showed multiple buildings or components, the focus was placed on the primary structure in view. This rigorous re-labeling process was crucial to eliminate any noise from the original labels in Φ -Net or other sources. For example, some Φ -Net images labeled as "moderate damage" were adjusted to "slight" if, under AeDES criteria, the damage seemed minor (e.g., hairline cracks only). Conversely, a few images labeled as just "moderate" in Φ -net were specified as "heavy" in our dataset if they actually showed collapse mechanisms. The outcome is that all 4,705 images now have a consistent damage level label in U, S, M, H, which will be used as the target output class for our classification models. In addition to the damage level, we also annotated some secondary information for quality control and potential future use. This includes the structural component or building type in the image (e.g., "masonry building" vs "RC frame", or "column" vs "wall" focus). While our immediate task in next chapters focuses on damage level classification, having these extra tags allows us to ensure that the dataset is balanced and to analyze results by category (for instance, checking if the model performs differently on masonry vs concrete buildings).

Table 4.5 summarizes the distribution of images by structural element and building typology within the dataset.

We also performed standardization of the images: all were resized to 224×224 pixels (while maintaining aspect ratio by cropping/padding as needed) to match the input requirements of common CNN architectures. In many cases, we also cropped or zoomed into the

Table 4.5. Dataset Composition: Structural Element Counts and Building Type Counts [94]

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		

region of interest (the damaged part) of the photo to eliminate irrelevant background and to ensure the damage is clearly visible. This was done carefully to not bias the model with pre-selection of damage, but to improve image clarity. For example, a wide street view photo of a building was cropped to just the building facade so that the cracks are discernible to both the model and a human inspector. By cleaning the images in this manner, we aimed to improve the signal-to-noise ratio for training.

4.5.3 Key Advantages of the Newly Compiled Dataset and Concluding Insights

Our newly curated dataset represents a significant improvement as a benchmark for advancing the performance of damage classification.

The final dataset we assembled is comparatively not smaller than task 7 Φ -Net (4,705 vs 4,600 images), and it is highly targeted to the problem of post-earthquake building damage classification. It combines the breadth of an international dataset (Φ -Net content from around the world) with the depth of regional expertise (INGV and ReLUIS content from Italy, plus the Turkey event data for diversity). Each image's label is grounded in engineering judgement, and the class definitions tie back to real damage assessment scales. We anticipate that a model trained on this curated dataset will have improved performance in classifying damage levels on new images, precisely because the training data is more consistent and relevant. The inclusion of Italian data and labeling standards means that the model is learning in the context it will be applied (especially if the deployment is for Italian civil protection or similar scenarios). In conclusion, this chapter surveyed the landscape of post-earthquake damage image datasets and introduced the custom dataset developed for this research. The Φ -Net provided a groundbreaking large pool of structural images, but due to its generality, it required adaptation for specific use. The INGV DFM contributed authoritative, intensity-linked photos from Italian earthquakes, and the IDEA dataset represents the cutting edge of annotated damage imagery useful for future advancements.

Building on these, our new dataset of 4,705 images was compiled to overcome Φ -Net's limitations and to ensure robust training data for our purposes. This new dataset will be the new baseline for the experiments in the next chapter, where we train and evaluate various convolutional neural network models for classifying building damage. By using this carefully curated dataset, we aim to achieve higher accuracy and better generalization in automated damage assessment, as will be demonstrated in the following chapters.

Summarizing, this dataset overcomes several well-known limitations found in currently available repositories:

- It includes a diverse range of structural typologies and damage classes, enhancing the model’s ability to generalize across different building types and damage scenarios;
- All images are annotated, not merely classified, enabling not only damage level recognition, but also building type and building element;
- The imagery is sourced from real post-earthquake reconnaissance, rather than controlled laboratory experiments or synthetic simulations, thus reflecting authentic damage conditions and real-world variability.
- The images have been cropped and zoomed to simulate the perspective of field engineers in situ, focusing on the key damaged element as they would when filling out assessment and operability forms.

In addition, our dataset integrates expert relabeling aligned with engineering standards such as the AeDES categorization, ensuring that severity labels (e.g. slight, moderate, heavy) hold direct operational meaning. Rigorous quality control, through multi-expert verification and metadata curation, further enhances label consistency and reliability. As a result, models trained on this dataset demonstrate improved accuracy and robustness (refer to Chapter 6) compared to those trained on less representative or incompletely annotated datasets. This makes our dataset not only suitable for this task for its content and context, but also for its scientific validity as a foundational resource for deep learning–based damage assessment.

Chapter 5

Experimental Setup and Benchmarking of Deep Learning Models for Post-Earthquake Damage Classification

The current chapter depicts the experimental framework that was designed to benchmark and compare the performance of various deep learning architectures in identifying the structural damage from RGB images of buildings that were impacted by earthquakes. The work aims at exploring the performance of various model families such as the traditional Convolutional Neural Networks (CNN) to lightweight architectures and Vision Transformers (ViTs) under differing training strategies and dataset arrangements.

The chapter opens with a benchmarking study on the Φ -Net dataset (§5.1), a large-scale reference dataset for classes of seismic damage. In this part, the description of the CNN strategy and chosen architectures takes place (§5.1.1), followed by information on the preprocessing pipeline and the data augmentation techniques implemented to mitigate the issue of class imbalance (§5.1.2). The section ends up with the comparative results (§5.1.3) discourse, where the discussion of different CNN families' pros and cons takes place.

The work presented has been subsequently extended to the curated dataset proposal (§5.2) that combines multiple data sources and is mainly concentrated on the Italian post-earthquake imagery. The first part of this work consists of an overview of the CNN training strategy and the selected models (§5.2.1), followed by a more detailed description of dataset construction and augmentation procedures and the reporting of the experimental outcomes that lead to the selection of the most effective training configuration (§5.2.2) and of the different architectures final results (§5.2.3) together with their brief discussion.

The third section of this report has been focused on lightweight CNNs (§5.3), which are basically conceptually designed to serve as a quick reference in the field of post-disaster assessment by rapidly showing the possibilities of their application in mobile devices and UAV- and smartphone-based scenarios for structural health monitoring, leading us further to the architectural designs of EfficientNet (§5.3.1) as well as to the wavelet-based data fusion strategies (§5.3.2). The description of a gradual unfreezing training method was next (§5.3.3), followed by a comparative discussion of results and the identification of the most suitable lightweight model (§5.3.4).

The next section considers the ViT (Vision Transformer) models (§5.4), a different approach to convolutions and introduce attention mechanisms instead. The differences between such modifications as B/16–B/32 and L/16–L/32 are outlined (§5.4.1–§5.4.2) with the emphasis on training strategy and performance analysis given in a separate subsection (§5.4.3). The last discussion (§5.4.4) equips us with the aid of the previous works to effortlessly identify the pros and cons of the ViT models in comparison to CNN-based benchmarks and helps to deliver a full understanding of their potential in post-earthquake building damage classification.

In the fifth chapter (§5.5) there is a discussion on some alternative ways of processing not only CNN-based primary methods (e.g., VGG16) or ViTs but also whether those alternatives bring advantages that are complementary or higher in terms of quality. They comprise such items as those related to the VGG16 neural net method for hierarchical classification (§5.5.1), decomposition of open source convolutional autoencoder (CAE)-based feature extraction with the use of traditional classifiers (§5.5.2), and the first attempts at building quaternion-based architectures (§5.5.3). The experimental results reveal that the newly proposed methods are generally on par with the baselines of CNNs, but they have to make concessions in terms of model complexity, computational cost, and interpretability. These approaches are not fully optimized or tested here. The development of exhaustive hyperparameter search, more comprehensive datasets for evaluations, and the extent of comparative experiments among the methods explored are planned for the future.

5.1 Benchmarking of CNN architectures for post-earthquake damage assessment using the Φ -Net dataset

The primary target of this chapter is to compare how various CNN designs influence the process of aftershock assessment of building damage, as well as checking the utility of the Φ -Net dataset as a benchmarking resource. Specifically, we quantify their capability to identify the correct damage level from RGB images of the damaged part of the structure. Our goals are:

- Comprehensive benchmarking: performing a large-scale benchmark of different CNN architectures for classification with a consistent methodology and the same dataset and test set as [25], using their training set to generate our train and validation sets.
- Optimal CNN selection: identifying the most suitable CNNs based on a variety of performance metrics, considering recall, precision, F1-score, and weighted F1-score for the four damage classes.
- Understanding limitations: gaining insights into the capabilities and limitations of both the neural network and the dataset.
- Harnessing open accessibility: exploiting the dataset's accessibility to benefit a wide range of users, to promote reproducibility of the results and justify the choice of the model best suited for the purpose.

5.1.1 Dataset

The experimental setup was based on the Φ -Net dataset [25], a large-scale collection of RGB images of structural elements subjected to varying degrees of seismic damage, already seen in the previous chapter. The dataset contains four categories: *Undamaged* (U), *Slight* (S), *Moderate* (M), and *Heavy* (H). Each image was resized to 224×224 pixels and normalized before the training phase. The original training set was split into training and validation subsets (75/25 ratio) to ensure reproducibility.

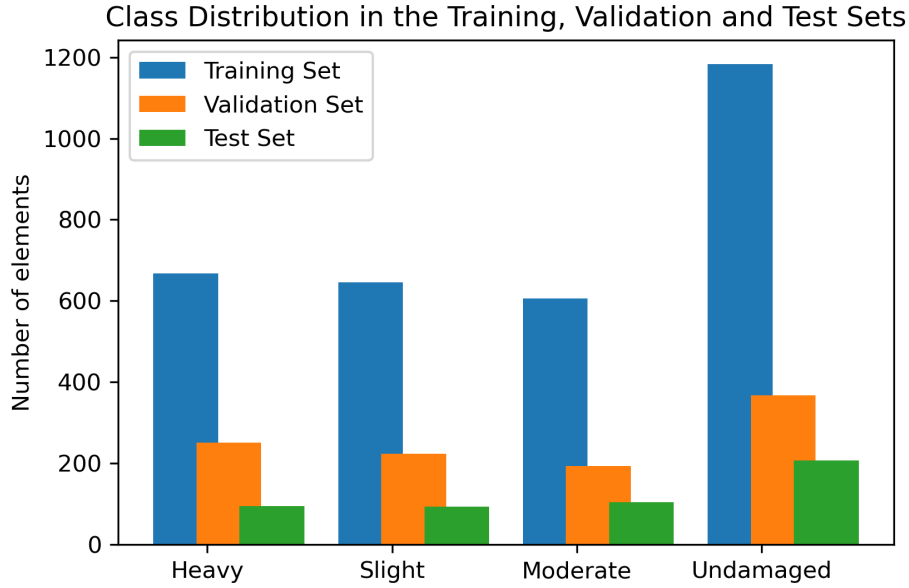


Figure 5.1. Distribution of images in the Φ -Net dataset.

Table 5.1. Distribution of images in the Φ -Net dataset.

Set	Damage				Total
	None	Slight	Moderate	Heavy	
Training	1183	606	646	668	3103
Validation	368	193	223	251	1035
Test	207	93	104	94	498
Total	1758	895	973	1013	4636

To improve generalization and mitigate overfitting, standard data augmentation was applied, including random flipping, rotations, and contrast variations during the training phase.

5.1.2 Training Strategy and CNN Models

To investigate the applicability of deep learning models for post-earthquake damage classification, we conducted a comparative analysis of ten well-known Convolutional Neural Network (CNN) architectures. These included classical models such as AlexNet and VGG16, as well as more advanced designs like ResNet, DenseNet, Xception, InceptionV3, and MobileNet.

Each network was implemented using the Keras library, which has been the common framework for all the models, ensuring consistency in the training phase. The purpose of this benchmarking exercise was to evaluate how different architectural choices, depth, connectivity patterns, and parameter efficiency, affect classification performance when dealing with structural damage images.

To ensure effective model training, the dataset underwent a series of pre-processing operations. Images were first normalized by rescaling their pixel values within the $[0,1]$ range, which enhances consistency across the dataset and supports stable convergence during training. The dataset was then randomly partitioned into training and validation subsets using the Scikit-Learn library, with a fixed seed (42) to guarantee reproducibility.

Subsequently, the data has been divided at random into subsets for training and validation using the Scikit-Learn library. A seed (42) was fixed to ensure replicability.

More specifically, 75% of the Φ -Net training set was kept for training and 25% for validation. This distribution allows one to depict a diverse sample for model learning while leaving the performance monitoring and overfitting control an unseen data.

The training process was conducted over 50 epochs, with an early stopping criterion after 10 consecutive epochs with no validation accuracy improvements.

Three different learning rate configurations were explored: two fixed values (10^{-4} and 10^{-5}) and an adaptive reduce-on-plateau (RoP) scheduler. The Adam algorithm was chosen as optimization algorithm and categorical cross-entropy as loss function, with softmax activation at the final dense layer.

All other hyperparameter settings were kept at their Keras default values to make sure that the experimental results are comparable.

Three different experimental scenarios were considered. The first two experiments included CNN models trained from the beginning with fixed learning rates of 10^{-4} and 10^{-5} respectively. Transfer learning was introduced in the third setup, which also combined the usage of ImageNet pre-trained weights [53] and RoP strategy. Consequently, the neural networks could leverage the reusable features that were already extracted from generic large-scale datasets, which led to faster convergence and better generalization for seismic damage classification. In this case, image normalization was done following ImageNet preprocessing standards to allow compatibility with pre-trained layers. The organization of the transfer learning process was in two stages. The first stage was the feature extraction step where the early convolutional layers were frozen, and only the classifier head was trained with a small learning rate of 10^{-4} . The second step was the fine-tuning phase, in which all layers were unfrozen and trained with a lower learning rate of 10^{-5} under the RoP strategy, allowing for controlled gradual adaptation of the pre-trained weights to the new task and at the same time avoiding severe weight changes. Data augmentation was evaluated to grow the sample diversity by employing the ImageDataGenerator module from TensorFlow. The main reason that only AlexNet and GoogLeNet received augmentation strategies was that pre-trained ImageNet weights were not natively available in Keras for

those architectures. Therefore, the synthetic transformations' ability to make up for the lack of transfer learning was tested. The different training experiments were implemented over both cloud and local hardware. Google Colab allowed the use of a Tesla T4 GPU (16 GB), while local simulations made use of a Tesla P40 GPU (24 GB). Employing both settings allowed a wide benchmarking campaign to be carried out, which made possible to measure training efficiency, computational time, as well as model scalability. The combination of these computational tools gave the team the opportunity to compare thoroughly between different architectures and training strategies, thus ensuring the reliability of the analysis for building damage after the earthquake. The researchers also experimented with transfer learning by setting the weights of the networks to pretrained on ImageNet and then retraining them on Φ -Net. This method was mainly effective to make up for the relatively small size of the dataset as compared to the standard computer vision benchmarks [26].

5.1.3 Discussion and Results

The comparative results of the ten evaluated CNN architectures are displayed in Table 5.2. Multiple training configurations for each model were tested and the results are reported in terms of accuracy, precision, recall, F1-score both for each class and as weighted averages, along with the average training time per epoch. Among all the networks that were under consideration, DenseNet201, VGG16, and ResNet101 stood out as the best performers with test accuracies of 73.09%, 72.49%, and 72.29%, respectively. Their weighted F1-scores (72.20%, 72.50%, and 71.96%) are in agreement with the strength of these architectures in distinguishing the structural damages across the four severity levels. The confusion matrices of these three best-performing models are given in Fig.5.2, while Table 5.3 shows their performance metrics. Most of the models, achieved a performance better than the 70% accuracy threshold as can be seen in Table 5.3 and this clearly demonstrates that CNN-based approaches are viable tools for damage classification of buildings. However, models without transfer learning that were trained from scratch had poor generalization and consequently they did not converge effectively. Complex architectures appeared less capable of fully exploiting their deep representation power given the size of the dataset, while simpler networks such as AlexNet, GoogLeNet, and especially VGG16, turned out to be more efficient in generalization and they could therefore take advantage of their smaller number of parameters and the lower overfitting risk. Despite these issues, the application of transfer learning led to a great enhancement of performance for every single architecture and made them all reach approximately the same accuracy level that is close to the state-of-the-art i.e. about 70%. During this exercise of benchmarking, VGG16 resulted as the most efficient and reliable model. It achieved the highest accuracy of 72.49% and the highest weighted F1-score of 72.50%.

On a class by class level, it showed good precision and recall for the "Undamaged" category (83.72% and 86.96%), worse for the "Slight" (61.11% and 59.14%) and "Moderate" (54.05% and 57.69%) damage classes, and the moderately good results for the "Heavy" class (80.49% and 70.21%). Besides the competitive accuracy, VGG16 was less demanding in terms of computational resources compared to the deeper models, which was visibly demonstrated by the reduced memory footprint and shorter inference times, attracting real-time application scenarios. Nevertheless, the problem areas persisted, with the main issue being the differentiating of slight and moderate damage groups, as misclassifications in these categories still appeared frequently. This limitation makes evident the need to not

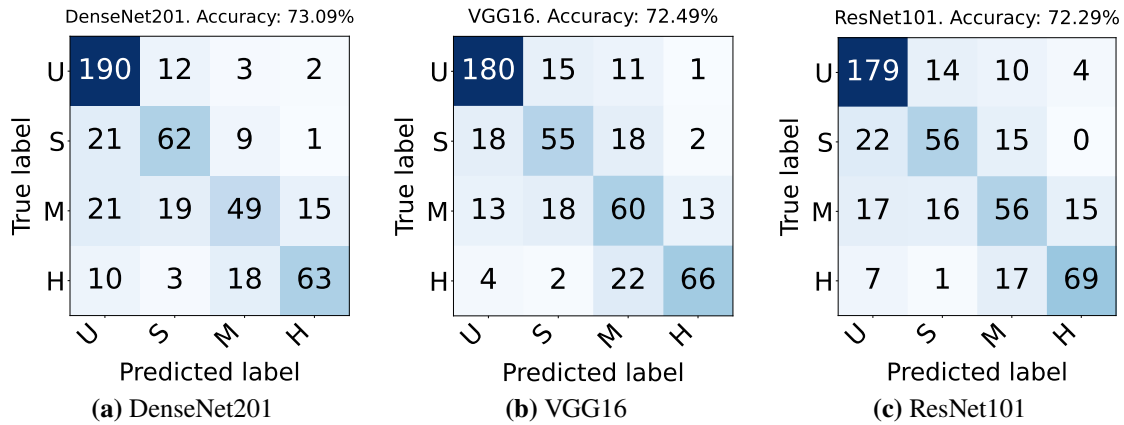


Figure 5.2. Confusion matrices for the best three models with highest weighted F1-score, i.e. DenseNet201, VGG16, and ResNet101. The four classes are: Undamaged (U), Slight (S), Moderate (M), and Heavy (H).

only increase the dataset volume but also improve the consistency of labeling and apply advanced augmentation methods. Although considerably limited performance was achieved with standard TensorFlow-based data augmentation, the scope of the improvements was small.

Overfitting was another recurring issue, partially alleviated by transfer learning, which confirmed itself as an indispensable strategy for reaching acceptable performance on limited datasets.

Moreover, from the computational point of view, the Tesla P40 GPU can be regarded as an efficient piece of hardware since the experiments fully supported it by showing that the training time could be reduced by 50% from it to the Tesla T4. For example, VGG16's average epoch duration was shortened from approximately 49 seconds on the T4 to 26 seconds on the P40, which is a manifestation of the importance of the hardware acceleration role in carrying out large-scale benchmarking studies.

Lastly, it is crucial to mention that the Φ -Net dataset incorporates various challenging factors that are typical in the deployment of aid in disaster zones. For instance, the images may contain occlusions, different lighting conditions and even a mixture of indoor and outdoor spaces. Although this complicates the classification, it is also more reflective of the actual working conditions. The next step will be to explore the possibility of bringing more preprocessing methods, for example, noise injection or adaptive contrast enhancement, into the dataset and extending it to include different types of buildings and damage patterns, with a particular emphasis on reinforced concrete and masonry structures as the current database trend.

The benchmarking exercise gave informations that showed that deeper architectures with skip connections or dense connectivity usually scored higher in terms of accuracy and robustness. Out of the ten models that were compared, VGG16, DenseNet201, and ResNet101 turned out to be the most efficient ones, obtaining weighted F1-scores of over 72% on the test dataset. As for simpler models like AlexNet and GoogLeNet, they demonstrated good generalization abilities, but their peak accuracy was lower, which is consistent with their smaller representational capacity.

Table 5.2. Results of the compared CNN architectures.

Model	LR	TL/DA	Accuracy		Precision			Recall			F1-score			Tr. Time					
			Val.	Test	Und.	Sl.	Mod.	Heav.	Wtd.	Und.	Sl.	Mod.	Heav.	Wtd.	T4	P40			
AlexNet	10 ⁻⁴	—	49.08	52.21	66.21	56.45	24.49	40.48	50.82	70.05	37.63	11.54	72.34	52.21	68.08	45.16	15.69	51.91	49.80
	10 ⁻⁵	—	51.01	53.21	62.93	45.24	37.50	48.31	51.56	70.53	40.86	23.08	60.64	53.21	66.51	42.94	28.57	53.77	51.78
	RoP	DA	52.17	56.63	67.15	48.28	42.86	54.17	56.10	67.15	45.16	34.62	69.15	56.63	67.15	46.67	38.30	60.75	56.09
GoogLeNet	10 ⁻⁴	—	54.30	54.62	72.38	54.41	44.12	40.88	57.18	63.29	39.78	28.85	78.72	54.62	67.53	45.96	34.88	53.82	54.09
	10 ⁻⁵	—	53.82	55.82	64.62	54.72	36.84	45.58	53.37	81.16	31.18	13.46	71.28	55.82	71.95	39.73	19.72	55.60	51.94
	RoP	DA	51.98	56.22	64.00	46.25	44.64	48.28	53.67	85.02	39.78	24.04	44.68	56.22	73.03	42.77	31.25	46.41	53.63
XceptionNet	10 ⁻⁴	—	67.34	36.35	47.73	28.26	23.73	32.35	36.18	50.72	27.96	26.92	23.40	36.35	49.18	27.96	26.92	23.40	36.09
	10 ⁻⁵	—	64.93	37.15	49.32	26.55	30.38	26.44	36.79	52.17	32.26	23.08	24.47	37.15	50.70	29.13	26.23	25.41	36.79
	RoP	TL	69.07	70.92	80.66	58.65	51.19	70.41	68.46	82.61	65.59	41.35	73.40	69.08	81.62	61.93	45.74	71.88	68.61
InceptionV3	10 ⁻⁴	—	66.76	31.53	47.83	26.45	22.52	21.90	33.66	32.70	34.41	24.04	24.47	31.53	41.84	29.91	23.26	23.12	32.20
	10 ⁻⁵	—	68.12	34.74	46.33	30.77	20.78	24.24	33.92	48.79	34.41	15.38	25.53	34.74	47.53	32.49	17.68	24.87	34.21
	RoP	TL	72.95	70.08	84.69	57.00	49.58	80.72	71.44	80.19	61.29	56.73	71.28	70.08	82.38	59.07	52.91	75.71	70.61
VGG16	10 ⁻⁴	—	54.49	58.63	70.04	49.44	38.61	61.73	58.06	76.81	47.31	37.50	53.19	58.63	73.27	48.35	38.05	57.14	58.22
	10 ⁻⁵	—	54.88	56.43	66.51	53.42	48.61	46.81	56.61	68.12	41.94	33.65	70.21	56.43	67.30	46.99	39.77	56.17	55.66
	RoP	TL	70.24	72.49	83.72	61.11	54.05	80.49	72.69	86.96	59.14	57.69	70.21	72.49	85.31	60.11	55.81	75.00	72.50
VGG19	10 ⁻⁴	—	54.39	56.43	68.10	45.16	39.06	50.00	54.34	76.33	30.11	24.04	74.47	56.43	71.98	36.13	29.76	59.83	54.18
	10 ⁻⁵	—	52.37	52.81	67.49	43.96	39.29	43.24	52.63	66.18	43.01	21.15	68.09	52.81	66.83	43.48	27.50	52.89	51.62
	RoP	TL	67.34	68.47	80.37	60.47	52.87	63.96	67.82	83.09	55.91	44.23	75.53	68.47	81.71	58.31	48.17	69.27	67.95
ResNet50	10 ⁻⁴	—	70.72	34.94	48.47	29.67	18.92	31.00	35.49	45.89	29.03	20.19	32.98	34.94	47.15	29.35	19.53	31.95	35.19
	10 ⁻⁵	—	68.40	35.14	49.02	26.47	23.66	26.26	35.22	48.31	29.03	21.15	27.66	35.14	48.66	27.69	22.34	26.94	35.15
	RoP	TL	68.79	71.69	78.32	62.63	63.51	71.72	71.05	85.51	66.67	45.19	75.53	71.69	81.76	64.58	52.81	73.58	70.96
ResNet101	10 ⁻⁴	—	69.76	38.35	50.81	39.76	26.87	29.17	36.65	45.41	35.48	34.62	29.79	36.32	47.96	37.50	30.25	29.47	36.30
	10 ⁻⁵	—	66.09	36.75	48.50	29.70	26.60	30.10	36.94	46.86	32.26	24.04	32.98	36.75	47.67	30.93	25.25	31.47	36.80
	RoP	TL	69.08	72.29	79.56	64.37	57.14	78.41	71.82	86.47	60.22	53.85	73.40	72.29	82.87	62.22	55.45	75.82	71.96
MobileNetV2	10 ⁻⁴	—	47.24	33.33	43.62	24.77	32.56	18.45	33.04	51.21	29.03	13.46	20.21	33.33	47.11	26.73	19.05	19.29	32.19
	10 ⁻⁵	—	45.60	33.53	42.43	21.31	33.33	16.52	31.70	62.32	13.98	5.77	20.21	33.53	50.49	16.88	9.84	18.18	29.63
	RoP	TL	67.05	70.68	81.17	58.59	55.17	73.03	69.99	87.44	62.37	46.15	69.15	70.68	84.19	60.42	50.26	71.04	70.18
DenseNet201	10 ⁻⁴	—	55.46	32.73	42.73	23.73	26.98	23.33	32.23	46.86	30.11	16.35	22.34	32.73	44.70	26.54	20.36	22.83	32.10
	10 ⁻⁵	—	56.63	31.93	42.11	21.78	24.49	24.17	31.25	46.38	23.66	11.54	30.85	31.93	44.14	22.68	15.69	27.10	30.97
	RoP	TL	70.05	73.09	78.51	64.58	62.03	77.78	72.33	91.79	66.67	47.12	67.02	73.09	84.63	65.61	53.55	72.00	72.20

Table 5.3. Best models performance. The last column is the averaged per-single image inference time, evaluated on the P40 GPU.

	F1-score	Accuracy	Parameters	Model Size	Inf. Time
VGG-16	72.50	72.49	14.8 M	170 MB	25.58 ms
DenseNet201	72.20	73.09	18.8 M	216 MB	35.17 ms
ResNet101	71.93	72.29	43.2 M	488 MB	36.82 ms

The research also revealed that transfer learning was a major factor in the improvement of the results in all the architectures, particularly when it came to dealing with the problem of scarce labeled data. Most models were not able to achieve good convergence without transfer learning, while fine-tuned networks were always above the 70% accuracy threshold. However, it was still difficult to distinguish *Slight* from *Moderate* damage levels, thus implying that the quality of the dataset and the range of augmentation being used need to be further enhanced.

In fact, this benchmarking exercise validated that CNNs are the best choice for visual-based seismic damage assessment. Specifically, the relationship between model complexity and efficiency that was demonstrated in VGG16 and DenseNet201 is a witness to the fact that models of moderate depth can give out both accuracy and computational feasibility. These findings serve as the baseline for the future research in automatic post-earthquake inspection frameworks [93].

5.2 Benchmarking of CNN architectures for post-earthquake damage assessment using the proposed dataset

With the dataset that was introduced and explained in Chapter 4, this section outlines the systematic benchmarking of CNN architectures dealing with post-earthquake building damage classification. In brief, the intent of this section is to utilize the proposed new dataset to:

- under the same experimental framework, test the performance of various CNN architectures;
- measure a number of classification metrics, such as recall, precision, and F1-scores;
- evaluate the contribution of data augmentation strategies and select the optimal augmented dataset;
- select the architectures that most appropriately combine performance and efficiency;
- analyze the influence of dataset cleaning and the removal of noisy samples;
- emphasize the intrinsic limits of both the models and the dataset, thus giving a very close view of the performance achievable.

Such a thorough evaluation fixes a very strong baseline for following improvements, which is essentially the way to lighter models and more advanced strategies for mobile and field applications.

5.2.1 Dataset and Data Augmentation

The experiments in this work are based on the dataset proposed and introduced in Chapter 4, which was particularly designed to overcome the limitations of existing collections. This dataset has been created by combining diverse sources, with the most significant contribution being from the PEER Hub ImageNet (Φ -Net) repository and the INGV Macroseismic Photographic Database (DFM), along with some post-earthquake photographic material from national and international surveys. The new collection scales to closer depict the kinds of structures and their damage patterns, as it can be seen from the progression of damage types from fully intact to collapse, moreover, giving a lot of attention to those reinforced concrete and masonry buildings which are the most common in Italy and the Mediterranean area. Owing to the significant imbalance in the four AeDES-derived damage categories, the dataset was subject to various augmentation and resampling procedures to not only improve class balance but also allow the model to generalize better. To begin with, oversampling was done on the classes with the least number of instances which increased the training set from 3,505 images to 4,828 samples. Eventually, geometric transformations were introduced to enable the gradual augmentation of the data. Firstly, there was an attempt to double the training size through horizontal flips which was later extended with the addition of random rotations of up to 10° , thus increasing the training size to 14,484 samples. Subsequently, the injection of Gaussian noise was introduced to cover various realistic acquisition conditions resulting in a more diverse dataset of 19,312 images. Ultimately, perspective transformations were identified as one of the approaches to introducing more variability albeit a minor one in terms of the improvement of performance.

All augmentations were implemented in Python, leveraging the NumPy and Albumentations libraries to ensure efficiency and reproducibility. To ensure an equitable and stable environment for benchmarking, the test set composed of 1,200 balanced samples across the four classes was kept intact. This augmentation pipeline was instrumental as the CNN architectures, on preliminary verification, took longer and were less stable when trained on the unaltered dataset. Consequently, the combination of different data sources with systematic balancing and augmentation turned out to be an effective strategy for the comparative evaluation of CNN architectures for post-earthquake damage classification.

Table 5.4. Different dataset compositions after augmentation.

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

5.2.2 Taining strategy And CNN Models

For the benchmarking activity, twelve state-of-the-art CNN architectures were implemented and tested. Among them, the VGG16 network [101] consistently emerged as the most

effective. VGG16 consists of 16 layers, 13 convolutional, and 3 fully connected and is very well-known for its compromise between the ease of understanding and the accuracy of results. Although it is not as new as the other deep networks, its relatively shallow depth and standardized design often make it particularly appropriate for medium-scale datasets, thus making it possible to achieve stable convergence and reliable classification performance.

In order to increase the model accuracy and robustness, we have made use of a transfer learning strategy that involved initializing the VGG16 architecture with pre-trained weights that were derived from the large-scale ImageNet dataset [100].

The fine-tuning process was executed in a supervised training setup with an integrated validation. In particular, 10% of the training data was held out for validation purposes. The model was trained on mini-batches of 16 images using the categorical cross-entropy loss function that was optimized with the Adam optimizer. The ReduceLROnPlateau scheduler was used as a dynamic learning rate policy, which automatically reduced the learning rate by half when the validation loss did not show any improvement for five consecutive epochs. The starting learning rate was 10^{-5} , and all other hyperparameters were kept to their Keras default values so as to be able to compare with other architectures.

Python was used for all the simulations. The computational setup was a standard workstation with an Intel Core i9-14900KF processor, 32 GB of RAM, and an NVIDIA GeForce RTX 3050 GPU with 8 GB of dedicated memory. This configuration was enough to handle the training and validation of the proposed models, providing a good balance between computational cost and experimental reproducibility.

5.2.3 Discussion and Results - Selection of the optimal augmented dataset for training

The classification metrics adopted to evaluate the effectiveness of the suggested models encompassed overall accuracy as well as per-class precision, recall, and F1-score, along with their weighted averages. The metrics were computed on the independent test set constituted by 1,200 images (see Table 5.1).

The reliability of using the VGG16 model is tied up with the necessity of considering both dataset balancing and the employment of various augmentation strategies, a fact that was revealed indisputably in the experiments. Table 5.5 gathers the results achieved across the training sets described in Table 5.4.

The model, when it was trained on the original imbalanced dataset, had a moderate performance with an acceptable level of results (86.25% overall accuracy) but still not optimal. The employment of the oversampling method (Training set 2) increased the number of samples in the minority classes, thereby bringing the dataset to a more balanced state and immediately producing these benefits not only in accuracy but in model generalization as well. More improvements were realized by gradually adding more and more geometrical transformation to the augmented part of the training data. Both horizontal flipping (Training set 3) and the introduction of small rotational perturbations (Training set 4) could have led to the model becoming more stable against the variations that could have occurred in the environment, thus reaching this better performance. The highest level of performance was attained by training set 5 that had combined flipping, 10° rotations, and Gaussian noise injection as augmentation techniques. Each one of these techniques contributed to diversity in the dataset without eliminating the key features that were even allowing the deep learning model to achieve an accuracy of 89.33%, i.e. almost twice the baseline. On the contrary, the

introduction of perspective distortions (Training set 6) had not been very effective to extend their benefits and even decreased the stability in several instances.

Table 5.5. Comparisons of VGG16 test set classification metrics (in percentage) using different data augmentation approaches.

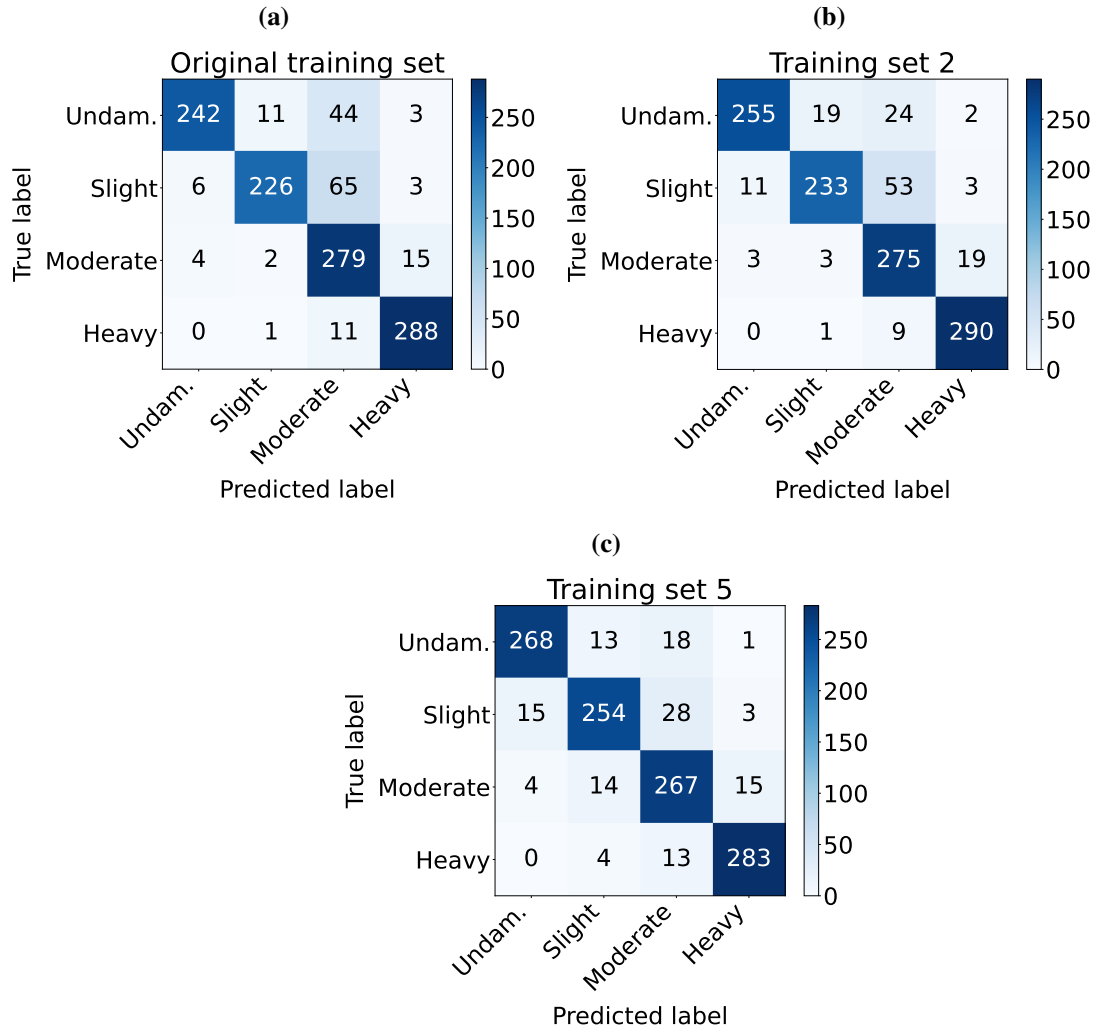
Training set	Accuracy	Precision	Recall	F1-score
Original training set	86.25	88.33	86.25	86.45
Training set 2	87.75	88.59	87.75	87.78
Training set 3	87.92	89.05	87.92	87.97
Training set 4	88.75	89.66	88.75	88.78
Training set 5	89.33	89.53	89.33	89.37
Training set 6	87.58	88.99	87.58	87.77

Table 5.6. 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.

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

The confusion matrices that correspond (Figure 5.3) show the changes in classification power across the datasets. A significant decrease in false positives is seen particularly in the cases of the incorrect labeling of "Undamaged" and "Slight" as "Moderate". For instance, the number of undamaged images that were falsely labeled as moderate were reduced from 44 (original set) to 18 (Training set 5). Moreover, the number of the slightly damaged images that were misclassified has decreased from 65 to 28. These advancements are crucial for the real world, as they significantly increase the trust of the automatic post-earthquake evaluations. Anyways, there are still some problems with consistent differentiation between "Slight" and "Moderate" indicating the need for further research in the future. Besides VGG16, the differences among the eleven other CNN architectures were also evaluated and their findings are summarized in Table 5.9. The models include AlexNet, GoogLeNet, DenseNet201, MobileNet-V2, EfficientNet-B0, two ResNet variants, XceptionNet, InceptionV3, and a baseline CNN from the literature [9] are the set of networks. These networks are the most common in structural health monitoring (SHM) and damage detection tasks [4, 93, 100,

Figure 5.3. 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).



40]. Accordingly, among all the contenders, VGG16 was always among those with the highest scores, yielding not only the highest accuracy but also the most balanced F1-scores, and computational efficiency in both training and inference. In essence, the findings bring out the potential of CNN-based deep learning methods in handling the process of building damage assessment that comes as a result of an earthquake. Transfer learning, methodical augmentation, and dataset equalization combined to form VGG16 model that was able to perform excellently with accuracies and F1-score higher than 89%. The presented results serve as an indication of the ability of deep models to offer fast, impartial, and reproducible assessments, thus, they become a powerful tool in support of field inspections that still remain the traditional way. Though there are many promising achievements, the mission to unravel the ambiguities between the intermediate damage levels still continues. One of the suggested directions for further research is the focus on improving the separability of the classes, which could be achieved by a multi-modal input like heat or laser sensor data or

by taking aid of an advanced architecture that can easily detect the subtle changes in the structure.

5.3 Light wheight CNN and DWT Data-Fusion

In the field of post-earthquake emergency management, the capacity to implement efficient deep learning models directly on mobile devices or on light hardware is extremely valuable. What a field engineer needs is not a heavy tool but a fast and reliable support tool on the spot, which can rapidly assess the damage and give the green light for the upcoming decision. This is the reason that this part of the thesis is the design and benchmarking of a lightweight CNN architecture which is the consideration of potential mobile deployment.

The foremost goals of this section may be outlined as follows:

- Mobile-oriented models: the study compact CNN architectures like EfficientNet-B0 that might realistically be embedded into mobile devices and remotely field engineers during on-site inspections, thus making them mobile devices assistant.
- Improved feature representation: to experiment on the application of Discrete Wavelet Transform (DWT) to be an input feature enrich method as well as for achieving higher classification accuracy in lightweight networks.
- Stability and reproducibility: training and testing to be conducted on the identical data set partitions as those in the previous section, thereby baselines CNNs direct comparability is guaranteed.

This section, through such an evaluation, aims at enlightening the engineering community about the applicability of the presented lightweight and wavelet-enhanced architectures as a possible bridge to finally witnessing research prototypes real-field mobile applications-structural damage assessment.

5.3.1 Efficientnet-B0 architecture

The EfficientNet-B0 architecture was selected as part of this study, due to its favorable compromise between accuracy and computational efficiency. As the comparative results of the last section have indicated, it outperforms MobileNetV2 over the proposed dataset in terms of accuracy and F1 score, while also sustaining a small enough design for applications with limited resources.

Traditional CNNs Ensemble decide to alter one dimension (depth, width, or resolution) at a time. Whereas EfficientNet compound scaling strategy [107] allows scaling small model to much bigger one with balanced dimensions and in a principled way. The results are higher accuracy with fewer parameters and less computational cost than traditional models.

The core of EfficientNet-B0 is an initial 3×3 convolutional layer, followed by a series of Mobile Inverted Bottleneck Convolution (MBConv) blocks, and then a final classification stage with a fully connected layer and softmax activation (Figure 5.4). The MBConv blocks, which come from MobileNet, are now in two versions: MBConv1 and MBConv6. These blocks use 1×1 convolutions, depthwise convolutions, and Squeeze-and-Excitation (SE) modules, the latter which adaptively reweights channel-wise feature responses to facilitate

that most informative features become more evident. In addition, the network uses the Swish activation function that eases the gradient flow as compared to the standard ReLU.

There are several reasons for the use of EfficientNet-B0:

- Lightweight architecture - the number of parameters is much lower than in classical architectures such as VGG16 or ResNet201, which makes it very suitable for hardware with limited computing power.
- A good trade-off between accuracy and efficiency - it can still perform well on a variety of tasks even though it is quite small, and for that reason, it is very practical for a mobile device whose battery life is limited.
- Compatibility of input sizes - the B0 version by default takes images of 224×224 pixels as input, and this is exactly the resolution of our dataset, so the model is naturally fitting without any resizing or other intermediate steps.

On the other hand, some weaknesses without should be mentioned. Even if EfficientNet-B0 is very efficient, its smaller depth and width in comparison with other larger EfficientNet variants such as B4 and B7 might result in limited capacity to recognize very complex patterns. We accept that compromise in our context, as we are more interested in testing the solutions that are scalable and mobile-ready, yet at the same time, we can maintain the reproducibility of the dataset partitions.

The EfficientNet-B0 design was the first choice in this research work largely due to its excellent trade-off with the accuracy and the computational side of things. As it was very clear from the comparison results of the previous part, it not only outclassed MobileNetV2 in terms of accuracy and F1-score on the new dataset, but it also maintained its small size which is ideal for applications with limited resources. While standard CNNs change depth, width, or resolution one at a time, EfficientNet comes up with a compound scaling method that adjusts three dimensions simultaneously and proportionally [107]. As a result, it achieves higher performance with significantly fewer parameters and less computational cost than the traditional models. Essentially, the core of EfficientNet-B0 is an initial 3×3 convolutional layer, followed by a series of Mobile Inverted Bottleneck Convolution (MBConv) blocks, and a final classification stage with a fully connected layer and softmax activation (Figure 5.4). The MBConv blocks, originating from MobileNet, are designed in two versions: MBConv1 and MBConv6. These blocks use 1×1 convolutions, depthwise convolutions, and Squeeze-and-Excitation (SE) modules, which recalibrate the channel-wise feature activations learned from the input data to allow the most informative features to be highlighted. On top of that, the network uses the Swish activation function which is a better gradient flow than the usual one when standard ReLU is employed. These are the reasons that led to the choice of EfficientNet-B0 in this research work:

- Lightweight design: the number of parameters is dramatically less than traditional architectures such as VGG16 or ResNet201, so it is possible to use it for applications with limited resources.
- Good accuracy-efficiency balance: this model attains a performance level that is comparable to other larger models in the field, despite its smaller size, which makes it perfect for environments with potential mobile deployment.

- Input size compatibility: the B0 version is designed for input images of 224×224 pixels, which is exactly the resolution of our dataset, thus a natural match occurs without any resizing procedures.

Although EfficientNet-B0 is very effective, the relatively smaller depth and width of EfficientNet-B0 as compared to other EfficientNet variants may result in its insufficient capability of capturing highly intricate patterns. This limitation is, however, a compromise in our study, where we aim at demonstrating the feasibility of using scalable and mobile-ready solutions while keeping consistency intact across the dataset partitions, and improving the performances of CNN with DWT.

Table 5.7. Comparison of parameter counts and input sizes for selected CNN architectures.

Architecture	Parameters (M)	Input size	Notes
VGG16	14.8	224×224	Classical deep CNN, high memory use
ResNet50	23.6	224×224	Residual connections, robust training
DenseNet201	18.3	224×224	Dense connections, high accuracy, heavy
MobileNetV2	2.3	224×224	Very lightweight, reduced accuracy
EfficientNet-B0	5.3	224×224	Balanced accuracy/efficiency

To recap, EfficientNet-B0 is indicative of a present-day solution that still manages to challenge conventional deep CNNs in terms of classification capabilities. However, it achieves this only with a small amount of computation. Also, its small size makes it especially interesting for possible mobile devices that will be used by field engineers during quick inspections after earthquakes.

5.3.2 Data-fusion techniques using Wavelet transforms

In addition to the standard RGB images, this work exploits features extracted through the Discrete Wavelet Transform (DWT), with the aim of enriching the input representation and improving the classification performance of the network.

The Wavelet Transform is a powerful tool for time–frequency analysis. Unlike the Discrete Fourier Transform (DFT), which decomposes a signal into sinusoidal basis functions spread over the entire domain, the DWT uses spatially localized basis functions obtained by dilations, translations, and scalings of a common *mother wavelet* [64]. In this study, the Haar wavelet was selected for its simplicity and computational efficiency, particularly suitable for capturing local edges and discontinuities.

The one-dimensional DWT is implemented through a two-channel filter bank:

- the *low-pass* output produces a down-sampled version of the signal, capturing its low-frequency content;
- the *high-pass* output retains the localized high-frequency information, such as discontinuities or sharp gradients.

The two-dimensional DWT is obtained by first applying the one-dimensional transform along the image rows and then along the columns. This process yields four sub-band matrices: the *approximate coefficients* (A), containing low-frequency information, and three

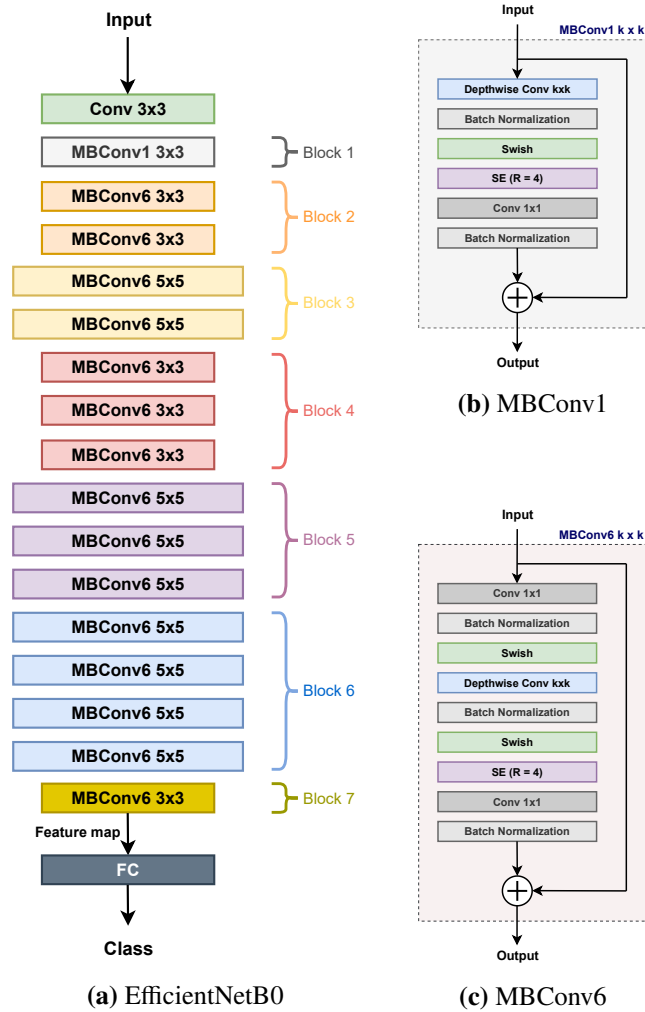


Figure 5.4. The EfficientNetB0 architecture (a), and details of the MBConv1 (b) and MBConv6 (c) blocks. Image adapted from [96]

sets of *detailed coefficients*—horizontal (H), vertical (V), and diagonal (D)—which encode high-frequency content in different orientations (Fig. 5.5) [81]. In the present work, these detailed coefficients were concatenated channel-wise to build a three-channel, image-like representation, later upsampled to 224×224 pixels for compatibility with the original RGB inputs.

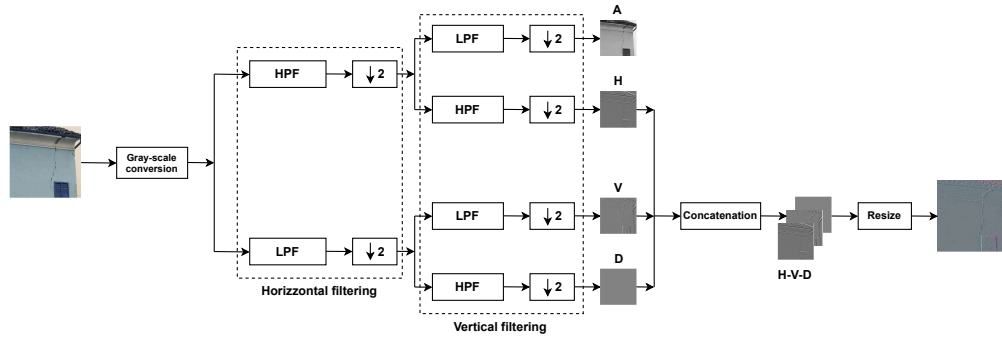


Figure 5.5. Schematic flow of the used decomposition implemented by the two dimensional discrete wavelet transform (DWT). LPF: low-pass filter, HPF: high-pass filter. A, H, V, and D represent the approximate, horizontal, vertical, and diagonal detailed coefficients. Image adapted from [96]

By combining the original images with the wavelet-based representation, two complementary *data-fusion strategies* were designed to integrate spatial and frequency-domain information (Fig. 5.6):

1. **Early Fusion (EF).** RGB images and wavelet-detailed coefficients are concatenated at the input level, generating a single unified input tensor. This allows the network to learn feature interactions jointly from both spatial and frequency information from the very first convolutional layers (Fig. 5.6a).
2. **Intermediate Fusion (IF).** In this case, RGB images and wavelet coefficients are first processed independently through two parallel branches of the EfficientNet-B0 backbone. Each branch specializes in extracting discriminative features from its input modality. The intermediate feature maps are then concatenated at a later stage and passed to a fully connected layer with 1,024 units, followed by the final softmax classifier (Fig. 5.6b). This strategy preserves modality-specific learning before integrating the information into a joint representation.

Both strategies aim to exploit the complementary nature of spatial and spectral-domain information. While EF promotes early interactions between modalities, IF enforces specialization before fusion. As discussed in the following sections, this dual approach allows testing which strategy is more effective in capturing subtle damage features, ensuring critical details are preserved and improving overall classification accuracy.

5.3.3 Training strategy: Gradual unfreezing

A *gradual unfreezing* strategy was used in the training procedure which was meant to guarantee strong convergence and reduce the risk of overfitting. As per the transfer-learning model, the convolutional core was first given the weights pre-trained on ImageNet, and the

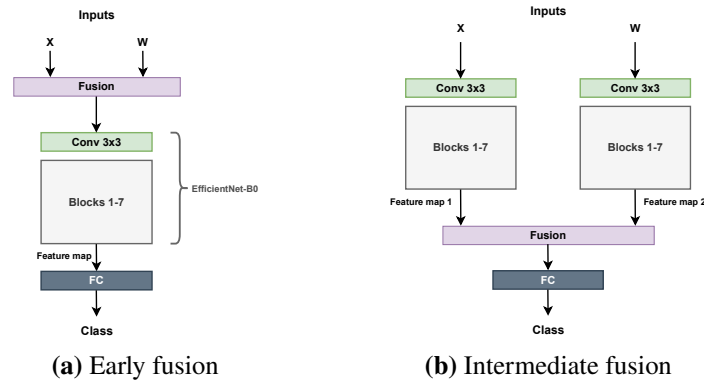


Figure 5.6. Proposed early (a) and intermediate (b) fusion strategies. X: input images, W: wavelet detailed coefficients, FC: fully connected layer. The EfficientNet-B0 architecture and details of blocks 1–7 are shown in Fig. 5.4a. Image adapted from [96]

network was adjusted in several stages, where the depth of the trainable layers increased progressively.

- **Dataset split and optimization.** The validation set was created by randomly choosing ten percent of the training data, while the remaining samples were used for training the model in a supervised manner. The batch size was 16 samples. The categorical cross-entropy loss was minimized using the Adam optimizer which was the chosen training method. To smooth out the convergence, the “Reduce LR on Plateau” scheduler was used: if the validation loss had reached its plateau for five epochs, then the learning rate was cut down by a factor of five. The learning rate was initially set to 10^{-4} and the rest of the Adam hyperparameters were at their default values.
- **Phase 1: partial fine-tuning.** Twenty last layers of the network were unfrozen and trained for 100 epochs, with the early-stopping patience of 10 epochs. All the frozen layers were left untouched.
- **Phase 2: deeper adaptation.** The last 100 layers were unfrozen and trained for another 60 epochs, keeping the same early-stopping criteria.
- **Phase 3: full fine-tuning.** The unfreezing and joint fine-tuning were conducted on the layers for 30 epochs, patience in the early stopping criterion was set to 5 epochs to prevent and mitigate overfitting issues.

During all three stages, batch-normalization layers were kept fixed in order to maintain the statistical characteristics transferred from ImageNet pre-training. This step-wise training method, called *gradual unfreezing*, had also been implemented with good results in other areas like natural language processing [43], and it turned out to be equally efficient here in stabilizing the network’s gradual adaptation to the seismic-damage field.

5.3.4 Discussion and results: Selection of the best model

The performance of the different fusion strategies was evaluated using standard metrics: overall accuracy, per-class precision, recall, F1-score, and their weighted averages. Sim-

ulations were executed on a workstation equipped with an Intel i9-14900KF CPU (14th generation), 32 GB RAM, and a GeForce RTX 3050 GPU with 8 GB of memory. The entire framework was implemented in Python using the Anaconda environment, and all source code is publicly available¹.

For what concerns the **data-fusion strategies**, three approaches were compared:

- EfficientNet-B0 trained only on RGB images;
- Early Fusion (EF), concatenating RGB images with wavelet-detailed coefficients at the input level;
- Intermediate Fusion (IF), processing RGB and wavelet representations in parallel branches before concatenation at a feature level.

The main findings can be summarized as it follows:

- The baseline RGB-only model achieved an accuracy of 84.4%.
- Early Fusion improved performance to 86.0%.
- Intermediate Fusion yielded the best results, with 86.9% accuracy.

The most challenging class was *Moderate*, which exhibited lower precision and F1-scores. However, the Intermediate Fusion strategy alleviated this issue: the F1-score increased from 0.788 (RGB-only) to 0.816 (IF). Confusion matrices (Fig. 5.7) confirmed that IF reduced the misclassification of “Slight” and “Undamaged” samples as “Moderate,” while the “Heavy” class was identified with particularly high reliability—a crucial feature for safety-critical applications.

Regarding the **comparison with state-of-the-art**, experimental outcomes were also compared to the performances of famous CNN architectures like AlexNet, GoogLeNet, DenseNet201, MobileNet-V2, as well as the CNN model by [9]. The fusion strategies suggested in this section had better performances than the baselines with accuracy improvements ranging from 3 to 5%, and with the respect to the simple EfficientNet-B0 ranging around 2-2.5%.

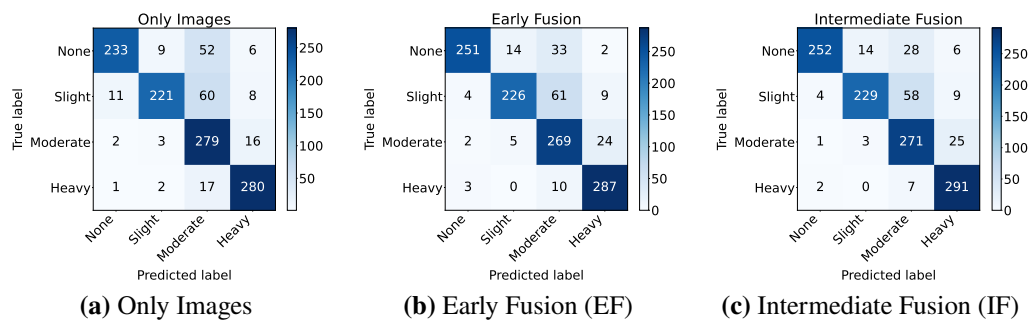


Figure 5.7. Confusion matrices related to the EfficientNet-B0 using: only images (a), early fusion (b), and intermediate fusion (c).

¹<https://github.com/mscarpiniti/Damage-Data-Fusion>

Table 5.8. Per-class results related to the classification of the damage level. Result for the three cases: (Only Images: EfficientNetB0 results using the RGB dataset; Early Fusion: EfficientNetB0 results using the RGB-DWT dataset with EF approach; Intermediate Fusion: EfficientNetB0 results using the RGB-DWT dataset with IF approach;

Class	Accuracy	Precision	Recall	F1-score
Only Images				
None		0.943	0.777	0.852
Slight		0.940	0.737	0.826
Moderate		0.684	0.930	0.788
Heavy		0.903	0.933	0.918
All classes	84.42	0.868	0.844	0.846
Early Fusion (EF)				
None		0.965	0.837	0.896
Slight		0.922	0.753	0.829
Moderate		0.721	0.897	0.799
Heavy		0.891	0.957	0.923
All classes	86.08	0.875	0.861	0.868
Intermediate Fusion (IF)				
None		0.973	0.840	0.902
Slight		0.931	0.763	0.839
Moderate		0.745	0.903	0.816
Heavy		0.879	0.970	0.922
All classes	86.92	0.882	0.869	0.870

Table 5.9. Comparisons of classification accuracy with the state of the art.

Approach	Accuracy	Precision	Recall	F1-score
Baseline CNN [9]	82.33	0.838	0.823	0.831
AlexNet [93]	82.75	0.840	0.828	0.827
DenseNet201 [93]	84.66	0.862	0.847	0.855
GoogLeNet [93]	84.75	0.858	0.848	0.853
MobileNet-V2 [91]	83.25	0.848	0.833	0.840
EfficientNet-B0 [106]	84.42	0.868	0.844	0.846
Proposed EF	86.08	0.875	0.861	0.868
Proposed IF	86.92	0.882	0.869	0.870

In the end, for the **best model selection**, the Intermediate Fusion approach combined with EfficientNet-B0 was finally chosen as the best solution based on multiple criteria. Its competence to combine co-existing local and spectral-domain info, while keeping stability for all classes, was a clear indication of the usefulness of wavelet-based data fusion. The accuracy of the final implementation was close to 87%, which means it can be effectively used in the field after an earthquake, in accordance with AeDES criteria for Italian seismic evaluations.

5.4 A Vision Transformer based Application

In this section, we present an application of Vision Transformer (ViT) architectures for the task of post-earthquake building damage classification, based on the concepts introduced in the paper of Dosovitskiy (2021) [21]. Four model configurations will be explored that vary in size and patch resolution, and they will be evaluated in performances on the curated multi-source dataset described in Chapter 5. The goal is to assess how well the transformer-based models can detect damage patterns in structures and further by testing their results against typical CNN-based methods.

5.4.1 B-16 and B-32 Models

The **Base** configuration of the Vision Transformer (ViT-B) represents the most compact version among the architectures considered in this study. It uses 12 transformer layers, a hidden size of 768, and a multi-layer perceptron (MLP) size of 3072 with 12 attention heads. Two variants are distinguished according to the patch size:

- **ViT-B/16:** the input image of 224×224 pixels is split into 16×16 patches, yielding 196 tokens. This configuration captures more fine-grained structural details, which is beneficial for tasks such as crack and spalling detection.
- **ViT-B/32:** the same input is split into 32×32 patches, resulting in 49 tokens. This version reduces computational cost but at the expense of spatial granularity.

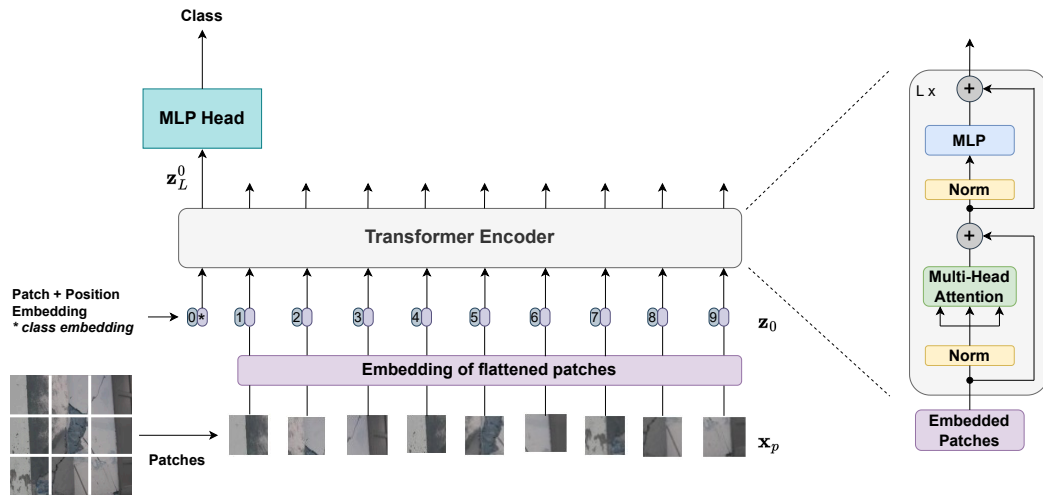


Figure 5.8. An overview scheme of the Vision Transformer architecture (ViT). Image adapted from [95]

Both variants balance model complexity with feasibility for training on relatively limited datasets, making them particularly suitable for proof-of-concept experiments in post-disaster contexts.

5.4.2 L-16 and L-32 Models

The **Large** configuration (ViT-L) extends the model depth and width, with 24 layers, a hidden size of 1024, an MLP size of 4096, and 16 attention heads. This results in more than 300 million trainable parameters, substantially increasing representational capacity. As for the base models, two patch sizes were explored:

- **ViT-L/16:** with 196 tokens per image, this model has the potential to extract highly detailed spatial and contextual features.
- **ViT-L/32:** with 49 tokens per image, this version is more efficient but less precise in capturing localized damage.

While these configurations are powerful, their performance is strongly influenced by the dataset size and diversity. In the context of post-earthquake imagery, where annotated data are limited, the advantage of the larger architectures may not always be fully exploited.

5.4.3 Training Strategy, Discussion and Results

All models were initialized from ImageNet-pretrained weights and fine-tuned on the target dataset of 19,312 augmented training samples. A validation split of 10% was used to monitor generalization. The training employed categorical cross-entropy loss, optimized with Adam and a learning rate scheduler using the “Reduce LR on Plateau” policy. Label smoothing with parameter 0.1 was introduced for additional regularization. The batch size was set to 32.

Summaries of the four models (see Table 5.10,) showed that the overall accuracies on the balanced test set were between 85.6% and 87.3%. The values of weighted F1-scores were within 0.867 to 0.883. More exactly:

- **ViT-B/16** was the most accurate model reaching 87.3% and 0.883 weighted F1-score. One of the reasons of its great performance was that it was very effective for the “Heavy” damage class and there it got an F1-score that was almost 0.95.
- **ViT-B/32** and **ViT-L/16** were not far behind as they performed at a comparable level and the accuracy of both were higher than 86% while their F1-scores were close to 0.868.
- **ViT-L/32** was a little low in overall accuracy (85.6%) but it still had good metrics for each class and the model was also very stable when it came to differentiating severe damage.

As seen from confusion matrices in Figure 5.9., the main source of misclassifications was the mixing with the closest neighboring levels of damage (for example, “Slight” and “Moderate”). This is because the subjectivity of such labels and the visual likeness of those cases that lie very close to the limit between two classes.

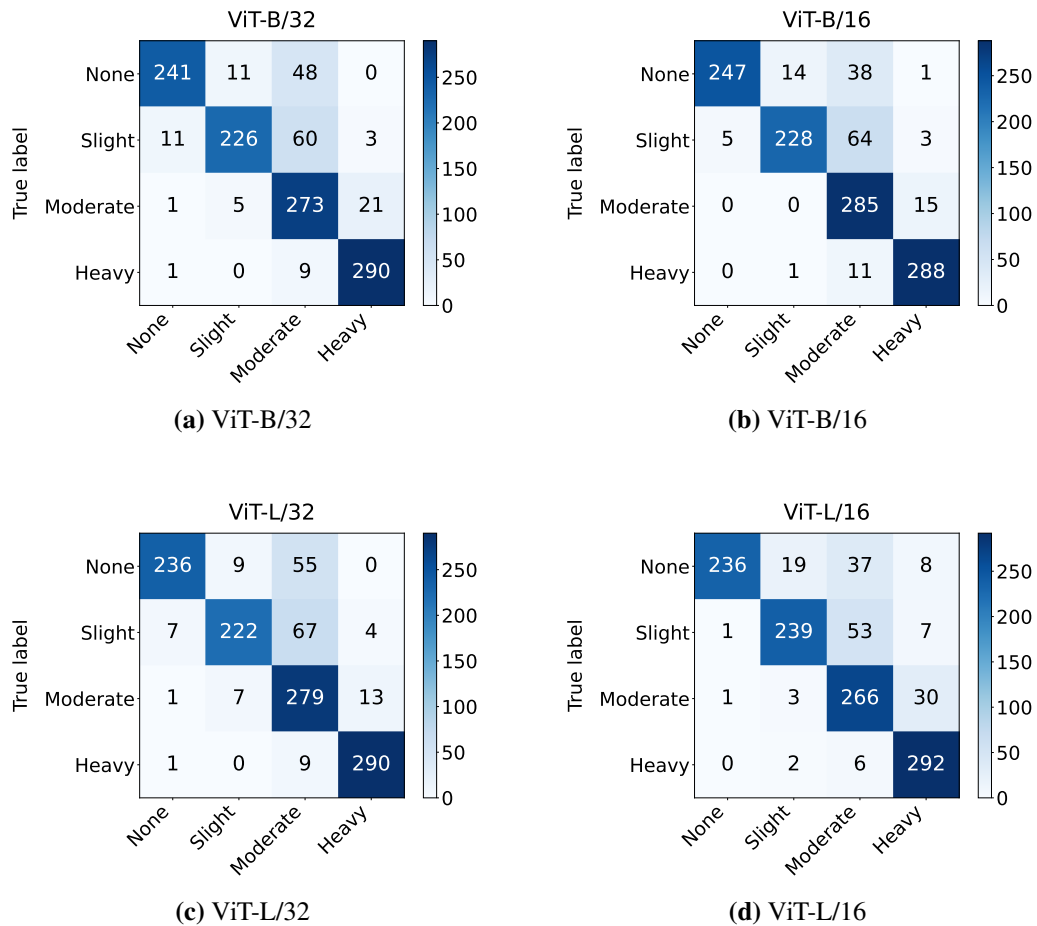


Figure 5.9. Confusion matrices related to: ViT-B/32 (a), ViT-B/16 (b), ViT-L/32 (c), and ViT-L/16 (d).

The experimental findings highlight several important aspects of applying ViTs in earthquake damage assessment:

Table 5.10. Per-class and weighted average classification results of the damage level provided by the considered ViT models.

Class	Accuracy	Precision	Recall	F1-score
ViT-B/32				
None		0.949	0.803	0.870
Slight		0.934	0.753	0.834
Moderate		0.700	0.910	0.791
Heavy		0.924	0.967	0.945
All classes	85.83	0.877	0.858	0.867
ViT-B/16				
None		0.980	0.823	0.895
Slight		0.938	0.760	0.840
Moderate		0.716	0.950	0.817
Heavy		0.938	0.960	0.949
All classes	87.33	0.893	0.873	0.883
ViT-L/32				
None		0.963	0.787	0.866
Slight		0.933	0.740	0.825
Moderate		0.680	0.930	0.786
Heavy		0.945	0.967	0.956
All classes	85.58	0.880	0.856	0.868
ViT-L/16				
None		0.992	0.787	0.877
Slight		0.909	0.797	0.849
Moderate		0.735	0.887	0.804
Heavy		0.867	0.973	0.917
All classes	86.08	0.875	0.861	0.868

- **Patch size:** it was found out that smaller patches (16×16) generally lead to better results as they keep the local structural details intact which is very important to separate the cracks or spalling that arise from the surface of the materials that seem to be intact.
- **Model size vs. dataset size:** although bigger models (ViT-L) have higher representational capabilities, they might not be able to demonstrate all their potential with a small amount of training data. In contrast, ViT-B/16 is the one that achieves the furthest balance between accuracy and computational resource leaving.
- **Comparisons with CNNs:** no matter which ViT variant was used, all of them outperform the results of the traditional CNN-based architectures (AlexNet, GoogLeNet, DenseNet201, EfficientNet-B0), with the improvement of performance going up to 5% both in terms of accuracy and F1-score. It confirms that transformers-based solutions can be efficient enough to be employed in structural damage classification tasks.
- **Error analysis:** most of the residual confusions were about the intermediate classes, while the cases at the extreme ends (“None” vs. “Heavy”) were distinguished with high accuracy. Hence, this point suggests that combining ViT with other situational aid (for example, metadata or multimodal fusion) can make the ambiguities even less.

In general, this section confirms that Vision Transformers are powerful models for post-disaster building assessment. The best model out of all the tested ones is ViT-B/16 as it offers a good balance between the effectiveness and the computational resource which makes it a very useful candidate for being deployed in the field of rapid emergency response where time is critical.

5.5 Other Approaches

In addition to the primary CNN-based classification methods (such as VGG16) used elsewhere in this work, we explored several alternative architectures and pipelines to assess whether they could offer improvements or complementary advantages. These “other approaches” include (i) hierarchical classifiers built around the best CNN (VGG16), (ii) convolutional autoencoder (CAE)-based feature extraction followed by traditional classifiers, and (iii) preliminary investigations into quaternion-based CNNs. Across these experiments, the performance obtained is broadly comparable to the CNN-based baselines, with some trade-offs in complexity, interpretability, or data demands. It is worth noting that these explored methods and architectures, while yielding promising and comparable results to our CNN baselines, are not exhaustively examined within the present thesis. Future work will involve deeper investigation into optimal configurations, hyperparameter tuning, and validation across larger, more diverse datasets to more fully assess their potential and limitations beyond the scope of this research.

5.5.1 Hierarchical CNN-based Classification

Here we constructed hierarchical classification pipelines using VGG16 as backbone. The idea was to decompose multiclass damage severity into a series of binary decisions: at first “damage vs no damage”, then “slight vs not slight”, etc., progressively refining until all severity classes are reached. This hierarchical strategy aimed to reduce confusion between

similar classes by focusing each binary stage on more separable pairs. In our experiments, the results from this hierarchical VGG16 classifier were similar in accuracy and F-score to the flat, direct multiclass VGG16 classification, though sometimes showing slightly better recall on certain classes (especially “slight” vs “moderate”).

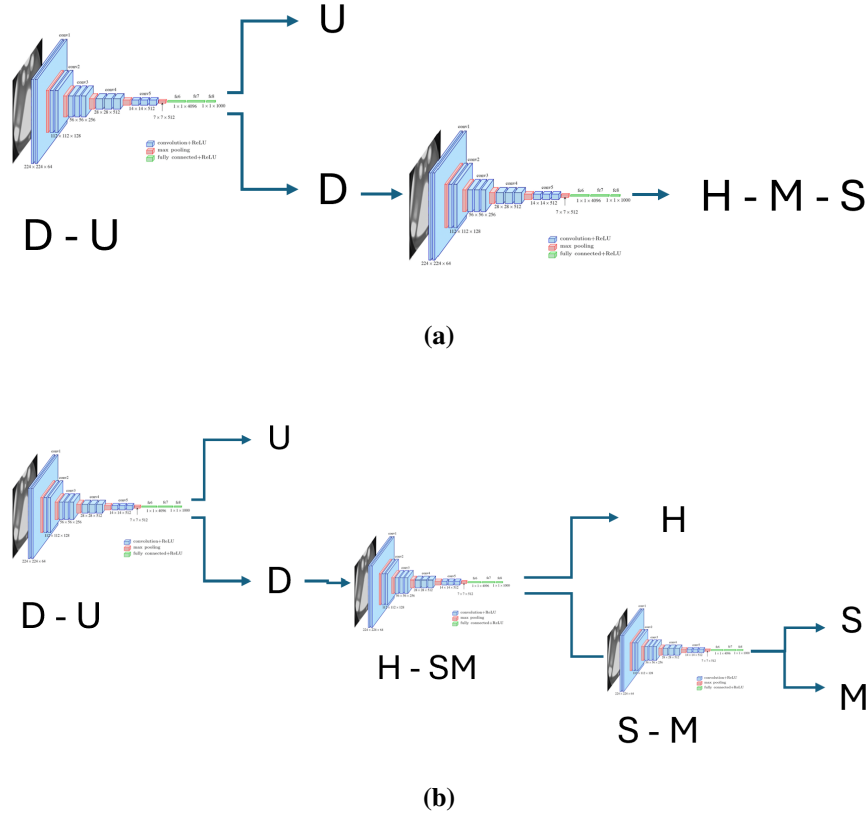


Figure 5.10. Schematic flows of the Hierarchical classification possible configurations: (a) first approach; (b) second approach. (H= Heavy, M=Moderate,S=Slight, U=Undamaged, D=Damaged)

5.5.2 CAE-based Feature Extraction with Traditional Classifiers

A second line of exploration involved convolutional autoencoders (CAEs). A CAE consists of an encoder that compresses the input image into a lower-dimensional latent space, and a decoder that reconstructs the input from that compressed representation. Several works in literature support the idea that CAE latent features can be effective in damage classification. For example, Maggipinto et al. (2018) used a CAE in a semi-supervised setting to clean noise and extract discriminative features, improving performance of downstream classifiers [62]. Similarly, Pintelas et al. (2021) demonstrated that CAEs can filter redundant information and produce latent embeddings that increase classification accuracy in image-based tasks [84]. Luo et al. (2023) applied CAE in unsupervised structural damage detection, useful in settings with limited labels [59].

In our pipeline:

- We trained three different CAE architectures using only RGB images for 100-200 epochs, varying whether pooling layers were used, and whether skip-connections were included.
- After training, we discarded the decoder and used the encoder (up to a flatten or global average pooling layer) to extract features (latent representations) from the images.
- These latent features were then fed into various classifiers (Dense layers, Random Forest, MLP, etc.) trained to predict the four damage classes.

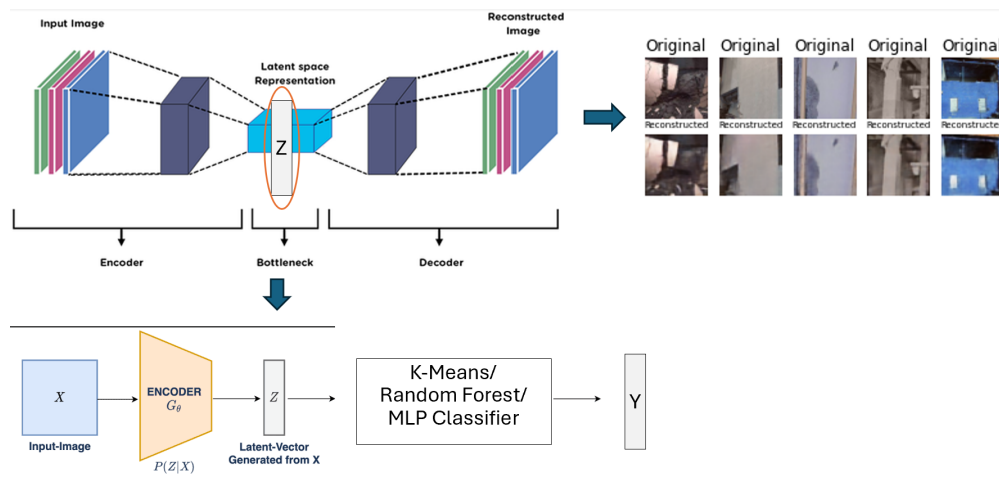


Figure 5.11. Schematic flow of the CAE classification.

Table 5.11. Summary of CAE Experiments on RGB Damage Images

CAE Variant	Pool / Skip	Epochs	Classifier	Accuracy	Other Metrics
CAE-1 (with pool- ing)	Yes/No	100	Dense softmax	83.94%	AUC=0.9753, MCC=0.7886
CAE-1 (with pool- ing)	Yes/No	200	Dense softmax	84.25%	AUC=0.9318, MCC=0.7925
CAE-2 (no pool- ing)	No/No	200	Dense softmax	77.47%	AUC=0.890, MCC=0.7043
CAE-3 (no pool- ing + skip connections)	No/Yes	200	Dense softmax	80.14%	AUC=0.882, MCC=0.7395
CAE-1 latent-features + Random Forest	Yes/No +	200	Random Forest	85.31%	F1=0.8592, AUC=0.9752, MCC=0.8084

Across the CAE variants tested:

- CAE1 (with pooling, trained for 200 epochs and using flatten) achieved 84.3% accuracy, high AUC (0.93), MCC 0.79.
- CAE2 (without pooling) and CAE3 (without pooling but with skip-connections) had lower performance (77-80% accuracy).
- Among classifiers, Random Forest fed with latent features from CAE1 gave the best results, closely matching performance of direct CNNs.

Thus, the CAE-based pipelines confirm that feature extraction via CAE can be a viable alternative, especially when one wants more compact or interpretable representations, although they did not clearly surpass the best CNN baselines in our experiments.

5.5.3 Quaternion CNNs (QCNN) and Quaternion-based Architectures

As a further exploratory direction, we considered the possibility of using quaternion-valued convolutional neural networks. Quaternion CNNs operate on inputs in a quaternion algebra domain, this means treating the color channels in a coupled way (e.g., treating RGB channels jointly as a quaternion), which can allow more efficient learning of inter-channel relationships. Recent studies have shown that QCNNs can sometimes achieve similar performance to real-valued CNNs with fewer parameters and greater robustness to rotations or color transformations [2]. While in this work we have not yet implemented a full quaternion VGG16 version, we evaluated relevant literature and consider QCNNs a promising future extension for damage classification tasks in RGB imagery.

5.5.4 Summary of Comparative Findings

Overall, the alternative methods explored (hierarchical CNNs, CAEs, QCNN architectures) performed in line with existing CNN baselines. The best results in our tests came from the CAE and Random Forest combination, achieving 84-85% accuracy, which is close to the VGG16 direct multiclass approach. The advantages of the CAE pipelines are more visible in lower data settings or when interpretability of latent features is needed. QCNNs remain to be fully assessed in the specific context of post-earthquake RGB damage imagery, but available literature suggests worthwhile potential.

Chapter 6

Post-Earthquake Residual Capacity Assessment

The results obtained in Chapter 5 have demonstrated the feasibility of using Convolutional Neural Networks (CNNs) to automatically classify damage levels in structural and non-structural elements of buildings after seismic events. Beyond the pure classification task, such information can play a crucial role within a **broader framework for seismic residual capacity assessment of damaged buildings**. In fact, linking visual evidence of damage to modifications of capacity reduction factors for damaged structural components provides a novel opportunity to embed AI-driven tools into seismic safety evaluation and decision-making processes. This chapter introduces a preliminary framework for integrating CNN-based damage classification into **post-earthquake residual capacity assessment**, building upon the insights of previous research studies and recent proposals for simplified analytical–mechanical methods [78, 94]. After a major earthquake, structural engineers must quickly evaluate not only the visible damage to buildings but also their residual capacity, the remaining strength and stiffness that determine if a structure can still safely resist future loads [75].

The reliance of traditional post-earthquake evaluation on prompt visual inspections and “tagging” (green/yellow/red tags for safe/limited use/unsafe) based on the subjective judgment of experts is well-known. Such a method certainly is effective when it comes to immediate safety checks; however, it does not generate a capacity loss measure of the building, which, in turn, is directly related to the re-occupancy, repair, or demolishing of the structure. On the other hand, a comprehensive and quantitative examination of the **residual capacity, the ability of a structure to sustain subsequent seismic events**, is a matter of vital information necessary to understand if a structurally compromised building can survive an aftershock or if it has to be retrofitted, or demolished and reconstructed. Interestingly enough, with respect to such matters as post-earthquake assessments, there are certain technological achievements, especially in the field of AI and computing. CNNs, for instance, as seen in this research, provide an image damage classification functionality that can be standardized and automated. By doing so, the whole process may become not only swift, but also objective and informative. As presented in this chapter, a model of post-earthquake residual capacity assessment has been described that allows for the integration of the so-called visual damage classification (and, in particular, the contemporary CNN-based methods) with the structural part of the system.

This chapter presents a framework for post-earthquake residual capacity assessment that integrates visual damage classification (including modern CNN-based methods) with structural analysis. A review of relevant studies is first provided, followed by a discussion on how CNN-based damage classification can be incorporated into the workflow. An illustrative case study on a reinforced concrete (RC) building demonstrates the approach, and the chapter concludes with considerations on linking damage classification to repair cost estimation and overall post-earthquake decision-making.

6.1 Framework Concept and Workflow

In current practice, the evaluation of the residual seismic capacity of earthquake-damaged structures relies on updating an analytical model to reflect the observed damage state. A well-established reference is FEMA 306 [23], which prescribes empirical capacity-reduction factors (so-called λ -factors) to modify the non-linear force–deformation relationships of structural components affected by damage. These modifiers reduce stiffness, strength, and ductility according to the severity of damage observed during post-event inspections. The updated model, when subjected to nonlinear static analysis (pushover), provides a new capacity curve representative of the damaged configuration, typically showing lower stiffness, reduced peak strength, and impaired ductility. Such curves are then combined with seismic demand spectra (e.g. through the Capacity Spectrum Method [15]) to evaluate the residual safety of the building. This conventional workflow involves three main steps: inspection, model updating, and residual performance assessment.

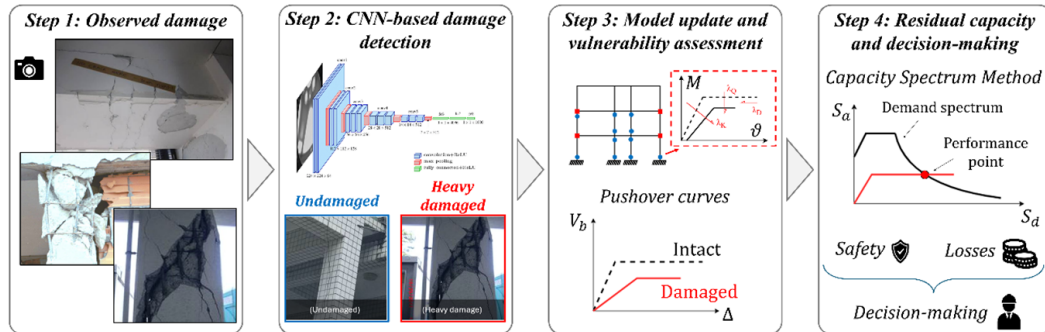


Figure 6.1. Flowchart for vision-based residual capacity assessment of earthquake-damaged buildings (Image from Saquella et al. 2025 [94]).

Building on this foundation, the proposed *vision-based framework* (Fig. 6.1) integrates automated image-based damage classification into the residual capacity assessment pipeline. The workflow is designed to couple rapid computer vision outputs with structural mechanics models, accounting also for classification uncertainty, as demonstrated in Pedone et al. [83]. The process can be summarized as follows:

1. **Post-earthquake data collection.** Immediately after an event, photographic evidence of structural damage is gathered via drones, ground surveys, or on-site inspections. Images capture cracks, spalling, bar exposure, and other indicators in structural and non-structural components.

2. **Automated damage detection and classification.** The collected images are processed using convolutional neural networks (CNNs), such as fine-tuned VGG16 models, which identify the type of component and assign a standardized damage state (e.g. no, slight, moderate, heavy, complete).
 - The procedure aligns with the AeDES classification scale commonly used in Italy.
 - Classification performance is quantified through a confusion matrix, which reports the probability of correct and incorrect labelling.
 - To propagate this uncertainty, Monte Carlo sampling is performed: for each component, damage states are sampled according to confusion-matrix probabilities, producing thousands of alternative realizations of the damaged building.

3. **Mapping to mechanical parameters.** Each classified damage state (or sampled state in the Monte Carlo analysis) is mapped to mechanical reduction factors:

$$K^{\text{dam}} = \lambda_K K^{\text{intact}}, \quad Q^{\text{dam}} = \lambda_Q Q^{\text{intact}}, \quad D^{\text{dam}} = \lambda_D D^{\text{intact}}$$

where λ_K , λ_Q , and λ_D reduce stiffness, strength, and ductility according to FEMA 306 guidelines [23]. Severe states (DS3–DS4) result in substantial reductions, while slight or moderate damage (DS1–DS2) lead to milder adjustments.

4. **Updating the structural model.** The as-built numerical model (intact configuration) is modified by applying the capacity-reduction factors to the plastic hinges of damaged members. Undamaged components retain their original properties. Through this process, thousands of post-earthquake variants of the building are generated, reflecting both observed damage and CNN misclassification uncertainty.
5. **Residual capacity estimation.** Nonlinear static analyses (pushover) are conducted on both intact and damaged configurations. Two complementary approaches are employed:
 - *Simplified analytical model (SLaMA).* Extended to account for post-earthquake conditions [68, 75], SLaMA offers fast capacity-curve estimation with acceptable accuracy.
 - *Refined numerical model.* A detailed pushover analysis of a refined numerical (software-based) model of the structure provides a benchmark to validate the simplified method.

The outcome is a family of capacity curves: intact vs. damaged, deterministic vs. uncertainty-aware (Monte Carlo realizations). Typically, damaged curves show decreased peak base shear and a reduction in terms of displacement and ductility capacity.

6. **Safety evaluation and decision support.** Capacity curves are compared with seismic demand spectra using the Capacity Spectrum Method [15]. The residual seismic safety is quantified through the IS-V index (Indice di Sicurezza–Vulnerabilità) [14], equivalent to the %NBS metric internationally [71].

- Clearly, the damaged configuration is expected to exhibit a lower (or at most the same) safety index compared to the intact configuration.
- Monte Carlo simulations allow evaluating the expected range of the safety index, while accounting for uncertainties in the CNN-based damage classification.

In conclusion, the proposed workflow combines the vision-based automated inspection with the mechanical model and the probabilistic treatment of classification uncertainty. This method makes it possible to provide a quantitative estimate of the residual seismic capacity in a very short time, thus, it effectively bridges the gap between rapid reconnaissance and detailed structural analysis. It is, therefore, a valuable tool for emergency management, risk classification, and selection of the most suitable retrofit strategies after a major earthquake.

6.2 Review of previous studies

Several contributions in the literature have highlighted the importance of linking observed damage to residual seismic capacity. FEMA 306 (1998) first introduced the concept of capacity reduction factors, relating observed damage in reinforced concrete and masonry structures to reductions in stiffness, strength, and ductility. Similar approaches were later adopted in Japan (JBDPA guidelines) and Turkey (TCIP methodology). In Europe, the EMS-98 and its Italian adaptation through the AeDES form also provided reference categories of damage, which can be operationally linked to structural parameter degradation.

More recent studies [94, 8, 88] proposed simplified analytical or mechanical procedures to derive reduction factors for reinforced concrete walls and frames. These works underline that damage-dependent recalibration of plastic hinges enables a realistic estimation of residual capacity. Other contributions (e.g., Ji et al. 2023, Miao et al. 2023) have investigated the use of AI to connect visual damage to mechanical degradation, showing the potential of combining data-driven classification and physics-based models.

Interest in post-earthquake residual capacity assessment has grown in recent years, and several studies in the literature have addressed aspects of this problem. After FEMA 306, research has focused on refining these reduction factors and extending the methodologies to different structural systems and damage scenarios. For example, researchers have derived plastic hinge modification factors for damaged RC components based on experimental data and numerical simulations. Ji et al. (2023)[50] note that incorporating observed crack characteristics (width, length, etc.) into a model via degradation factors can improve accuracy in predicting residual capacity. Ceccarelli et al. (2021) [8] performed extensive nonlinear analyses on reinforced concrete wall structures to investigate their residual capacity after earthquakes and the impact on seismic losses. In their study, stiffness and strength reduction coefficients were quantified for various damage levels in structural walls, and these were applied in a case-study building to assess post-earthquake performance. They found that the loss of lateral stiffness/strength due to damage can lead to significantly higher economic losses as the damage level increases. In fact, even moderate damage (if not accounted for in analysis) could hide a substantial loss of capacity in terms of stiffness, meaning future low-intensity events might cause disproportionately large losses when residual capacity is low. This highlighted the importance of accounting for residual capacity in loss estimation frameworks.

Several studies have specifically connected observed damage states with fragility and loss models. Rossi et al. (2022) [88] investigated how prior earthquake damage and subsequent repair interventions influence the expected annual losses of RC buildings with wall systems. Their work in Earthquake Spectra showed that neglecting the effects of unrepaired damage can underestimate future losses, whereas timely repairs can mitigate long-term risk. These findings reinforce that residual capacity (or lack thereof) has economic consequences over a building's remaining life. On the modeling front, researchers like Polese et al. (2012) [85] and Marder et al. (2020)[65] have proposed procedures to update building models after damage and evaluate their fragility to aftershocks. Chiu et al. (2021) [12] developed an approach to quantify reduction factors for damaged RC columns using a database of experiments, providing empirical validation to methods like FEMA 306. Pedone et al. (2025) [83] have recently contributed to this field by introducing a vision-based damage classification tool to support residual capacity assessment. In that study, a CNN model was used to rapidly classify damage on RC frame components, and the information was linked to analytical models to evaluate post-earthquake performance. It was demonstrated that automated image-based damage recognition could be integrated with structural analysis to produce residual capacity estimates, laying the groundwork for the integrated framework discussed in this chapter.

In summary, prior studies have:

- (1) established methods to incorporate observed damage into structural models (via stiffness/strength reduction factors),
- (2) highlighted the impact of residual capacity on economic loss and decision-making [8], and
- (3) begun to explore automated or semi-automated means of damage assessment (e.g., using computer vision) to enhance post-earthquake evaluation [28].

These contributions provide the scientific basis for developing an integrated, AI-assisted residual capacity assessment workflow.

6.3 Perspectives for integration with CNN-based classification

By implementing the residual capacity method as the framework to incorporate the CNN classification pipeline that uses the image data as input, the process of safety evaluation may become partly automated.

For instance:

- As a result of the CNNs prediction, the main advantage will be the characterizing of the damage rapidly, objectively, and through the usage of a set algorithm (which is the element level).
- From these damage categories, it will be possible to directly link the λ parameters of stiffness, strength, and plasticity, thus achieving automatic conversion of the structural model without undergoing an interpretation step by human inspectors.
- A new structural model that can almost be updated in a real-time manner so that an emergency response system with a short reaction time can be supported.

This kind of adoption is indicative of a new era, in which the AI technologies and the traditional approaches of structural engineering, this hybrid framework, can overcome such problems thanks to their data-driven and physically consistent nature.

Hence, by CNN-powered damage classification working in the aftermath of an earthquake, the capability of conducting fast mass evaluations can be realized. The key idea is to use AI to handle the time-consuming task of damage classification, supporting engineers to focus on structural analysis and decision-making. The combination of these two is the subject of most recent research. Ji et al. (2023) [50] have developed a computer vision-based framework that is able not only to detect but also to classify various categories of damage induced by the earthquake in a 3-story reinforced concrete test structure. They then took this extracted information further to estimate the residual capacity of the structure. Imaging produced from the shaking table tests was taken as input for a deep CNN that processed the data and found crack locations as well as their widths. Next, these geometric parameters were used to feed a structural model made for the prediction of the available load-bearing capacity. Along the same lines, Miao et al. (2023) [69] proposed a deep learning method for the identification of material property degradation of earthquake-damaged RC columns. In general terms, the strategy is based on the use of CNN features as the main source of information about damage and the only direct way to report the reduction of the mechanical parameters of which the research team becomes the estimator. Consequently, the capacity of the damaged column is quantitatively lowered, and the team can predict it. Obviously, these investigations are crucial in the sense of bridging the gap between computer vision and structural analysis. The researchers have achieved a feat in automated damage detection which has dual purposes: to make tagging of damage rapid, and more importantly, to streamline the implementation of engineering models for generation of realistic performance evaluations [50].

The CNN-based classification has been designed to perfectly fit into the suggested scheme of this chapter. As a front-end “damage sensor”, the CNN essentially reinterprets the raw images into a format that the engineering model can easily comprehend which is structured damage data. The approach has the following positives:

- **Speed and Scale:** Concretely, with the help of a pre-trained CNN, one can easily get through hundreds of pictures in a matter of minutes. This in turn makes the whole assessment process very fast as one can basically do the evaluations of multiple buildings in just a few minutes. Following an earthquake, such a scenario is very likely to happen as it is always extremely hard for engineers to cover the whole urban area in such a short time period. Hence, the role that machines like AI can play in this situation is to let the human experts focus on the buildings that are most affected by simply locating and automatically flagging those with severe damage levels.
- **Objectivity and Consistency:** In addition to the already highlighted features of CNNs, it is worthwhile noting that they also contribute heavily to minimizing the human factor in the classification process, which is undeniably one of the biggest sources of error in inspection results. The use of standardized damage criteria in combination with the employment of CNNs results not only in the neutralization of the human factor but also in a higher degree of reliability of the metrics obtained for residual capacity analysis.
- **Data-Driven Insights:** The huge amount of image data processed by CNNs can pinpoint patterns (for instance, most columns in the ground floor are found to be

“extremely” damaged) that concentrate the structural assessment on the likely critical weaknesses. So essentially, what it does is that it grants a detailed damage map of the building which on turn can be used to update the analytical models with a higher accuracy than if it was done by a basic manual survey.

- **Near Real-Time Decision Support:** By combining rapid classification with pre-modeled structures, it is conceivable to generate near-real-time estimates of residual capacity for important buildings. For instance, within hours of an event, authorities could have a preliminary report that a particular building has lost 30% of its lateral load capacity and poses a collapse risk if strong aftershocks occur, information that is invaluable for public safety decisions.

6.4 Illustrative Application

This section demonstrates an end-to-end workflow that integrates computer-vision-based damage recognition with post-earthquake residual-capacity assessment for an existing reinforced-concrete (RC) frame structure. The pipeline includes: (i) definition and modelling of a representative case-study RC building; (ii) simulation of a realistic post-earthquake damage scenario via nonlinear time-history analysis (NLTHA); (iii) image-based component damage classification using a convolutional neural network (CNN) and explicit propagation of classification uncertainty; (iv) pre- and post-event nonlinear static (pushover) analyses using a simplified analytical method (SLaMA) as well as a refined numerical model; and (v) spectrum-based safety evaluation following the Italian “Sismabonus” guidelines (IS-V index, equivalent to %NBS), with uncertainty quantification. Capacity-reduction factors for damaged members follow FEMA 306 recommendations [23], and the capacity spectrum method (CSM) is used for demand–capacity checks [15, 71, 14].

6.4.1 Case-Study RC Building: Geometry, Detailing, and Materials

The case study is a three-storey RC moment-resisting frame (MRF) representative of pre-1970s Italian construction, i.e., gravity-load design without formal capacity-design principles. The plan features three-bay MRFs in one principal direction and two-bay frames in the orthogonal direction (see Fig. 6.2). The structural weaknesses typical of that era include: (i) poor transverse confinement, (ii) joint panels without transverse reinforcement, and (iii) plain round bars with hooked ends for beam longitudinal reinforcement. Member detailing (illustrative values consistent with archival practice) is: beam $\varnothing 6/15$ cm; column ties $\varnothing 8/15$ cm; joint cores unconfined.

Mean material properties used in the models are: concrete cylinder strength $f'_c = 14.41$ MPa and steel yield stress $f_y = 340.51$ MPa. The building is assumed to be sited in a high-seismicity area in Italy (soil type B); a representative peak ground acceleration (PGA) level for the site hazard context is $0.307g$ (for the scenario definition below).¹

6.4.2 Simulated Post-Earthquake Scenario via NLTHA

To obtain a realistic pattern of a component damage, we perform a 2D NLTHA on the interior longitudinal frame modelled with lumped plasticity approach using the structural software

¹These values reflect the scenario adopted for method demonstration and are consistent with [83].

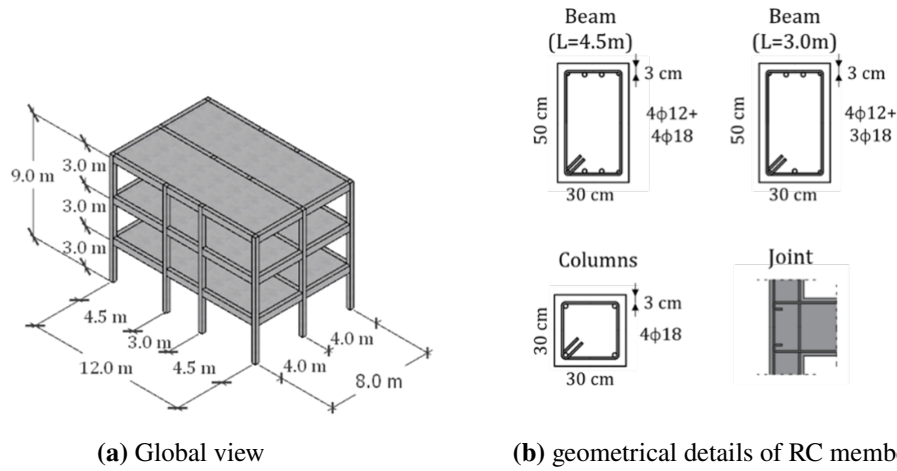


Figure 6.2. Case study RC building: (a) global view and (b) geometrical details of RC members. Image from [83].

RUAUMOKO [7], following the modelling framework in [82]. A recorded accelerogram from the SIMBAD database [103] is employed: the 1980 Irpinia earthquake, waveform ID 414 (event ID 46, station BGI, $M_w = 6.9$, epicentral distance 21.79 km), Y direction with $PGA_Y = 0.13g$. This “proof-of-concept” dynamic excitation produces interstorey drift ratios concentrated primarily in the first and second storeys (Fig. 6.3), consistent with the absence of capacity design and a soft/intermediate-storey mechanism.

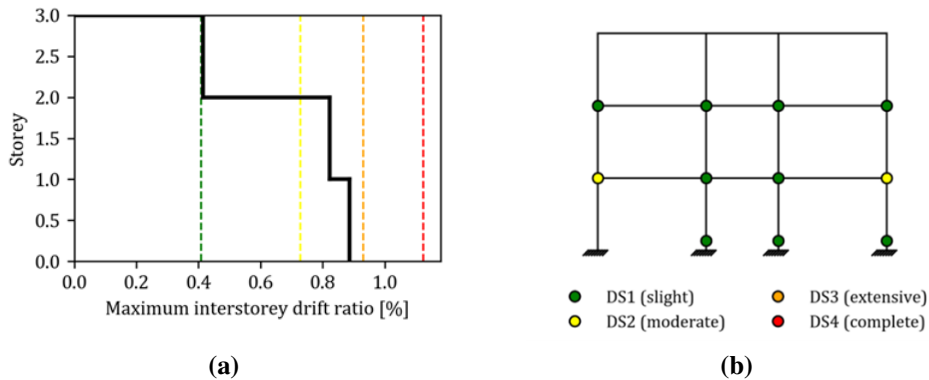


Figure 6.3. Simulated earthquake scenario: (a) maximum interstorey drift ratios and (b) observed earthquake damage to structural members. Image adapted from [83]

Damage state mapping. Component-level damage states (DS) are assigned by post-processing peak inelastic deformations:

- DS1 (*slight*): onset of yielding;
- DS2 (*moderate*): 0.50 of ultimate plastic deformation capacity;
- DS3 (*extensive*): 0.75 of ultimate plastic deformation capacity;

- DS4 (*complete*): attainment of ultimate plastic deformation.

Under the selected record, the pattern is typical of non-ductile RC frames: moderate damage at exterior beam–column joints at the first storey; slight damage in base columns and remaining joints at the first two storeys; negligible/no damage at the third storey (Fig. 6.3b). This distribution will be used to define the *true* (numerical) post-event state for validation of the CNN-based pathway.

6.4.3 CNN-Based Damage Classification and Explicit Uncertainty Modelling

A CNN trained on earthquake damage images is used to classify component images into DS0 (undamaged)–DS4 (complete). Beyond deterministic labels, we explicitly propagate *classification uncertainty* by using the model’s confusion matrix (Fig. 6.4), as done in [83].

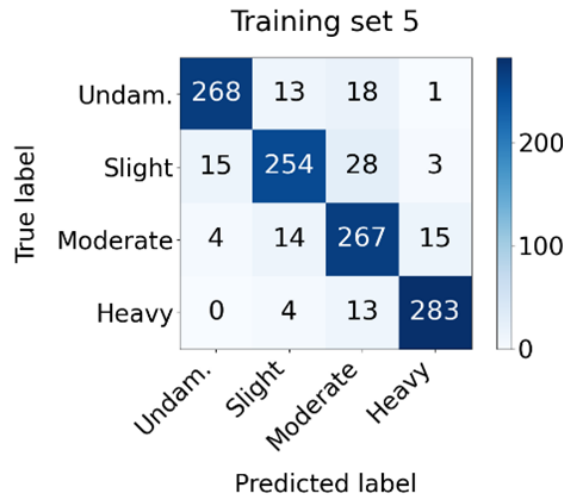


Figure 6.4. CNN (VGG-16) confusion matrix results, from chapter 5, used for uncertainty propagation.

Let $\mathbb{P}(\widehat{DS} = j \mid DS = i)$ denote the confusion-matrix probability that a true class i is predicted as j . A Monte Carlo (MC) scheme generates $N = 2,000$ alternative *realizations* of the damaged structure by sampling, for each beam–column–joint subassembly, a random label $\widehat{DS}$ from the row of the confusion matrix corresponding to the nominal (observed) class DS .

Capacity-reduction factors. Each sampled damage state $\widehat{DS}$ at a component is associated with capacity modifiers λ drawn from FEMA 306 tables [23]. In the structural model, we apply strength, stiffness, and ductility reductions to the corresponding plastic hinges:

$$Q_i^{\text{dam}} = \lambda_Q(\widehat{DS}) Q_i^{\text{intact}}, \quad K_i^{\text{dam}} = \lambda_K(\widehat{DS}) K_i^{\text{intact}}, \quad D_i^{\text{dam}} = \lambda_D(\widehat{DS}) D_i^{\text{intact}} \quad (6.1)$$

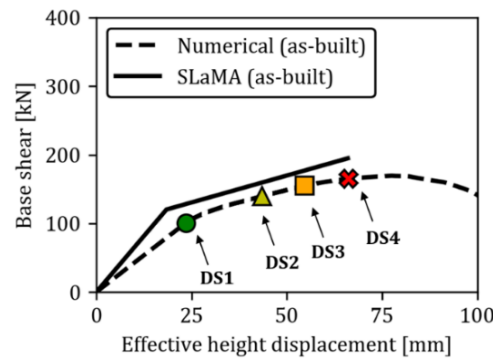
This produces N post-event structural variants reflecting plausible CNN mislabelling patterns.

6.4.4 Nonlinear Static Analyses: SLAMA vs. Refined Numerical Model

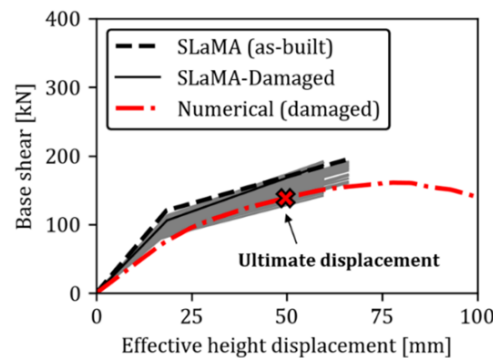
We evaluate global capacity for both intact and damaged configurations by (i) a simplified analytical/mechanical procedure (SLAMA) and (ii) a refined numerical pushover, enabling

cross-validation.

SLaMA method and its post-event extension. The SLaMA procedure is adopted owing to its demonstrated accuracy and speed. We use the post-earthquake extension proposed in [68], which improves the estimate of initial stiffness and incorporates damage-induced strength/ductility deterioration. For the intact configuration, SLaMA and the refined numerical pushover yield comparable capacity curves (Fig. 6.5(a), showing the expected low-ductility global behaviour consistent with non-ductile detailing. For the damaged configuration, we apply the λ -factors from Eq. (6.1) to the hinges of the affected members. For each of the $N = 10,000$ MC realizations, a SLaMA capacity curve is produced, yielding a *family* of post-event capacity curves (Fig. 6.5(b)). When benchmarked against the refined numerical pushover of the “true” damaged state, the SLaMA-based strength is in good agreement, whereas ultimate displacement tends to be overestimated, a known bias also noted in [83].



(a)



(b)

Figure 6.5. Results of the nonlinear static analyses for the structure in its (a) as-built and (b) damaged configuration. Image adapted from [83].

6.4.5 Spectrum-Based Safety Evaluation (IS-V/%NBS)

Seismic safety is quantified by the IS-V index prescribed by the Italian “Sismabonus” guidelines [14], which is equivalent to the %NBS measure used internationally [71]. Demand–capacity checks at the performance point are obtained via the Capacity Spectrum Method (CSM) [15]. For the intact state, both SLaMA and refined numerical capacity curves are assessed; for the post-event state, the CSM is repeated across all N MC realizations to propagate CNN-induced uncertainty to IS-V (Fig. 6.5(b).)

Representative results: Consistent with [83], the intact configuration yields an IS-V of approximately 0.61 with close agreement between SLaMA and the refined model (relative error $\sim 10\%$). In the damaged state, the IS-V drops to about 0.52 (numerical reference), evidencing a meaningful reduction of residual safety. The SLaMA-based IS-V for the damaged case shows a larger relative error (up to $\sim 15\%$), primarily due to the overestimation of ultimate displacement, yet remains sufficiently accurate for rapid appraisal. Notably, the dispersion due to CNN misclassification (visualised as $\pm 3\sigma$ bands across the N realizations) is modest: the uncertainty in labelling does *not* significantly affect the decision-relevant safety assessment, supporting the feasibility of the proposed image-to-capacity workflow for early-phase emergency evaluations [83].

6.4.6 Discussion

The integrated workflow captures key attributes of post-earthquake appraisal for non-ductile RC frames:

1. **Realistic damage patterns from NLTHA.** Concentration of drift demand in lower storeys emerges naturally under a recorded input (Irpinia 1980), generating the joint and column damage patterns typical of pre-code frames.
2. **Component-level updating via FEMA factors.** The λ -factor strategy (Eq. 6.1) yields physically interpretable reductions in both stiffness, strength and ductility at plastic hinges, consistent with FEMA 306 [23].
3. **Uncertainty-aware vision-to-capacity link.** Confusion-matrix–driven MC sampling provides a principled way to propagate CNN misclassification to global capacity and safety; in this application, the effect on IS-V is limited (§6.4.5).
4. **Two-track capacity evaluation.** SLaMA offers fast, sufficiently accurate estimates for both intact and damaged states; the refined pushover provides a validation target and highlights biases (e.g., SLaMA’s tendency to overestimate ultimate displacement in damaged configurations).
5. **Decision metrics.** IS-V/%NBS at the CSM performance point translates engineering results into codified, actionable safety classes for authorities and stakeholders [14, 71].

6.4.7 Limitations and Future Work

The present application deliberately focuses on a 2D interior frame under a single recorded accelerogram for clarity and traceability. Future extensions will consider: (i) 3D models

including torsion and plan irregularities; (ii) multi-record suites with hazard-compatible scaling; (iii) broader validation on buildings with documented post-event image datasets; and (iv) automated image-to-element mapping with geometry-aware registration to streamline the digital-to-analytical integration. The proposed uncertainty propagation can naturally accommodate improved CNNs, alternative confusion matrices, and Bayesian updating as field data accrue.

6.4.8 Implementation Notes (Practical Replicability)

- **Dynamic analysis.** Lumped-plasticity modelling in RUAUMOKO [7], excitation: SIMBAD Irpinia 1980, waveform ID 414, station BGI, $PGA_Y = 0.13g$ [103].
- **Damage states.** DS thresholds defined on fractions of ultimate plastic deformation (DS1: yield; DS2: 0.50; DS3: 0.75; DS4: ultimate).
- **CNN uncertainty.** Confusion matrix from the trained classifier (Fig. 6.4); MC size $N = 10,000$.
- **Capacity modifiers.** $\lambda - factors(\widehat{DS})$, per FEMA 306 [23].
- **Capacity evaluation.** SLaMA with post-event extension [68] and a refined numerical pushover.
- **Safety metric.** IS-V/%NBS via CSM per [15, 14, 71].

6.5 Integration of Damage Classification with Repair Cost Estimation

An important extension of post-earthquake assessment is estimating the repair costs associated with the damage, often as part of a loss assessment. Integrating damage classification data with cost estimation models can provide stakeholders with economic metrics (like monetary loss or repair time) in addition to structural safety metrics. In traditional loss estimation (e.g., FEMA P-58 methodology or HAZUS), damage states of components are mapped to repair costs and downtimes. The approach proposed in this chapter can enhance that process by supplying consistent, AI-derived damage states for all components, which then feed into a cost model.

The classified damage levels (from CNN or inspection) for each structural component can be linked to unit repair costs. For example, if a column is classified as heavy damage (cracked with spalling), the repair action might involve concrete removal and jacketing, which has a certain cost per column. Summing up the costs for all damaged components gives an estimate of the total repair cost for the building. Because the CNN-based survey is comprehensive, it ensures no damaged element is overlooked in the cost calculation, potentially improving accuracy. Moreover, the severity gradation (slight, moderate, heavy) can be tied to different repair scopes (e.g. epoxy injection for slight cracks vs. member replacement for severe damage), which have vastly different costs.

Prior studies have shown the importance of considering residual capacity in loss calculations. Ceccarelli et al. (2021) extended residual capacity assessment to evaluate economic losses for RC wall buildings [8, 77, 76]. They found that accounting for the reduced

capacity (and thus higher vulnerability) of a damaged structure leads to higher estimated losses. In their case, a multi-story building with wall damage saw increasing economic loss as the damage level worsened, and even low-level earthquakes could incur substantial losses if the structure's capacity was already degraded from prior events [80]. This implies that if a building is not repaired after an initial quake, its expected losses in the next event rise sharply. Rossi et al. (2022) [88] further examined how repair or non-repair alters a building's risk profile. By incorporating scenarios with and without repair, they quantified differences in expected annual loss: buildings that were repaired to restore capacity had lower future losses, whereas those left in a damaged state had a much higher loss potential over time. These insights emphasize that post-earthquake repair decisions have economic consequences, and thus cost estimation should be part of the assessment framework.

Integrating damage classification with cost estimation can be done as follows: once the CNN provides the breakdown of damage states per element, a repair cost algorithm assigns a cost to each element based on its type and damage state. For instance, a severely damaged column might be assigned the cost of installing a steel jacket or even replacement, a moderately damaged beam might be assigned the cost of fiber wrapping or epoxy injection, etc. Non-structural damages (if classified by the CNN, say falling plaster or cracked infills) can also be quantified for repair.

The sum of these single costs represents the total repair cost. Moreover, someone can also calculate a damage ratio (repair cost as a percentage of replacement cost of the building). In the event that this ratio is higher than the limit (commonly around 40-50%), the building can be seen as being without economical value for repair, thereby suggesting the option of demolishing and rebuilding instead.

In a building example from a case study, the cost model is assumed to show that the necessary works on 4 columns that are severely damaged and other minor repairs would total around 30% of the building's replacement value. This is a considerable cost, but it's very possible that the repairs would still make sense, particularly if the building has some importance. Though, if the destruction were more extensive (or the cost went over, e.g., 50% of replacement), then the stakeholders would be more likely to decide towards demolition. By these means, the residual capacity framework incorporating such cost measures becomes a full-fledged decision-support tool: the tool not only lets in details on whether the building is safe or not but also the probable cost to get it back to safety. Decision-makers (owners, governments) can thus balance safety and economic issues simultaneously.

Effective integration of such a feature requires reliable data on repair costs associated with different damage levels. This information can be derived from past earthquake repair records or construction cost databases. With time and more data collection (potentially even with AI damage estimation and repair tracking), cost prediction accuracy will increase. Eventually, connecting automated damage assessments with loss models could result in almost immediate estimates of economic impact after an earthquake. For instance, it is conceivable to have a city-wide system wherein AI processes drone images of building exteriors; within hours, the officials not only get a safety assessment of each building but also an order-of-magnitude repair cost for each, thereby deciding on the allocation of resources to areas in most need.

6.6 Conclusive Insights

This chapter introduced and discussed an integrated framework that combines automated vision-based damage classification with established structural engineering methods for residual seismic capacity assessment. The proposed approach directly addresses two of the most critical post-earthquake questions: “*Is the building still safe?*” and “*What will it take to repair or replace it?*”. By linking CNN-based damage detection to analytical model updating, the framework enables the derivation of quantitative indicators of residual capacity, going beyond qualitative rapid tagging.

The workflow can be summarized in the following steps:

1. **Damage detection and classification.** CNN models process photographic surveys (drone-based or manual) and automatically classify the type of component and its damage state (aligned with AeDES scales). This step reduces reliance on subjective field inspections.
2. **Mapping to mechanical parameters.** Classified damage states are translated into empirical reduction factors (λ_K , λ_Q , λ_D) for stiffness, strength, and ductility, following FEMA 306 [23] and subsequent proposals. This step operationalizes vision outputs into engineering inputs.
3. **Updating the structural model.** The analytical model of the as-built structure is updated to its post-earthquake configuration. Plastic hinge properties or strength coefficients of damaged members are modified according to the mapped reduction factors, while intact components retain their original properties.
4. **Residual capacity estimation.** Nonlinear static analyses (pushover curves) or equivalent procedures are conducted to quantify the building’s degraded capacity. Intact and damaged capacity curves can then be compared against seismic demand spectra using the Capacity Spectrum Method (CSM) [3].
5. **Decision-making support.** The outputs provide engineers and authorities with quantitative, uncertainty-aware measures of residual capacity. These results support triage decisions such as immediate re-occupancy, restricted use, urgent repair, or demolition.

From this workflow, several key insights emerge:

- *Automated classification is reaching maturity.* CNNs have demonstrated accuracy levels close to 90% in classifying damage severity from images, making them a reliable first step in large-scale post-disaster evaluations.
- *Damage-to-mechanics mapping is essential.* Translating qualitative labels into quantitative λ -factors for stiffness, strength, and ductility ensures that the structural model reflects the true residual state of the building.
- *Uncertainty must be explicitly addressed.* Confusion-matrix-based Monte Carlo sampling allows misclassification probabilities to be propagated through the model, resulting in families of capacity curves and safety indices rather than single deterministic values [83].

- *Analytical and numerical models are complementary.* Simplified approaches such as SLaMA [68] enable rapid estimates, while refined pushover analyses serve as benchmarks, highlighting conservative or non-conservative biases.
- *Decision metrics matter.* Quantifying safety through indices such as IS-V (%NBS) [14, 71] translates engineering results into actionable classifications for policymakers and stakeholders.

Looking ahead, the prospects of this integrated approach are promising. Future developments will likely focus on:

- improving CNN performance across different structural typologies and extending classification to non-structural components;
- coupling real-time image acquisition (via drones, UAVs, or street cameras) with digital-twin environments;
- validating the framework with experimental and field datasets, including comparisons between AI-assisted predictions and actual building performance under aftershocks;
- refining the mapping of vision outputs to mechanics, incorporating advanced degradation models for joints, bond-slip, and shear failures.

To sum up, the combination of AI-driven vision systems and structural mechanics can be a game-changer in the way post-earthquake assessment is performed. Operations can thus be expedited, not only emergency response but also the efficient allocation of resources, and strategies of resilience for the built environment can be supported. The proposed workflow, which unites rapid reconnaissance with detailed engineering evaluation, is seen as a definite step towards the future integration of robust digital twins for earthquake-prone cities and AI-augmented seismic risk management.

Conclusions and Future Efforts

Conclusions

This thesis has presented a vision-based framework for post-earthquake building damage assessment that combines the strengths of modern deep learning with established engineering analysis. The research was articulated around three main contributions:

1. **Creation of a new earthquake damage dataset:** a new curated dataset was developed by integrating multiple sources (PEER Φ -Net, INGV, ReLUIS, Turkey 2023 surveys, and scientific archives) and re-annotating all images according to the Italian *AeDES* guidelines. In addition, in various experiments, data augmentation and oversampling strategies were applied to overcome class imbalance, resulting in a balanced dataset of more than 19,000 images and overcoming many limitations. This dataset can provide a unique resource tailored to the Italian seismic engineering context and in the future could be a fundamental asset for training AI models in this domain.
2. **Benchmarking of CNN and Vision-Transformers architectures:** several CNNs were systematically evaluated, with particular focus on VGG16, because of its excellent performance in the testing phase. In fact, despite its simplicity compared to more recent architectures, the VGG16 demonstrated the best results when fine-tuned on the curated dataset, outperforming more complex networks in terms of stability and classification accuracy. These experiments confirmed that CNNs and ViTs are reliable tools for distinguishing damage levels (in our research comprising None, Slight, Moderate and Heavy classes), achieving accuracies up to 89-90%. Among the highlights of this work we noticed that architecture selection should be adapted to dataset size and domain constraints, and that VGG16 remains a strong baseline for damage assessment tasks in post-earthquake events.
3. **Integration with residual capacity assessment:** in the last chapter of this thesis, an outline of a conceptual workflow was presented, illustrating the application of convolutional neural networks (CNN) in classification, mapping the results to mechanical reduction factors, and then feeding them into structural models. This workflow, going around the pushover analysis and capacity curve update of the impaired structures, is thus a first step towards a feasible estimation of the residual seismic capacity. Methodologically, this stage is an innovation, signaling a possible contribution of computer vision-based deep learning tools in structural engineering tasks and further facilitating the delivery of outputs (safety indices and residual capacity estimates) of practical use.

Among other aspects, this thesis also showed the potential of data fusion based on the wavelet to complement RGB images, demonstrated the possibility to use AI-augmented workflows on example use cases, and highlighted the importance of digital data for ensure the reproducibility and upscaling of damage assessment.

Future Efforts

Although the results obtained are encouraging, several directions for further research remain open and constitute natural extensions of this thesis:

- **Semi-synthetic image generation:** A method of augmenting the training process with limited but well-controlled variations of damaged and undamaged structures by using a tunable semi-synthetic dataset could be a step further from the recent work [20]. This may not only solve the problem of the shortage of the real post-earthquake images but also facilitate the CNNs' ability to generalize.
- **YOLO-based damage detection:** The use of object detection models like YOLO can be a first step to locating the damaged parts (for instance, beams, columns, infilling walls) in UAV or drone images. Subsequently, CNNs can allocate these local patches into different damage levels, thereby a two-stage workflow (detection → classification) as work of [80] can be realized.
- **Expansion with IDEA dataset:** The inclusion of new datasets like IDEA will extend the range of building typologies, locations, and damage manifestations, which in turn will make the models more robust.
- **Component-specific neural networks:** Rather than a multi-class single classifier, a set of specialized CNNs could be created for different structural elements (for instance, a network only trained on columns). Such a division is expected to bring about the rise of sensitivity and decrease the number of confusions between adjacent classes, especially for those components that are considered as critical.
- **Integration of Large Language Models (LLMs):** The outputs of LLMs could be combined with those of CNNs to make it a fully automatic system that produces descriptive text reports, which are summaries of the damage type and severity in a human-readable format. The feature would be very helpful in the communication process between engineers, authorities, and other non-expert stakeholders.
- **Hybrid deep learning models:** The research in the future should explore the hybrid architectures which combine CNNs with attention mechanisms or Vision Transformers (ViTs). Multi-scale recognition methods can have a positive effect on the identification of faint and diverse damage patterns.
- **Domain adaptation and semi-supervised learning:** The differences between various areas and the need to label less data can be solved by the use of transfer learning and domain adaptation techniques. Moreover, the use of unlabeled images in semi-supervised methods would significantly increase the amount of data available for use.

- **Digital twin integration:** In the end, the suggested process could be the workflow of the digital twin of the city, where the seismic state of the building stock is updated continuously from the real-time image capture (by UAVs, street-level cameras, or satellites), thus enabling the predictive simulations and the proactive risk management.

In conclusion, this thesis has shown that the synergy between vision-based AI and structural engineering can lead to faster, more objective, and more scalable post-earthquake assessments. By focusing on CNNs, with particular emphasis on VGG16, and by introducing a new dataset specifically curated for earthquake damage classification, this work contributes both methodological innovations and practical tools. These advances represent a significant step toward AI-augmented seismic risk management and lay the foundation for future applications in resilience-oriented urban planning and civil protection operations.

Publications Related to This Thesis

1. S. Saquella et al. *Benchmarking of CNN Architectures for Post-Earthquake Damage Assessment*. In: 32nd Italian Workshop on Neural Networks WIRN 2024. Vietri sul Mare, Italy, June 2024.
2. S. Saquella et al. *Building damage level classification using deep learning: a CNN-based approach for post-earthquake structural assessment*. In: 10th International Conference COMPDYN 2025. Rhodes Island, Greece, 2025.
3. S. Saquella et al. *Post-earthquake Damage Assessment of Buildings Exploiting Data Fusion*. In: International Joint Conference on Neural Networks (IJCNN). Rome, Italy, 30 June–5 July 2025.
4. S. Pampanin, S.Saquella et al. *Towards an Integrated Interdisciplinary Approach for Large-Scale Modelling, Simulation and Management of Natural Disasters: A Virtual Case Study Urban Area*. In: COMPDYN 2025 – 10th ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering. Ed. by M. Papadrakakis and M. Fragiadakis. Rhodes Island, Greece, June 2025.
5. S. Saquella et al. *Post-earthquake Building Damage Level Classification using the Vision Transformer*. In: Proceedings of the 34th Italian Workshop on Neural Networks WIRN 2025. Smart Innovation, Systems and Technologies. Springer, 2025.
6. L. Pedone, S.Saquella et al. *Integrating Machine Learning Techniques into Post-Earthquake Residual Capacity Assessment of Reinforced Concrete Structures*. In: XX ANIDIS Conference. 2025.

Acknowledgements and Thanks

This research is supported by the PNRR (Piano Nazionale di Ripresa e Resilienza) funding scheme (PNRR MUR project CN 00000013 (B83C22002940006)/CN1) as part of the National Research Centre CN1 on “High-Performance Computing, Big Data and Quantum Computing” — Spoke 5: Environment & Natural Disasters [46].

I acknowledge the helpful support of the Sapienza team and my supervisor Prof. Giovanni Laneve. in particular I wish to thank Giulia Angelucci, Livio Pedone and Professor Michele Scarpiniti. Along with my wife, Claudia, and all the people that encouraged me.

Reference:

Spoke 5 — Environment & Natural Disasters, National Research Centre CN1, ICSC: description available at “Spoke 5 – Environment & Natural Disasters” page of the ICSC website. [46]

Bibliography

- [1] A. ALOISIO et al. *Machine-learning-aided regional post-seismic usability prediction of buildings: 2016-2017 Central Italy earthquakes*. In: Journal of Building Engineering 91 (2024), p. 109526. DOI: 10.1016/j.jobe.2024.109526.
- [2] G. ALTAMIRANO-GÒMEZ and C. GERSHENSON. *Quaternion Convolutional Neural Networks: Current Advances and Future Directions*. In: Advances in Applied Clifford Algebras 34.4 (2024), p. 42. DOI: 10.1007/s00006-024-01350-x.
- [3] ATC. *Seismic evaluation and retrofit of concrete buildings*. Tech. rep. ATC-40. Redwood City, CA, USA: Applied Technology Council, 1996.
- [4] M. AZIMI, A. ESLAMLOU, and G. PEKCAN. *Data-Driven Structural Health Monitoring and Damage Detection through Deep Learning: State-of-the-Art Review*. In: Sensors 20 (2020), p. 2778. DOI: 10.3390/s20102778.
- [5] C. BAGGIO et al. *Field manual for post-earthquake damage and safety assessment and short term countermeasures (AeDES)*. Tech. rep. EUR, 22868. European Commission—Joint Research Centre—Institute for the Protection and Security of the Citizen, 2007.
- [6] B. BORZI et al. *IRMA Platform for the Calculation of Damages and Risks of Italian Residential Buildings*. In: Bulletin of Earthquake Engineering (2020). DOI: 10.1007/s10518-020-00924-x.
- [7] A. J. CARR. *RUAUMOKO2D – the Māori god of volcanoes and earthquakes. Inelastic analysis finite element program*. In: Proceedings of [Conference/Workshop]. Christchurch, NZ, 2016.
- [8] C. CECCARELLI et al. *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.
- [9] Y.-J. CHA, W. CHOI, and O. BÜYÜKÖZTÜRK. *Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks*. In: Computer-Aided Civil and Infrastructure Engineering 32.5 (2017), pp. 361–378. DOI: 10.1111/mice.12263.
- [10] G. CHEN et al. *Case study of the 2023 Turkey–Syria doublet*. In: ResearchGate prepublication (2025).
- [11] C.-K. CHIU et al. *Experimental Quantification on the Residual Seismic Capacity of Damaged Reinforced Concrete Column Members*. In: International Journal of Concrete Structures and Materials 13.17 (2019). DOI: 10.1186/s40069-019-0338-z.

- [12] 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*. In: Earthquake Engineering & Structural Dynamics 50.3 (2021), pp. 756–776. DOI: 10.1002/eqe.3355.
- [13] F. CHOLLET. *Xception: Deep Learning with Depthwise Separable Convolutions*. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017, pp. 1251–1258. DOI: 10.1109/CVPR.2017.195.
- [14] E. COSENZA et al. *The Italian guidelines for seismic risk classification of constructions: technical principles and validation*. In: Bulletin of Earthquake Engineering 16.12 (2018), pp. 5905–5935.
- [15] A. T. COUNCIL. *Seismic Evaluation and Retrofit of Concrete Buildings*. Tech. rep. ATC-40. Redwood City, CA, USA: Applied Technology Council, 1996.
- [16] S. D’ASCOLI et al. *ConViT: Improving Vision Transformers with Soft Convolutional Inductive Biases*. In: Proceedings of the 38th International Conference on Machine Learning (ICML). 2021, pp. 2286–2296.
- [17] A. DEMIR et al. *Post-earthquake structural damage assessment, lessons learned, and addressing objections following the 2023 Kahramanmaraş, Turkey earthquakes*. In: Bulletin of Earthquake Engineering 23 (2025), pp. 1107–1127. DOI: 10.1007/s10518-024-02092-8.
- [18] M. DOLCE et al. *Observed Damage Database of Past Italian Earthquakes: the Da.D.O. Web-GIS*. In: Bollettino di Geofisica Teorica ed Applicata 60.2 (2019), pp. 141–164.
- [19] M. DOLCE et al. *Seismic Risk Assessment of Residential Buildings in Italy*. In: Bulletin of Earthquake Engineering (2020). DOI: 10.1007/s10518-020-01009-5.
- [20] P. DONDI et al. *Post-earthquake structural damage detection with tunable semi-synthetic image generation*. In: Engineering Applications of Artificial Intelligence 147 (2025), p. 110302. DOI: 10.1016/j.engappai.2025.110302.
- [21] A. DOSOVITSKIY et al. *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. In: International Conference on Learning Representations (ICLR 2021). 2021.
- [22] A. DOSOVITSKIY et al. *An image is worth 16x16 words: Transformers for image recognition at scale*. In: arXiv preprint arXiv:2010.11929 (2020).
- [23] FEMA. *Evaluation of earthquake damaged concrete and masonry wall buildings – Basic procedures manual*. Tech. rep. Federal Emergency Management Agency, Washington DC, US, 1998.
- [24] C. FERRARIS, G. AMPRIMO, and G. PETTITI. *Computer Vision and Image Processing in Structural Health Monitoring: Overview of Recent Applications*. In: Signals 4.3 (2023), pp. 539–574. DOI: 10.3390/signals4030029.
- [25] Y. GAO and K. M. MOSALAM. *Deep Transfer Learning for Image-Based Structural Damage Recognition*. In: Computer-Aided Civil and Infrastructure Engineering 33 (2018), pp. 748–768. DOI: 10.1111/mice.12363.

- [26] Y. GAO and K. M. MOSALAM. *PEER Hub ImageNet (Φ -Net): A Large-Scale Multi-Attribute Benchmark Dataset of Structural Images*. Tech. rep. 2019/07. Pacific Earthquake Engineering Research Center (PEER), University of California, Berkeley, 2019. DOI: 10.55461/QFKL9711.
- [27] Y. GAO and K. M. MOSALAM. *PEER Hub ImageNet: A Large-Scale Multiattribute Benchmark Data Set of Structural Images*. In: *Journal of Structural Engineering* 146.10 (2020), p. 04020198.
- [28] Y. GAO et al. *Multiattribute multitask transformer framework for vision-based structural health monitoring*. In: *Computer-Aided Civil and Infrastructure Engineering* 38.17 (2023), pp. 2358–2377. DOI: 10.1111/mice.13067.
- [29] Y. GAO and K. M. MOSALAM. *PEER Hub ImageNet (PHI-Net)*. 2020. URL: <https://apps.peer.berkeley.edu/phi-net/>.
- [30] G. GIUFFRIDA et al. *The Value of Multiple Earthquake Missions: the EEFIT L'Aquila Case*. In: *Bulletin of Earthquake Engineering* (2014).
- [31] I. GOODFELLOW et al. *Generative Adversarial Nets*. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2014, pp. 2672–2680.
- [32] I. GOODFELLOW, Y. BENGIO, and A. COURVILLE. *Deep Learning*. MIT Press, 2016. URL: <http://www.deeplearningbook.org>.
- [33] K. GOPALAKRISHNAN et al. *Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection*. In: *Construction and Building Materials* 157 (2017), pp. 322–330. DOI: 10.1016/j.conbuildmat.2017.09.110.
- [34] G. GRÜNTAL, ed. *European Macroseismic Scale 1998 (EMS-98)*. Vol. 15. Cahiers du Centre Européen de Géodynamique et de Séismologie. Luxembourg: Centre Européen de Géodynamique et de Séismologie, 1998, p. 99. DOI: 10.2312/EMS-98.full.en.
- [35] K. HE et al. *Deep Residual Learning for Image Recognition*. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.
- [36] K. HE et al. *Deep residual learning for image recognition*. In: *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 770–778.
- [37] G. E. HINTON and R. R. SALAKHUTDINOV. *Reducing the Dimensionality of Data with Neural Networks*. In: *Science* 313.5786 (2006), pp. 504–507. DOI: 10.1126/science.1127647.
- [38] G. HINTON, A. KRIZHEVSKY, and S. WANG. *Transforming Auto-Encoders*. In: *Proceedings of the International Conference on Artificial Neural Networks (ICANN)*. 2011, pp. 44–51. DOI: 10.1007/978-3-642-21735-7_6.
- [39] S. HOCHREITER and J. SCHMIDHUBER. *Long Short-Term Memory*. In: *Neural Computation* 9.8 (1997), pp. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.
- [40] Z. HONG et al. *Classification of Building Damage Using a Novel Convolutional Neural Network Based on Post-Disaster Aerial Images*. In: *Sensors* 22.15 (2022), p. 5920. DOI: 10.3390/s22155920.

- [41] V. HOSKERE et al. *Vision-based structural inspection using multiscale deep convolutional neural networks*. In: Proceedings of the 3rd Huixian International Forum on Earthquake Engineering for Young Researchers. Urbana-Champaign, IL, USA, 2018.
- [42] A. G. HOWARD et al. *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. In: arXiv preprint arXiv:1704.04861. 2017.
- [43] J. HOWARD and S. RUDER. *Universal Language Model Fine-tuning for Text Classification*. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne, Australia, 2018, pp. 328–339. DOI: 10.18653/v1/P18-1031.
- [44] J. HOWARD and S. RUDER. *Universal language model fine-tuning for text classification*. In: arXiv preprint arXiv:1801.06146 (2018).
- [45] G. HUANG et al. *Densely Connected Convolutional Networks*. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017, pp. 4700–4708. DOI: 10.1109/CVPR.2017.243.
- [46] F. ICSC. *Spoke 5 – Environment & Natural Disasters*. 2025. URL: <https://www.supercomputing-icsc.it/en/spoke-5-environment-natural-disasters-en/>.
- [47] A. ILKI et al. *The Modified Post-Earthquake Damage Assessment Methodology for TCIP (TCIP-DAM-2020)*. In: Advances in Assessment and Modeling of Earthquake Loss. Ed. by S. AKKAR et al. Springer Tracts in Civil Engineering. Springer, Cham, 2021, pp. 85–107. DOI: 10.1007/978-3-030-68813-4_5.
- [48] ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA (INGV). *DFM – Database Fotografico Macrosismico*. 2024. URL: <https://dfm.ingv.it/>.
- [49] J. JASIEWICZ and T. F. STEPINSKI. *Geomorphons — a pattern recognition approach to classification and mapping of landforms*. In: Geomorphology 182 (2013), pp. 147–156.
- [50] X. JI et al. *Vision-based seismic damage detection and residual capacity assessment for an RC shaking table test structure*. In: Earthquake Engineering & Structural Dynamics 52.3 (Mar. 2023), pp. 806–827. ISSN: 0098-8847. DOI: 10.1002/eqe.3788.
- [51] H. KIM et al. *Crack and noncrack classification from concrete surface images using machine learning*. In: Struct. Health Monit. 18 (2019), pp. 725–738.
- [52] T. N. KIPF and M. WELLING. *Semi-Supervised Classification with Graph Convolutional Networks*. In: International Conference on Learning Representations (ICLR). 2017.
- [53] A. KRIZHEVSKY, I. SUTSKEVER, and G. E. HINTON. *ImageNet classification with deep convolutional neural networks*. In: Communications of the ACM 60.6 (2012), pp. 84–90.
- [54] Y. LECUN et al. *Gradient-based learning applied to document recognition*. In: Proceedings of the IEEE 86.11 (1998), pp. 2278–2324. DOI: 10.1109/5.726791.

- [55] B. LI and K. M. MOSALAM. *Seismic Performance of Reinforced-Concrete Stairways During the 2008 Wenchuan Earthquake*. In: *Journal of Performance of Constructed Facilities* 27.6 (2012), pp. 721–730. DOI: 10.1061/(ASCE)CF.1943-5509.0000389.
- [56] X. LIANG. *Image-based post-disaster inspection of reinforced concrete bridge systems using deep learning with Bayesian optimization*. In: *Computer-Aided Civil and Infrastructure Engineering* 34 (2019), pp. 415–430. DOI: 10.1111/mice.12425.
- [57] Z.-L. LIU et al. *A crack detection and quantification method using matched filter and photograph reconstruction*. In: *Scientific Reports* 15 (2025), p. 25266. DOI: 10.1038/s41598-025-08280-z.
- [58] M. E. LOCCHI et al. *A large fault partially reactivated during two contiguous seismic sequences in Central Italy*. In: *Tectonophysics* 877 (2024), p. 230284.
- [59] K. LUO and L. IE-BIN. *Building a Vision Transformer-Based Damage Severity Classifier with Ground-Level Imagery of Homes Affected by California Wildfires*. In: *Fire* 7.4 (2024), p. 133. DOI: 10.3390/fire7040133.
- [60] 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: *COMPADYN Proceedings*. 2019, pp. 930–943.
- [61] M. MAEDA and D. E. KANG. *Post-Earthquake Damage Evaluation for Reinforced Concrete Buildings Based on Residual Seismic Capacity*. In: *13th World Conference on Earthquake Engineering*. Struct. volume. 2004.
- [62] M. MAGGIPINTO et al. *DeepVM: A Deep Learning-based Approach with Automatic Feature Extraction for 2D Input Data Virtual Metrology*. In: *Electronics* 8.3 (2019), p. 321. DOI: 10.3390/electronics8030321.
- [63] S. MALLAT. *A theory for multiresolution signal decomposition: The wavelet representation*. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11.7 (1989), pp. 674–693.
- [64] S. MALLAT. *A Wavelet Tour of Signal Processing: The Sparse Way*. 3rd. Academic Press, 2009.
- [65] K. MARDER et al. *Post-earthquake assessment of moderately damaged reinforced concrete plastic hinges*. In: *Earthquake Spectra* 36.1 (2020), pp. 299–321. DOI: 10.1177/8755293019878192.
- [66] J. MASCI et al. *Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction*. In: *Proceedings of the 21st International Conference on Artificial Neural Networks (ICANN)*. 2011, pp. 52–59. DOI: 10.1007/978-3-642-21735-7_7.
- [67] M. MATTEONI et al. *Multi-scale and multi-refinement framework for seismic risk assessment of urban areas*. In: *Procedia Structural Integrity* 64 (2024), pp. 2005–2012.
- [68] M. MATTEONI et al. *Simplified Analytical/Mechanical Procedure for the Residual Capacity Assessment of Earthquake-Damaged Reinforced Concrete Frames*. In: *9th International Conference COMPADYN 2023*. Athens, Greece, 2023.

- [69] Z. MIAO et al. *Deep learning-based evaluation for mechanical property degradation of seismically damaged RC columns*. In: Earthquake Engineering & Structural Dynamics 52.8 (July 2023), pp. 2498–2519. ISSN: 0098-8847. DOI: 10.1002/eqe.3749.
- [70] K. M. MOSALAM, S. M. TAKHIROV, and S.-H. PARK. *Applications of Laser Scanning to Structures in Laboratory Tests and Field Surveys*. In: Structural Control and Health Monitoring 21.1 (2014), pp. 115–134.
- [71] NEW ZEALAND SOCIETY FOR EARTHQUAKE ENGINEERING (NZSEE), STRUCTURAL ENGINEERING SOCIETY OF NEW ZEALAND (SESOC), and NEW ZEALAND GEOTECHNICAL SOCIETY (NZGS). *The Seismic Assessment of Existing Buildings: Technical Guidelines for Engineering Assessments (“Red Book” Version 1)*. Tech. rep. Wellington, New Zealand: Ministry of Business, Innovation, Employment (MBIE), and Earthquake Commission (EQC), 2017. URL: https://design.resilience.nz/assets/Documents/Overview_of_the_Seismic_Assessment_Guidelines_20170703.pdf.
- [72] P. D. OGUNJINMI et al. *Rapid Post-Earthquake Structural Damage Assessment Using Convolutional Neural Networks and Transfer Learning*. In: Sensors 22 (9 2022), p. 3471. DOI: 10.3390/s22093471.
- [73] A. L. C. OTTONI et al. *Tuning of Data Augmentation Hyperparameters in Deep Learning to Building Construction Image Classification with Small Datasets*. In: International Journal of Machine Learning and Cybernetics 14.1 (2023), pp. 171–186. DOI: 10.1007/s13042-022-01555-1.
- [74] Ç. F. ÖZGENEL and A. GÖNENÇ SORGUÇ. *Concrete Crack Images for Classification*. Version v2. 2019. DOI: 10.17632/5y9wdsg2zt.2. URL: <https://data.mendeley.com/datasets/5y9wdsg2zt/2>.
- [75] 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. Ed. by S. AKKAR et al. Springer Tracts in Civil Engineering. Springer International Publishing, 2021, pp. 3–25. DOI: https://doi.org/10.1007/978-3-030-68813-4_1.
- [76] S. PAMPANIN. *Towards the Practical Implementation of Performance-Based Assessment and Retrofit Strategies for RC Buildings: Challenges and Solutions*. In: 4th International Conference on Smart Monitoring, Assessment and Rehabilitation of Civil Structures (SMAR 2017). Zurich, Switzerland, 2017.
- [77] S. PAMPANIN et al. *Residual Capacity and Repairing Options for Reinforced Concrete Buildings*. Tech. rep. Contract 2012-UOC-02-NHRP. Natural Hazard Research Platform, 2015.
- [78] S. PAMPANIN et al. *Towards an Integrated Interdisciplinary Approach for Large-Scale Modelling, Simulation and Management of Natural Disasters: A Virtual Case Study Urban Area*. In: COMPDYN 2025 - 10th ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering. Ed. by M. PAPADRAKAKIS and M. FRAGIADAKIS. Rhodes Island, Greece, June 2025.
- [79] S. J. PAN and Q. YANG. *A Survey on Transfer Learning*. In: IEEE Transactions on Knowledge and Data Engineering 22.10 (2010), pp. 1345–1359. DOI: 10.1109/TKDE.2009.191.

- [80] X. PAN and T. YANG. *Postdisaster image-based damage detection and repair cost estimation of reinforced concrete buildings using dual convolutional neural networks*. In: Computer-Aided Civil and Infrastructure Engineering 35 (2020), pp. 495–510. DOI: 10.1111/mice.12549.
- [81] P. PARIDA and N. BHOI. *Wavelet based transition region extraction for image segmentation*. In: Future Computing and Informatics Journal 2.2 (2017), pp. 65–78. DOI: 10.1016/j.fcij.2017.10.005.
- [82] L. PEDONE et al. *Energy-based procedures for seismic fragility analysis of mainshock-damaged buildings*. In: Frontiers in Built Environment 9 (2023), p. 1183699.
- [83] L. PEDONE et al. *Integrating Machine Learning Techniques into Post-Earthquake Residual Capacity Assessment of Reinforced Concrete Structures*. In: XX ANIDIS Conference. 2025.
- [84] E. PINTELAS, I. E. LIVIERIS, and P. E. PINTELAS. *A Convolutional Autoencoder Topology for Classification in High-Dimensional Noisy Image Datasets*. In: Sensors 21.22 (2021), p. 7731. DOI: 10.3390/s21227731.
- [85] M. POLESE et al. *Residual capacity of earthquake damaged buildings*. In: Proceedings of the 15th World Conference on Earthquake Engineering. Lisboa, Portugal, 2012.
- [86] S. PUJOL, I. BEDIRHANOGLU, C. DÖNMEZ, et al. *Quantitative evaluation of the damage to RC buildings caused by the 2023 southeast Turkey earthquake sequence*. In: Earthquake Spectra 40.1 (2024), pp. 505–530. DOI: 10.1177/87552930231211208.
- [87] A. RADFORD, L. METZ, and S. CHINTALA. *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*. In: International Conference on Learning Representations (ICLR). 2016.
- [88] A. ROSSI, C. DEL VECCHIO, and S. PAMPANIN. *Influence of earthquake damage and repair interventions on expected annual losses of reinforced concrete wall buildings*. In: Earthquake Spectra 38.3 (2022), pp. 2026–2060. DOI: 10.1177/87552930211072878.
- [89] D. E. RUMELHART, G. E. HINTON, and R. J. WILLIAMS. *Learning representations by back-propagating errors*. In: Nature 323.6088 (1986), pp. 533–536. DOI: 10.1038/323533a0.
- [90] S. SABOUR, N. FROSST, and G. E. HINTON. *Dynamic Routing Between Capsules*. In: Advances in Neural Information Processing Systems (NeurIPS). 2017, pp. 3856–3866.
- [91] M. SANDLER et al. *MobileNetV2: Inverted Residuals and Linear Bottlenecks*. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2018). Los Alamitos, CA, USA, 2018, pp. 4510–4520. DOI: 10.1109/CVPR.2018.00474.
- [92] M. SANDLER et al. *MobileNetV2: Inverted residuals and linear bottlenecks*. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, pp. 4510–4520.
- [93] S. SAQUELLA et al. *Benchmarking of CNN Architectures for Post-Earthquake Damage Assessment*. In: 32nd Italian Workshop on Neural Networks (WIRN 2024). Vietri sul Mare, Italy, June 2024.

- [94] S. SAQUELLA et al. *Building damage level classification using deep learning: a CNN-based approach for post-earthquake structural assessment*. In: 10th International Conference COMPDYN 2025. Rhodes Island, Greece, 2025.
- [95] S. SAQUELLA et al. *Post-earthquake Building Damage Level Classification using the Vision Transformer*. In: Proceedings of the 34th Italian Workshop on Neural Networks (WIRN 2025). Smart Innovation, Systems and Technologies. Springer, 2025.
- [96] S. SAQUELLA et al. *Post-earthquake Damage Assessment of Buildings Exploiting Data Fusion*. In: International Joint Conference on Neural Networks (IJCNN). Rome, Italy, 30 June–5 July 2025, 2025.
- [97] F. SCARSELLI et al. *The Graph Neural Network Model*. In: IEEE Transactions on Neural Networks 20.1 (2009), pp. 61–80. DOI: 10.1109/TNN.2008.2005605.
- [98] I. SENALDI et al. *IDEA: Image database for earthquake damage annotation*. In: Data in Brief 61 (2025), p. 111733. ISSN: 2352-3409. DOI: <https://doi.org/10.1016/j.dib.2025.111733>. URL: <https://www.sciencedirect.com/science/article/pii/S2352340925004603>.
- [99] H. SEZEN et al. *Performance of Reinforced Concrete Buildings During the August 17, 1999 Kocaeli, Turkey Earthquake, and Seismic Design and Construction Practice in Turkey*. In: Engineering Structures 25.1 (2003), pp. 103–114. DOI: 10.1016/S0141-0296(02)00121-9.
- [100] A. SHAMILA EBENEZER et al. *Identification of Civil Infrastructure Damage Using Ensemble Transfer Learning Model*. In: Advances in Civil Engineering 2021.1 (2021), p. 5589688. DOI: 10.1155/2021/5589688.
- [101] K. SIMONYAN and A. ZISSERMAN. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. In: International Conference on Learning Representations. 2015.
- [102] K. SIMONYAN and A. ZISSERMAN. *Very deep convolutional networks for large-scale image recognition*. In: arXiv preprint arXiv:1409.1556 (2014).
- [103] C. SMERZINI et al. *Ground motion record selection based on broadband spectral compatibility*. In: Earthquake Spectra 30.4 (2014), pp. 1427–1448.
- [104] C. SZEGEDY et al. *Going Deeper with Convolutions*. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015, pp. 1–9. DOI: 10.1109/CVPR.2015.7298594.
- [105] C. SZEGEDY et al. *Going deeper with convolutions*. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015, pp. 1–9.
- [106] M. TAN and Q. LE. *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. In: Proceedings of the 36th International Conference on Machine Learning. Long Beach, 2019, pp. 6105–6114.
- [107] M. TAN and Q. LE. *EfficientNet: Rethinking model scaling for convolutional neural networks*. In: International Conference on Machine Learning (2019).
- [108] TIME MAGAZINE. *Thousands of Buildings Collapsed in Turkey. Devastation Was Preventable, Experts Say*. In: Time (2023). URL: <https://time.com/6254024/turkey-syria-earthquake-building-2023/>.

- [109] E. TINTI et al. *A kinematic source-time function compatible with earthquake dynamics*. In: Bulletin of the Seismological Society of America 95.4 (2005), pp. 1211–1223.
- [110] H. TOUVRON et al. *Training data-efficient image transformers & distillation through attention*. In: Proceedings of the 38th International Conference on Machine Learning. Ed. by M. MEILA and T. ZHANG. Vol. 139. Proceedings of Machine Learning Research. PMLR, 2021, pp. 10347–10357.
- [111] N. WANG et al. *Damage Classification for Masonry Historic Structures Using Convolutional Neural Networks Based on Still Images*. In: Computer-Aided Civil and Infrastructure Engineering 33 (2018), pp. 1073–1089. DOI: 10.1111/mice.12411.
- [112] Y. XIE. *Deep Learning in Earthquake Engineering: A Comprehensive Review*. In: arXiv preprint arXiv:2405.09021 (2024).
- [113] H. YAMADA et al. *Evaluation of Building Damage due to Natural Disaster using CNN and GAN*. In: Structural Health Monitoring- The 9th Asia-Pacific Workshop on Structural Health Monitoring, 9APWSHM 2022. Association of American Publishers. 2023, pp. 67–75. DOI: 10.21741/9781644902455-9.
- [114] Q. ZHANG et al. *Feature extraction of face image based on LBP and 2-D Gabor wavelet transform*. In: Mathematical Biosciences and Engineering 17.2 (2020), pp. 1578–1592. DOI: 10.3934/mbe.2020082.
- [115] B. ZOPH et al. *Learning Transferable Architectures for Scalable Image Recognition*. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018, pp. 8697–8710. DOI: 10.1109/CVPR.2018.00907.