E-26
Kongres 2026
Congress 2026

doi: 10.67563/ASESCONF2026064M

Marko Marinković[1], Zoran Stojadinović[2], Miloš Kovačević[2], Mladen Nikolić[3], Dejan Marinković[2], Đorđe Nedeljković[2], Zoran Pucanović[2], Filip Đorđević[2], Marija Ivanović[2], Nevena Simić[2]

DEVELOPMENT OF FRAMEWORK FOR ASSESSMENT OF LOSSES AFTER EARTHQUAKE AS A FUNCTION OF QUICK RECOVERY – RELAR PROJECT

Summary: The paper presents the RELAR project, with the aim of improving the assessment of losses and the recovery process after an earthquake. The project is based on the application of machine learning methods and image recognition techniques to speed up the damage assessment process and increase the accuracy of remediation cost estimates. The limitations of traditional approaches are pointed out, which are often time-consuming and prone to errors due to the lack of data and their rigidity. RELAR introduces innovative solutions that enable fast and reliable assessments, independent of the availability of ground acceleration records. The ultimate goal of the project is the development of algorithms, models and guidelines for risk reduction and faster recovery.

Keywords: seismic risk, damage assessment, repair costs, recovery

RAZVOJ OKVIRA ZA PROCENU GUBITAKA NAKON ZEMLJOTRESA U FUNKCIJI BRŽEG OPORAVKA – PROJEKAT RELAR

Rezime: U radu je prikazan projekat RELAR, sa ciljem unapređenja procene gubitaka i procesa oporavka nakon zemljotresa. Projekat se zasniva na primeni metoda mašinskog učenja i tehnika prepoznavanja slika radi ubrzanja postupka procene oštećenja i povećanja tačnosti procene troškova sanacije. Ukazuje se na ograničenja tradicionalnih pristupa, koji su često vremenski zahtevni i skloni greškama usled nedostatka podataka i njihove rigidnosti. RELAR uvodi rešenja koja omogućavaju brze i pouzdane procene, nezavisno od dostupnosti zapisa o ubrzanju tla. Krajnji cilj projekta je razvoj algoritama, modela i smernica za smanjenje rizika i brži oporavak.

Ključne reči: seizmički rizik, procena štete, troškovi popravke, oporavak

[1] University of Belgrade, Faculty of Civil Engineering, Belgrade, Serbia, mmarinkovic@grf.bg.ac.rs
[2] University of Belgrade, Faculty of Civil Engineering, Belgrade, Serbia
[3] University of Belgrade, Faculty of Mathematics, Belgrade, Serbia

1. INTRODUCTION

After a damaging earthquake, the first engineering question is usually simple to state and difficult to answer: where is the damage, how serious is it, and what should be done next? In a large urban area, conventional building-by-building inspection remains indispensable, but it is also slow, demanding and difficult to scale. The shortage of qualified inspectors is most pronounced precisely when decisions have to be made quickly, under repeated aftershocks and incomplete information.

The RELAR project develops a framework in which field observations, image analysis, repair-cost estimation and recovery planning are treated as parts of one decision chain. The intention is not to replace the engineer. It is to reduce the amount of repetitive screening work, organise the incoming evidence and provide a transparent basis for prioritising detailed inspections. This distinction is important. A model may identify a suspicious image region, but the final decision on structural safety still requires engineering judgement and, where necessary, an on-site investigation.

Rapid post-earthquake building assessment serves as a critical first line of defense in emergency management, focusing primarily on immediate life-safety and emergency occupancy decisions. The fundamental objective of this early phase is to provide real-time support for emergency response, enforce access control in affected zones, and establish clear priorities for follow-up actions. It is intentionally not designed to define the exact technical scope or execution strategies of subsequent repair works. This initial safety evaluation relies heavily on visual inspections and standardized, qualitative classification protocols. A prominent operational baseline is ATC-20 (Procedures for Post-Earthquake Safety Evaluation of Buildings), which outlines frameworks for rapid and detailed engineering inspections to assign structures into distinct safety categories: safe, restricted use, or unsafe [1]. These procedures are explicitly optimized for large-scale, rapid deployment under stressful emergency conditions, where conservative, safety-first judgments systematically take precedence over detailed damage quantification.

In a similar vein, related screening approaches like FEMA P-154 (Rapid Visual Screening of Buildings for Potential Seismic Hazards) introduce structured checklists and score-based systems to identify buildings with latent seismic vulnerabilities based on readily observable structural features. Although originally designed for pre-disaster screening, these rapid visual screening (RVS) concepts are frequently adapted for post-earthquake environments due to their logistical speed and procedural simplicity. In recent years, academic research has increasingly focused on leveraging computer-vision algorithms and machine-learning tools to augment these rapid safety assessments. For instance, image-based classifiers are actively investigated for the automated assignment of damage states (DS) to streamline high-level emergency decision-making. However, these data-driven methodologies remain closely aligned with the objectives of safety tagging and broad damage recognition, rather than practical repair definition. Consequently, rapid safety assessment represents a vital but inherently restricted decision layer; it successfully determines immediate building accessibility, while complex questions regarding exact remediation methods, resource allocation, and reconstruction scheduling are deferred to subsequent assessment layers operating at much finer temporal and informational scales.

In contrast, probabilistic loss estimation frameworks and portfolio-scale analyses address an entirely different decision-making domain. Their primary goal is to quantify the aggregated economic consequences of seismic damage across individual structures, commercial portfolios, or entire urban regions, thereby guiding long-term strategic planning, risk financing, and public policy. Well-established methodologies rooted in Performance-Based Earthquake Engineering (PBEE), such as the FEMA P-58 framework and its accompanying Performance Assessment Calculation Tool (PACT), formalize this analytical layer [2, 3]. They combine probabilistic seismic demand models, component-level fragility curves, and consequence functions to estimate expected repair costs, structural downtime, and secondary financial losses [2]. Within this framework, uncertainties across ground motion characteristics, structural responses, and physical damage states are explicitly propagated to yield probabilistic metrics like loss-exceedance curves [2].

At the portfolio scale, these models rely heavily on data abstraction and aggregation to remain scalable [3, 4]. While highly effective for regional screenings and policy-level prioritization, recent developments also explore post-event cost estimation frameworks that incorporate multiple influencing factors to improve repair cost prediction [5]. However, this abstraction inevitably masks critical spatial variations at the building level. Recent probabilistic and data-driven approaches have demonstrated that repair cost can be predicted with reasonable accuracy at the damage-state level using cost matrices [6]. In parallel, advances in satellite imagery [7], LiDAR [8], and machine-learning-based classification [9] have enabled rapid, large-scale estimation of damage severity. Yet, categorical damage states fail to capture the actual spatial distribution of damage within a single structure, and classification is routinely governed by the single most severe defect observed, even if confined to an isolated area. Consequently, a building classified into a severe damage state due to localized failures might require less total repair volume than a building experiencing widespread, low-level damage. This fundamental distinction is lost within a single severity label, demonstrating that regional-scale loss models cannot seamlessly translate into executable, site-specific construction instructions where decisions depend entirely on the precise layout, volume, and practical constructability of repairs. This limitation was also observed in prior work on damage-state-based repair-cost prediction, where accurate aggregate estimates were achieved despite substantial variability arising from damage extent [6].

Extensions of PBEE have incorporated post-earthquake functional recovery and resilience, linking damage states to downtime and restoration processes [10, 11]. These approaches enhance understanding of how individual buildings contribute to broader recovery, but they remain grounded in probabilistic performance metrics rather than executable repair definitions.

This paper focuses on the connection between two components of the framework: image-based damage recognition and observation-based quantification of repair works. The first component locates visible damage in photographs. The second translates measurable damage attributes into repair methods, quantities and costs. Their integration avoids a common discontinuity in loss models, where a building is first assigned to a broad damage state and then linked to an average repair ratio that may have little connection with the actual observed defects.

2. RELAR ASSESSMENT CONCEPT

The proposed workflow starts with photographs collected during a rapid field survey. Each image should be traceable to a building, storey, room and wall or structural element. The photograph is divided into overlapping patches, which are classified as damaged or undamaged. Because the position of every patch within the original image is known, the output is not only a class label but also an approximate spatial distribution of visible damage.

The classified regions are then combined with information on the affected component and the geometry of the defect. Engineering rules select an appropriate repair method and calculate a component-level bill of quantities. Local prices for labour, materials and equipment are subsequently applied to obtain repair costs. Building-level results can be aggregated to a portfolio and used as input for recovery planning. Figure 1 shows RELAR concept.

The procedure is deliberately modular. The damage classifier may be improved or replaced without changing the repair logic, while local prices can be updated without retraining the image model. This separation also makes it easier to identify uncertainty. Errors originating from poor image quality, classification, geometric scaling and engineering assumptions can be examined separately rather than hidden within a single loss ratio.

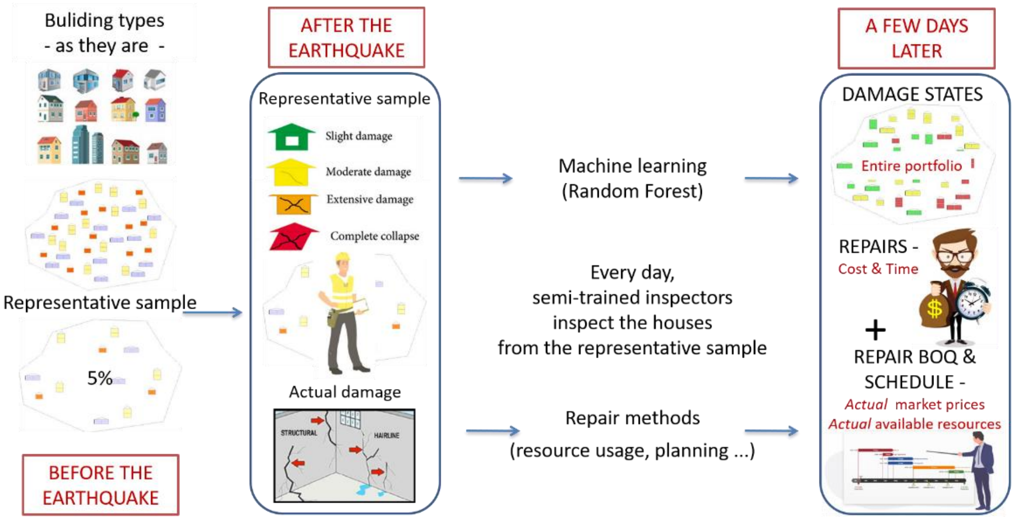

Figure 1: RELAR concept

3. IMAGE DATA AND ANNOTATION

3.1. PETRINJA EARTHQUAKE IMAGE DATABASE

The image database used in the present stage of the work originates from a post-earthquake reconnaissance mission following the 29 December 2020 Mw 6.4 Petrinja earthquake in Croatia. The photographs were captured during the SUZI-SAEE (Serbian Association for Earthquake Engineering) reconnaissance mission and were provided by one of the co-authors of this paper, Marko Marinković. The field team documented damaged and apparently undamaged buildings in the epicentral region, including both exterior and interior views. The source photographs are high-resolution images, typically between 3000 and 4000 pixels on the shorter side.

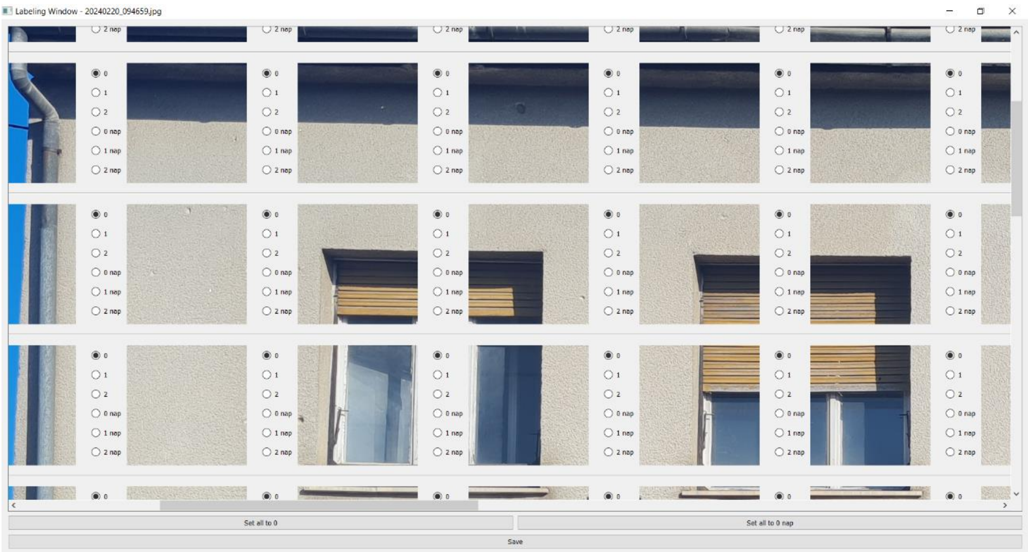

Figure 2: Photo patches

For model development, each photograph was divided into 224 × 224 pixel patches using a sliding window with 10% overlap (Figure 2). The overlap was introduced to reduce the probability that a crack or a local failure would fall exactly on the boundary between two patches. The resulting dataset contains approximately 15,000 patches, with a broadly balanced representation of damaged and undamaged image regions.

3.2. LABELLING PROCEDURE

A dedicated labelling application was developed to keep the original image and the patch grid connected. The main window allows the annotator to move through the photograph set and check the labelling status, while the secondary view presents the automatically generated patches and the available labels. This arrangement proved useful because an isolated patch can be ambiguous: a line that resembles a crack may in fact be a joint, cable, shadow or boundary between finishes. Seeing the parent photograph provides the context needed for a defensible label.

Four annotators independently reviewed the patches: two structural engineers and two engineers from mathematical and technical disciplines. Each patch was labelled as damaged or undamaged. Final labels were obtained by majority voting, while tied cases were discussed. Patches associated with low confidence were set aside rather than forced into the binary dataset. This conservative choice reduced the amount of noisy training data and preserved difficult examples for later development of multi-level damage classes.

Four independent reviewers evaluated each patch to classify it as damaged or undamaged. The panel consisted of two structural engineers (serving as experts) alongside two data and mathematical science engineers (acting as non-experts). The team was directed to assign a damaged label if a patch displayed obvious structural failure or if a combination of several small defects indicated major deterioration. To measure the level of consensus between pairs of reviewers, we utilized Cohen's kappa coefficient, calculated as:

$$\kappa = \frac{P_0 - P_e}{1 - P_e} \qquad (1)$$

Here, P_0 represents the actual observed agreement rate between two reviewers, and P_e represents the probability of agreement occurring purely by chance [12]. Based on the framework by [13], κ values below 0 signify agreement worse than chance, 0.00 to 0.20 indicate slight agreement, 0.21 to 0.40 show fair agreement, 0.41 to 0.60 represent moderate agreement, 0.61 to 0.80 reflect substantial agreement, and 0.81 to 1.00 denote nearly perfect agreement.

Table 1 outlines the specific consensus rates achieved across our dataset. The kappa metric between the two domain experts reached 0.800, which marks the highest point of the substantial agreement bracket. Meanwhile, the agreement rate between Expert 1 and Non-expert 1 hit 0.733 (substantial agreement), while Expert 1 and Non-expert 2 shared a lower score of 0.562 (moderate agreement). The majority of reviewer discrepancies occurred on patches featuring minor superficial issues, including hairline fractures or tiny patches of peeling plaster.

Table 1: Inter-annotator (experts vs non-experts) agreement using Cohen's κ

–	Expert 2	Non-expert 1	Non-expert 2
Expert 1	0.800	0.733	0.562

4. DAMAGE CLASSIFICATION MODELS

4.1. CONVOLUTIONAL NEURAL NETWORKS

Two convolutional neural networks were considered. ResNet-18 was selected as a well-established baseline, while ConvNeXt-Tiny represented a more recent convolutional architecture with design choices influenced by vision transformers. Both networks were initialised with ImageNet-pretrained weights and fine-tuned on the Petrinja patch dataset.

The source photographs were divided at image level into training, validation and test groups in an 80/10/10 proportion. Keeping all patches from one photograph in the same group was essential. A random patch-level split would place visually almost identical neighbouring patches in both training and test sets and would therefore give an unrealistically favourable estimate of performance. Training used binary cross-entropy loss, stochastic gradient descent with momentum, a batch size of 64 and early stopping based on validation loss.

4.2. ZERO-SHOT VISION-LANGUAGE MODEL

A large vision-language model was also examined in a zero-shot setting. Gemma 3 27B received an image patch together with a short prompt asking for a description of visible damage. The response was converted into a binary decision by parsing terms associated with cracking, spalling, damage or an intact surface. No project-specific fine-tuning was applied. The experiment was therefore intended to test whether a general multimodal model can transfer its visual knowledge to post-earthquake imagery without a dedicated training set.

4.3. RESULTS AND ERROR PATTERNS

ConvNeXt-Tiny gave the best overall result, with an F1-score of approximately 0.85 for the damaged class and an accuracy close to 87%. ResNet-18 achieved an F1-score of about 0.83 and an accuracy near 85%. For ConvNeXt-Tiny, precision and recall were approximately 0.88 and 0.82, respectively. The zero-shot Gemma model achieved an F1-score of approximately 0.79. Its recall was relatively high, but this came with a larger number of false positive detections.

Table 2: Comparative performance of the evaluated image models

Model	Training mode	F1-score	Main observation
ResNet-18	Fine-tuned	≈ 0.83	Stable baseline; subtle cracks may be missed
ConvNeXt-Tiny	Fine-tuned	≈ 0.85	Best balance of precision and recall
Gemma 3 27B	Zero-shot	≈ 0.79	High sensitivity, but more false alarms

The inspection of misclassified patches was as informative as the aggregate scores. Fine hairline cracks and slight spalling were common false negatives for the convolutional models. Dirt, peeling paint and textured backgrounds generated false positives. The vision-language model was particularly sensitive to debris, foliage and irregular patterns that it associated with damaged scenes. These errors show why model output should be treated as screening information rather than as a direct safety classification.

5. FIELD IMAGE ACQUISITION REQUIREMENTS

The quality of a damage model cannot compensate for a photograph that lacks geometric context. For this reason, the RELAR workflow includes a practical acquisition protocol. Each image should be assigned, whenever possible, to a project or building identifier, floor, room, wall or element and image type (Figure 3). A wall overview is preferred to an isolated close-up because the overview preserves the relationship between the defect and the component.

For wall damage, at least one full-height vertical edge should be visible from floor to ceiling (Figure 4). This edge acts as the principal geometric reference for later scaling and image rectification. A near-frontal image is preferable, but an oblique view is acceptable in narrow rooms or for long walls, provided that the reference edge and the complete damage region remain visible. Close-up photographs are useful for interpreting crack width or local crushing, but they should supplement rather than replace the overview image.

Metadata / naming convention example

Field	Example
Project ID	RELA-023
Building ID	B01
Floor	02
Room	R205
Wall ID	W03
Image type	Wall overview
Filename	RELA023_B01_F02_R205_W03_overview_01.jpg

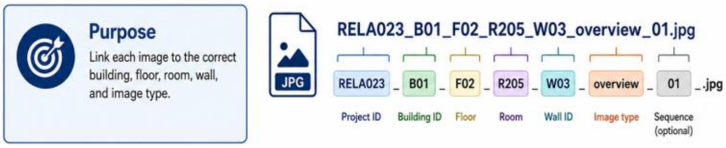

Figure 3: Spatial referencing and a recommended image filename structure

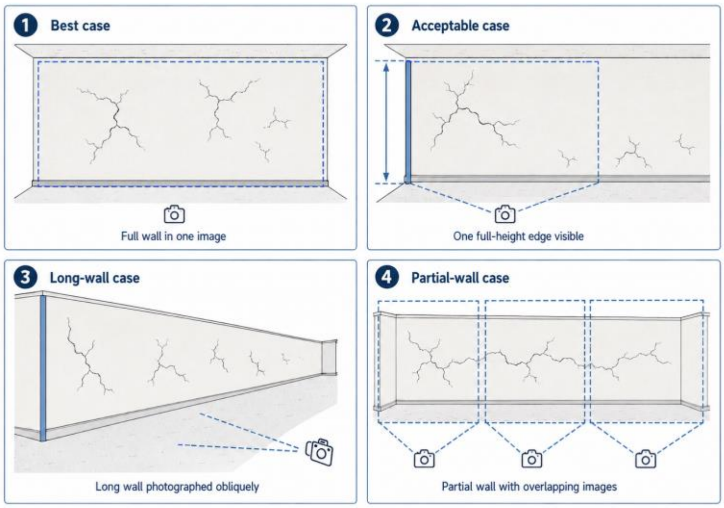

Figure 4: Preferred framing options for wall photographs

Obstructions must also be recorded. Furniture, installations, debris or finishes can hide part of the wall, and a hidden area should never be interpreted automatically as undamaged. The field record should therefore include the best available overview, additional angled images and a note describing the limited visibility. These seemingly minor instructions are important because they determine whether the photographs can later support metric crack-length estimation and repair quantity calculation.

6. FROM OBSERVED DAMAGE TO REPAIR COST

6.1. LIMITATIONS OF FIXED DAMAGE-STATE MATRICES

The original RELAR concept envisaged a repair-cost matrix indexed by building typology, damage state and damage extent. Such matrices are useful for regional scenarios, but they become difficult to apply when the objective is to reproduce the actual repair works in a particular building. Two walls assigned to the same nominal damage state may require different interventions because of crack width, depth, distribution, material, access or finishing requirements.

A fixed matrix also hides the engineering reasoning behind the cost. The analyst retrieves a value from a cell, but the connection between the visible defect and the assumed work items is not always clear. Calibration can reduce this problem, although a large number of typologies and conditional cases quickly makes the matrix cumbersome and difficult to update.

6.2. OBSERVATION-DRIVEN DECISION LOGIC

The implemented approach therefore starts from observable damage attributes. The affected component, crack geometry, local material degradation and relevant constructability constraints are evaluated through a decision tree. Each branch represents an explicit engineering condition, such as a crack-width or crack-depth threshold, while each terminal node identifies a repair method and its associated work items.

The output is a component-level bill of quantities expressed in physical units. Labour, materials and equipment are then monetised using local unit prices and resource-consumption rates. The cost of a component is obtained as the sum of its work items, and the building cost is obtained by aggregation. Costs may also be normalised by affected area or gross floor area for comparison across buildings and for portfolio-level analysis.

This formulation does not remove uncertainty, but it makes it visible. A user can trace the calculated cost back to the observed damage, selected repair method, measured quantity and unit price. The logic can also be revised when local construction practice changes. In this sense, the decision tree can be understood as a generalisation of a repair matrix: each path through the tree corresponds to a specific combination of damage conditions and repair actions, but the conditional dependencies remain explicit.

7. DISCUSSION

The current results demonstrate that image-based screening can provide useful information for rapid post-earthquake assessment, especially when the number of incoming photographs exceeds the capacity of available engineers. The best-performing convolutional model correctly distinguished damaged and undamaged patches in most cases, while the zero-shot vision-language model showed that useful transfer is possible even without project-specific training. At the same time, neither approach is sufficiently reliable to support autonomous safety decisions.

The principal practical challenge is the transition from an image-level indication to an engineering quantity. A binary damaged/undamaged label is helpful for triage but does not by itself define a repair. The next development step is therefore a richer description of the defect: affected element, crack trace, width range, surface loss, local crushing and visibility. Structured field acquisition and spatial referencing are as important as model architecture because they provide the geometric and contextual information needed for this transition.

Another challenge is transferability. The Petrinja dataset reflects local building types, materials, finishes, camera conditions and damage patterns. Performance may decrease in a different region or after an earthquake producing other failure modes. The proposed framework addresses this issue by keeping human review in the loop and by allowing the classifier to be updated while the repair logic and cost database remain separate. Future validation should include additional earthquakes, external test sets and a direct comparison between model-derived quantities and bills of quantities prepared by experienced engineers.

8. CONCLUSIONS

The RELAR framework connects rapid field imagery, machine-learning-based damage screening and engineering quantification of repair works. On the Petrinja patch dataset, ConvNeXt-Tiny achieved the best classification result, with an F1-score of approximately 0.85, while ResNet-18 provided a slightly lower but stable baseline. Gemma 3 27B achieved an F1-score of approximately 0.79 without domain-specific training, although with a higher false-positive rate.

The work also showed that classification accuracy is only one part of a usable post-earthquake system. Images must be spatially referenced and must preserve enough component geometry for measurement. Repair quantities should be derived from observable damage and explicit engineering rules rather than from opaque average cost ratios. The proposed modular structure makes assumptions traceable and allows image models, repair rules and local prices to be improved independently.

The intended role of the system is rapid screening and decision support. Human engineers remain responsible for final safety evaluations, complex damage interpretation and escalation of uncertain cases. With this limitation clearly stated, the integration of image analysis and observation-based repair estimation can shorten the path from field evidence to practical recovery planning.

ACKNOWLEDGEMENTS

This research was supported by the Science Fund of the Republic of Serbia, Grant No. 7038, Rapid Earthquake Loss Assessment and Recovery Framework – RELAR. The authors acknowledge the members of the reconnaissance and annotation teams who contributed to the Petrinja image database.

REFERENCES

[1] Applied Technology Council. (1989). ATC-20: Procedures for Post-Earthquake Safety Evaluation of Buildings. Redwood City, CA: Applied Technology Council. [2] Federal Emergency Management Agency. (2018). FEMA P-58-1: Seismic Performance Assessment of Buildings, Volume 1 – Methodology. Washington, DC: FEMA.

[2] Federal Emergency Management Agency. (2018). FEMA P-58-2: Seismic Performance Assessment of Buildings, Volume 2 – Implementation Guide. Washington, DC: FEMA.

[3] Silva, V., Crowley, H., Pagani, M., Monelli, D., & Pinho, R. (2014). Development of the OpenQuake engine, the Global Earthquake Model's open-source software for seismic risk assessment. Natural Hazards, 72(3), 1409-1427.

[4] Erdik, M., Sesetyan, K., Demircioglu, M. B., Hancilar, U., & Zulfikar, C. (2011). Rapid earthquake loss assessment after damaging earthquakes. Soil Dynamics and Earthquake Engineering, 31(2), 247-266.

[5] Domingo, N., et al. (2025). Post-event cost estimation frameworks for building repair works incorporating multi-influencing factors. Journal of Construction Engineering and Management (In Press/Early Access).

[6] Stojadinović, B., et al. (2022). Damage-state-based repair-cost prediction for building portfolios: Opportunities and variability. Earthquake Engineering & Structural Dynamics, 51(8), 1845-1865.

[7] Booth, E., Saito, K., Spence, R., Madabhushi, G., & Boston, M. (2011). The use of remote sensing for post-earthquake damage assessment. Proceedings of the Institution of Civil Engineers-Structures and Buildings, 164(2), 123-133.

[8] Foroughnia, F., et al. (2024). Machine learning and satellite-based remote sensing for scalable post-earthquake damage severity mapping. Remote Sensing of Environment, 302, 113940.

[9] Mangalathu, S., Sun, H., Nweke, C. C., Yi, Z., & Burton, H. V. (2020). Classifying structural damage from images using deep learning and machine learning approaches. Automation in Construction, 113, 103144.

[10] Terzic, V., Merrifield, R., & Rogers, B. (2022). Linking component damage to functional recovery downtime: A performance-based framework for resilience. Earthquake Spectra, 38(3), 1920-1945.

[11] Burton, H., Deierlein, G., Lallemant, D., & Lin, T. (2016). Framework for incorporating probabilistic functional recovery into performance-based earthquake engineering. Structural Safety, 60, 23-36.

[12] Cohen, J. 1960. "A Coefficient of Agreement for Nominal Scales." Educ. Psychol. Meas., 37. https://doi.org/10.1177/001316446002000104.

[13] McHugh, M. L. 2012. "Interrater Reliability: The Kappa Statistic." Biochemia Medica, 22(3), 276-282. https://doi.org/10.11613/bm.2012.031.