13th National Conference on EARTHQUAKE ENGINEERING
PORTLAND, OREGON | JULY 13-17, 2026
Resilience Through Community Engagement and Technology

Interpretable AutoML for Post-Earthquake RC Building Damage: Insights from the 2023 Türkiye Sequence

M.A. Hariri-Ardebili[1,2] and S. Sattar[3]

ABSTRACT

Post-earthquake reconnaissance produces rich yet heterogeneous information that is hard to synthesize quickly for triage and decision-making. We summarize a two-part study of the February 2023 Türkiye earthquake sequence: (i) structured field data collection on >240 reinforced-concrete buildings across the most affected provinces; and (ii) an interpretable automated machine learning (AutoML) pipeline that classifies damage and extracts physically meaningful drivers via SHAP (Shapley Additive Explanations). Using 15 building and shaking features, the multi-class task (light/moderate/severe) achieved ~0.85 test accuracy with Random Forest and CatBoost after addressing class imbalance. A complementary binary task for critical damage utilized composite capacity–style indices (column/wall index and density) in conjunction with maximum PGA, PGV, and CAV (peak ground acceleration/velocity, and cumulative absolute velocity). It achieved test accuracy above 0.95. Interpretation indicates that building features dominate light damage, whereas ground motion measures — especially CAV and PGV — govern severe/critical damages. The overall recommendation is that if column and wall indices remain below 0.07 and 0.08, respectively, they will positively contribute to the likelihood of critical structural damage.

Introduction

Post–earthquake reconnaissance is typically carried out by multidisciplinary teams (engineers, geologists, seismologists, etc.) using on–site inspections, geospatial analysis, and modern sensing to document the physical effects of strong ground motion [1]. While indispensable for learning and response, the resulting products are often qualitative—driven by field notes, photographs, and expert judgment—and thus challenging to synthesize rapidly at scale for triage and decision–making [2].

This paper reports on a two–week reconnaissance mission following the February 6, 2023, Türkiye earthquake sequence and presents a compact, interpretable data analytics workflow to convert field observations into actionable insights. We focus on supervised, automated machine learning (AutoML) to (i) classify structural damage severity in reinforced–concrete (RC) buildings and (ii) extract transparent drivers of damage using SHAP (Shapley Additive Explanations). The aim is not to replace engineering assessment but to augment it with reproducible, quantitatively validated screening rules that can support rapid post–event tagging and prioritization.

[1] Researcher, College of Computer, Mathematical and Natural Sciences, University of Maryland, College Park, MD 20742, USA
[2] PREP Researcher, Engineering Laboratory, National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA (amin.hariri@nist.gov)
[3] Group Leader, Research Structural Engineer, Engineering Laboratory, National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA

Hariri-Ardebili MA, Sattar S. Interpretable AutoML for Post-Earthquake RC Building Damage: Insights from the 2023 Türkiye Sequence. *Proceedings of the 13th National Conference in Earthquake Engineering*, Earthquake Engineering Research Institute, Portland, OR. 2026.

Our contributions are threefold: (1) a curated building–level dataset coupling core geometric/architectural attributes with event–specific intensity measures; (2) a lightweight AutoML pipeline (with imbalance handling) that attains competitive accuracy for both multi–class and binary "critical" damage targets; and (3) physics–consistent interpretation that surfaces threshold–like behaviors (e.g., capacity–style indices and height transitions) directly valuable for practitioners. Throughout, we intentionally minimize ML theory and emphasize results, figures, and thresholds that can be operationalized in the field.

Reconnaissance Overview and Dataset

On February 06, 2023, a magnitude M_w 7.8 mainshock struck the East Anatolian Fault. A M_w 7.5 event followed about nine hours later. A M_w 6.4 event occurred on February 20th. Over 108,000 buildings collapsed or were severely damaged, resulting in major socio-economic losses [3].

A multi-institution team (including ACI Committee 133 members and NIST staff) surveyed the region in late March–early April 2023, traveling across cities such as Gaziantep, Nurdağı, Antakya, Adıyaman, Malatya, and Elbistan [4]. The building stock is predominantly reinforced concrete (RC). The team focused on RC frames (often a commercial first story with residential stories above), RC shear wall systems (primarily post-2000), and tunnel-form buildings. More than 240 RC buildings were documented. The team produced plan sketches (columns, walls, and spans) and assigned structural and nonstructural damage states. Three-dimensional scans were created for selected cases. Strong–motion records were obtained from AFAD (Türkiye's Disaster and Emergency Management Authority) through its TADAS system. The nearest stations to each building were used to estimate shaking. Fig. 1 shows the main shock locations, surveyed buildings, and stations [5].

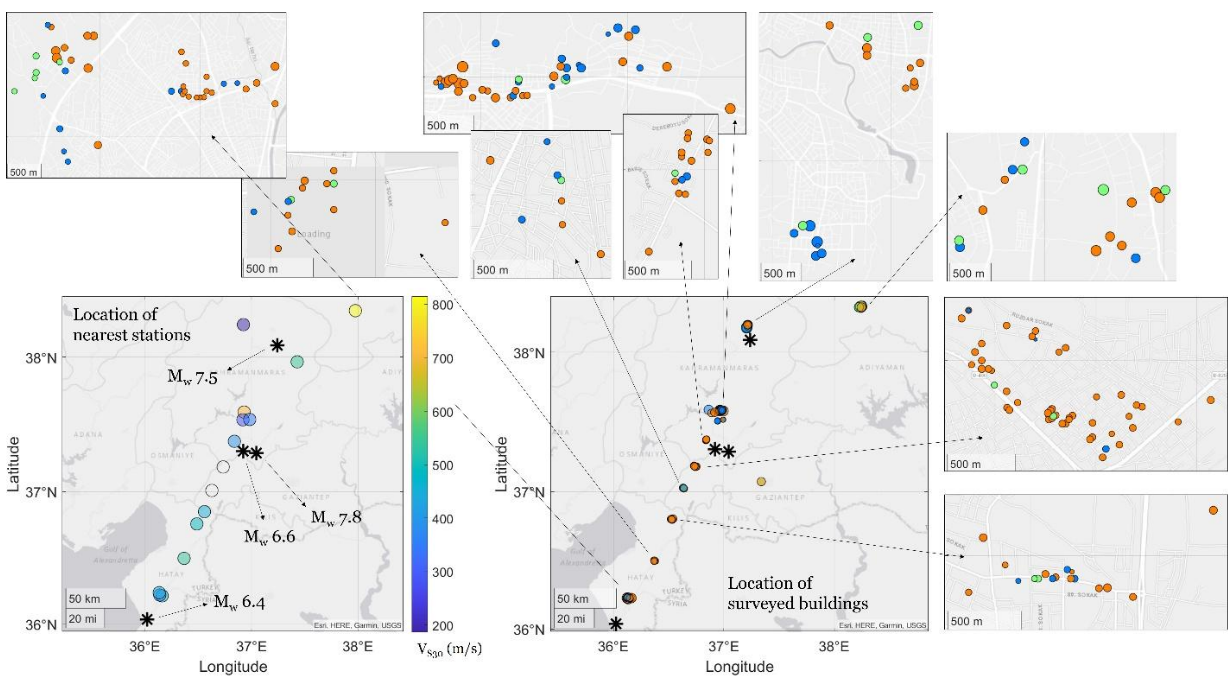


Figure 1. Mainshocks (asterisks) and surveyed buildings from the 2023 Türkiye earthquake sequence. Point colors indicate qualitative damage level: green = light, blue = moderate, orange = severe.

Structural damage was categorized as light, moderate, or severe [4]. To avoid extreme imbalance, cases with collapse/critical status were merged into the severe category. In the structural set, about 25% were light, 10% were moderate, and 65% were severe. The low number of moderate cases is a challenge for learning. We curated 15 raw features: building age; total height; typical floor area; total floor area above the critical floor; column area above the critical floor; concrete wall areas (minimum and maximum, combined into strong/weak directions); masonry wall areas (minimum and maximum); and, for the M_w 7.8, 7.5, and 6.4 events, PGA and epicentral distance at the building.

Seven intensity measures (IMs) were estimated at each building: PGA, PGV, spectral acceleration (S_a) at 0.3, 0.6, and 1.0 s, Arias Intensity, and CAV. Values were computed with a Bayesian spatial correlation update of ground motion model (GMM) predictions [6]. IMs are highly correlated (often a correlation coefficient of R>0.8-0.9). For small datasets, parsimony is important. We therefore emphasize PGA in the multi–class task and use max PGV and max CAV alongside max PGA in the binary task.

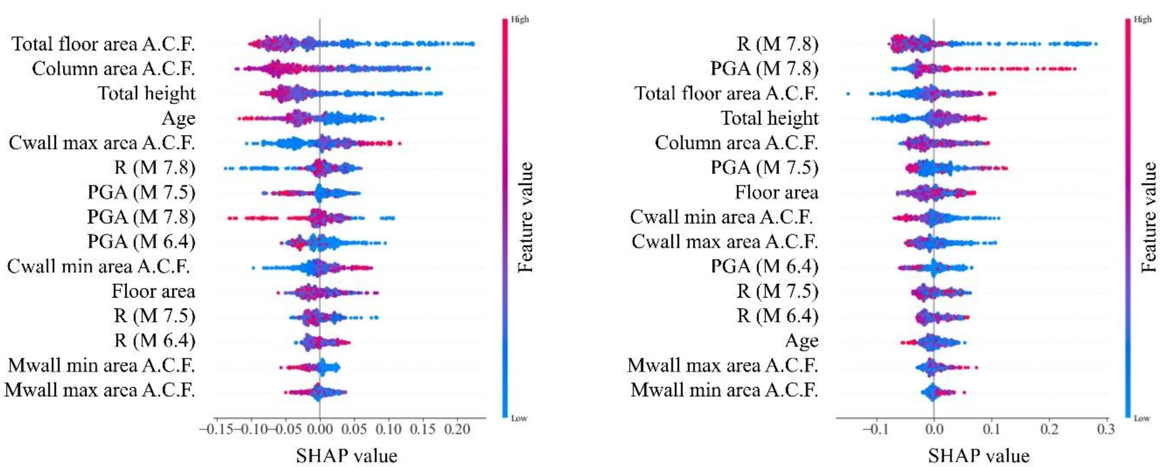
Methods in Brief (AutoML + SHAP)

We compare common tabular classifiers (Naïve Bayes, k-nearest neighbors, Decision Tree, Random Forest, AdaBoost, XGBoost, Light Gradient Boosting Machine, and CatBoost). Data are split 70/30 (train/test) with stratification by class [7]. We use 10 a 10-fold CV (cross–validation) on the training set for hyperparameter tuning. Since the moderate class is the minority class, we apply SMOTE only within the training folds. We report balanced accuracy and macro–F1 scores, as well as overall accuracy. Preprocessing is kept to a minimum to preserve engineering meaning.

We explain model outputs with SHAP [8]. The method assigns an additive contribution to each feature for each prediction, and the attributions sum to the model output. We use: (i) global importance (mean absolute SHAP) to rank drivers; (ii) beeswarm plots to show how feature values push predictions up or down; and (iii) dependence plots to reveal threshold–like behavior and interactions. We frame two supervised tasks. The multi–class task predicts light/moderate/severe from the 15 raw features. The binary task predicts critical vs. non–critical using composite capacity–style indices (Column Index, Wall Index, Column Density, Wall Density) and total height, together with max PGA, max PGV, and max CAV at the site.

Results: Multi-class

Global SHAP results indicate that two building features—column area above the critical floor (ACF) and total floor area above the critical floor—are among the top drivers, together with epicentral distance and PGA for the M_w 7.8 event. In the class–specific SHAP beeswarm plots of Fig. 2, each dot represents one building; the dot color encodes the feature value (red = high, blue = low), and the horizontal position indicates the SHAP value (right = increases the probability of the plotted class, left = decreases it). Vertical spread shows how widely the feature's effect varies across buildings (often due to interactions with other inputs). For light damage, smaller buildings (lower height and smaller areas) and larger concrete wall areas push dots to the right (higher likelihood of the light class), while higher shaking near the site pushes dots to the left (lower likelihood of light damage). For severe damage, higher PGA and closer distance push dots right; less concrete wall area and greater height also shift predictions toward the severe class.



(a) Light damage (b) Severe damage

Figure 2. SHAP plots for light and severe structural damage.

The dependence plots in Fig. 3 display SHAP value versus the raw feature on the x–axis (colored by a second feature to reveal interactions). Points to the right with positive SHAP indicate that larger feature values increase the probability of the class; a crossing near SHAP = 0 marks a threshold where the feature's contribution changes sign. Fig. 3 shows dependence plots for the severe class. We observe: (i) a height transition around 20 m where the SHAP value crosses zero; (ii) an inflection near PGA~0.6 g; (iii) a distance break near ~30 km; and (iv) a wall–area effect that decreases SHAP up to about 5 m^2 and then levels off. These features can be converted into simple screening cues for field use.

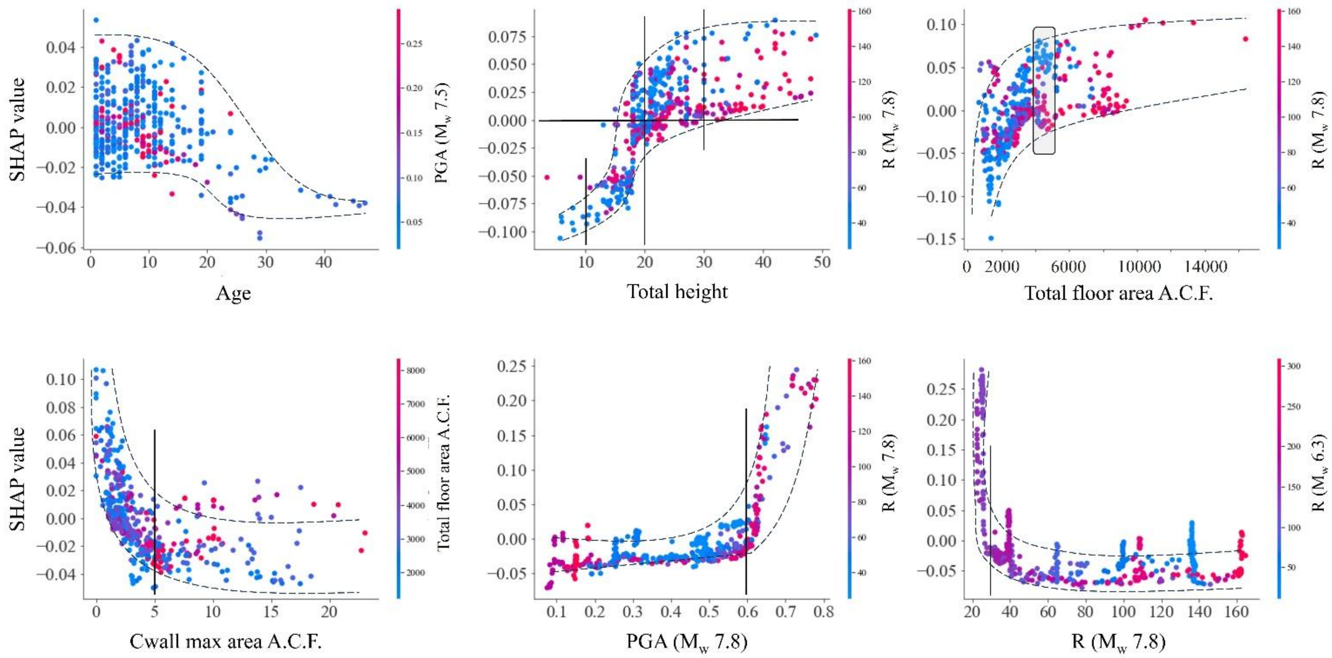


Figure 3. Dependence plots for selective input variables with respect to severe structural damage.

Summary

__What is new__? The workflow connects field data to transparent, threshold–level screening rules in a few steps. It strikes a balance between accuracy and interpretability, utilizing features that practitioners can quickly measure or estimate.

__Operational use__. For three–class damage, expect ~0.75 test accuracy from off–the–shelf AutoML, and about ~0.85 after oversampling the minority class. For critical/non–critical decisions, accuracy exceeded 0.95 when using capacity–style indices and maximum PGA/PGV/CAV. For triage, prioritize buildings with a small Column Index (CI) or Wall Index (WI), taller height, and large PGV/CAV at the site. If only the PGA is available, the ~0.6 g cue and a distance of ~30 km are helpful when combined with building features.

__Limitations__. We approximate cumulative shaking with the maximum IM per site. Mainshock–aftershock sequences may lead to nonlinear damage growth. Some material and geotechnical details were unavailable (e.g., foundation conditions and soil–structure interaction). The scarcity of moderate cases affects multi–class calibration even after oversampling.

Disclaimer

Certain commercial equipment, instruments, or materials are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology (NIST), nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

References

1. R. Spence, The full-scale laboratory: The practice of post-earthquake reconnaissance missions and their contribution to earthquake engineering: The third Nicholas Ambraseys lecture, Perspectives on European Earthquake Engineering and Seismology: Volume 1 (2014) 1–52.
2. M. A. Hariri-Ardebili, M. S. Speicher, Reconnaissance-informed post-earthquake functional recovery: Observations and challenges, Earthquake Spectra 41 (2025) 88–125.
3. M. A. Hariri-Ardebili, H. Tosun, Dams in the wake-up call of the 2023 Türkiye earthquake sequence: Insights from observed damages, risk assessment, and monitoring, International Journal of Disaster Risk Reduction 102 (2024) 104284.
4. S. Pujol, I. Bedirhanoglu, C. Donmez, J. Dowgala, M. Eryilmaz-Yildirim, K. Klaboe, F. B. Koroglu, R. Lequesne, B. Ozturk, L. Pledgera, S. E., Quantitative lessons from the earthquake sequence that affected south-east Türkiye in 2023, Earthquake Spectra 40 (2024) 505-530.
5. M. A. Hariri-Ardebili, S. Sattar, Data-driven insights into post-earthquake reconnaissance findings: 2023 Türkiye earthquake sequence, Earthquake Spectra 41 (2025) 58–87.
6. R. Pretell, S. Brandenberg, J. Stewart, Global and regionalized spatial correlation models for ground motion intensity measures, In Progress (2023).
7. X. He, K. Zhao, X. Chu, AutoML: A survey of the state-of-the-art, Knowledge-Based Systems 212 (2021) 106622.
8. S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Advances in Neural Information Processing Systems 30 (2017).