

Chapter 10

Turkiye Earthquake to Maui Wildfire: Generalizability of Aerial Rapid Post-Hazard Assessment Tools



Mohammad Hesam Soleimani-Babakamali, Rojiar Soleimani, Onur Avci, and Ertugrul Taciroglu

Abstract One well-known challenge in deploying parametric data-driven models is the generalizability to test data, which is also true for rapid post-hazard damage assessment methods. Generalizability to test data—refers to how well a parametric data-driven model, trained on specific sets of data, can adapt to new, unseen data. In the context of rapid post-hazard damage assessment methods, this problem arises because the model is typically trained on data from past disasters, which may not always reflect the full range of conditions or damage patterns seen in new events. The rapid post-hazard damage assessment tools are pre-trained on one or more “source” domains and are applied to “target” regions that can cover vast urban textures and different hazard-based damage manifestations, *i.e.*, the challenge of generalizability. To further explore and better understand such challenges, we study two recent major events: the 2023 Türkiye earthquake (source) and the Maui wildfire (target) incidents. The difference in hazard type and region highlights the challenges of generalizability and enables us to evaluate the effectiveness of several devised data and Machine Learning measures in addressing them. The proposed strategies include augmenting image channels with damage proxy maps, data fusion, deep ensemble learning, and Test Time Augmentation, which are needless for fine-tuning or labels from the target. The developed measures significantly improved damage detection with and without severity classification, with F1 scores rising from 0.71 to 0.87 and from 0.40 to 0.82 compared to the typical damage assessment pipelines, respectively

Keywords Natural hazards · Data fusion · Aerial damage assessment · Ensemble learning · Disaster response and recovery

Introduction

The increased urbanization and climate change have expanded human settlement exposures to natural incidents [1], causing considerable socio-economic consequences globally [2], [3], [4]. Following such incidents, response and recovery stages can alleviate these impacts and pave the road for quick socio-economic recovery [5]. The recent emergence of Artificial Intelligence (AI) and the broad application and ongoing research in the field of computer vision have allowed the employment of sources, such as satellite imagery, which become available post-hazard rapidly to accommodate these measures, offering great scalability and cost-effectiveness compared to traditional in-situ evaluations, either human-based or computer vision-

Mohammad Hesam Soleimani-Babakamali

Postdoctoral Research Associate, Civil & Environmental Eng Department, University of California, Los Angeles, CA 90095

e-mail: soleimanisam92@g.ucla.edu

Rojiar Soleimani

Ph.D. Student, Institute of Urban Design and Landscape Architecture, Vienna University of Technology, Vienna 1040, Austria

e-mail: rojiar.babakamali@tuwien.ac.at

Onur Avci

Assistant Professor, Department of Civil and Environmental Engineering, Engineering Sciences Building, 1306 Evansdale Drive, West Virginia University, Morgantown, WV 26506

(Corresponding author)

e-mail: onur.avci@mail.wvu.edu

Ertugrul Taciroglu

Professor, Civil & Environmental Eng Department, University of California, Los Angeles, CA 90095

e-mail: etacir@g.ucla.edu

based but with in-situ data acquisition, *e.g.*, 3D point clouds or Unmanned Aerial Vehicle (UAV)-based imagery. A rapid initial damage assessment also helps the response phase by guiding assessment efforts to more damage-prone areas, saving time and other resources. Satellite imagery offers fast availability, *e.g.*, public sources such as the MAXAR Open Data Program¹, which, for instance, publicly published the Maui wildfire [6] pre- and post-hazard satellite images one day after the incident. Satellite imagery coverage also offers unprecedented completeness, *i.e.*, coverage of the earth's surface, which becomes necessary as natural hazard events may occur at any point globally. Integrating AI-powered computer vision with satellite imagery can facilitate the automation of damage assessment, enabling a more efficient and accurate response to natural disasters. Still, computer vision tools for post-hazard damage assessment come with several limitations, including lower data resolution (details), which impacts the generalizability in downstream post-hazard tasks, compared to data sources such as field survey data with UAVs [7], street view images [8], and 3D point clouds [9].

Novel post-hazard rapid damage assessment methodologies have been proposed in the literature regarding two concepts: image segmentation [10] and object detection [11]. Image segmentation performs pixel-wise classification, *i.e.*, outputs patches of pixels representing a category (*e.g.*, building damage states). Object detection models detect and localize target objects on images through bounding boxes (*e.g.*, a damaged building). Both models are typically empowered with convolutional neural networks (CNNs) or other Deep Learning architectures, learning to identify patterns and features associated with the provided (damage-related) labels. One such challenge is the generalizability to unseen damage manifestations and urban textures. That is, when a model is trained on data from one type of hazard, it may not generalize well to another hazard with different characteristics. This issue is critical because post-hazard damage assessments often need to be quick and accurate in order to effectively allocate resources and prioritize rescue efforts. If a model overfits to its training data, it may fail to provide reliable assessments in future disaster scenarios, leading to inaccurate conclusions and suboptimal responses.

Although some studies have touched on generalizability issues with satellite images as the only visual input, they have focused mainly on small-scale regions, such as neighborhoods (*e.g.*, [12], [13], [14]) that do not provide significant variation in urban texture necessary to understand potential real-world generalizability challenges. Given the availability of state-of-the-art deep learning (DL) segmentation architectures, this study aims to enhance the generalizability of rapid (satellite image-based) damage assessment frameworks built upon these architectures. Specifically, the objective is to identify challenges associated with those frameworks' applications to unseen urban textures and diverse hazards and then develop data-oriented and Machine Learning-based strategies to overcome those challenges. To that end, a dataset comprising the 2023 Kahramanmaraş earthquake sequences in Türkiye [15] and the Lahaina wildfire on Hawaiian island of Maui [6] is constructed. The earthquake data serves as training (**source**) data, while the wildfire data simulates a real-world rapid damage assessment test area (**target**) without labeled data. These two locations exhibit significantly different urban textures and are affected by distinct predominant natural hazards, pushing state-of-the-art models to their limits and demonstrating real-world generalizability challenges. Various methods are designed and implemented to enhance generalizability during the training and testing. Devised measures include data fusion, *i.e.*, augmenting RGB channels with different damage proxy maps as additional data channels, deep pixel-wise ensemble learning across models with diverse architectures and inputs, and Test Time Augmentation (TTA) strategies.

Methodology

Considering a typical post-hazard damage assessment methodology using image segmentation models, a typical pipeline for the procedure is shown in Figure 1. This pipeline requires a pre-trained damage assessment model, a target region from which input data, namely satellite images, are captured, and the building footprints to overlay the damage maps and quantify damaged structures. A "measure" is introduced for each module to aid generalizability to the target area. To explore the pipeline and these measures, we first elaborate on assembling the Maui wildfire and Kahramanmaraş earthquake data, including the extraction and pre-processing of the images. The following subsection revolves around the Deep Learning (DL) network architectures, image data fusion, and training of these models using Kahramanmaraş (Türkiye) earthquake sequence data.

Satellite Imagery and Channel Augmentation

The first incident studied here is the 2023 Türkiye earthquake that struck southern and southeast Türkiye on February 6, 2023, near the city of Gaziantep [15]; this dataset serves the "source" role. These events affected at least 11 provinces and caused an estimated economic loss of around 110 billion dollars [16]. The "target" incident, namely the Maui (Hawaii)

¹MAXAR, Maxar open data program, <https://www.maxar.com/open-data>, accessed: 2023-09-10.

wildfire, began on the morning of August 8, 2023, followed by another outbreak later that afternoon, causing significant damage to the city of Lahaina [6]. For both events, MAXAR released pre- and post-hazard satellite images for the affected regions within one day, demonstrating the practicality of satellite images for rapid post-hazard damage assessment purposes. These images became available shortly on Google Maps API and are downloaded for this study. For the Turkiye incident, the city of Antakya, located in the heavily affected Hatay province [17], offers abundant satellite images for training damage assessment models and is thus chosen to train the source models. Lahaina city is chosen from the Maui incident as the major impacted region. The satellite images come in a single large raster, and to make them suitable for input to the downstream Machine Learning models, given the maximum resolution of public MAXAR open data, images are taken with a width and height of 640 pixels, almost equal to 150-by-150-meter blocks. Accordingly, the ($\sim 35 \text{ km}^2$, Figure 2a) Lahaina city region is divided into 1551 blocks, and the Antakya region ($\sim 128 \text{ km}^2$, Figure 2b) is divided into 5,679 blocks with their associated pre- and post-hazard images stored. The images are downloaded from Google and Bing Maps (see Figure 3), which are MAXAR-based yet more conveniently retrievable through their application programming interfaces.

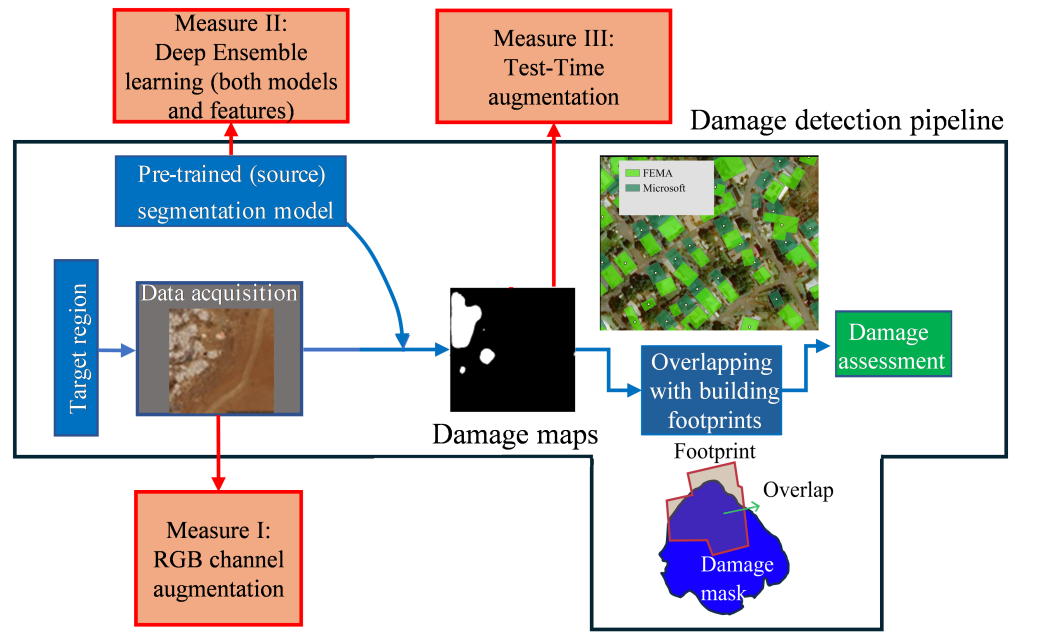


Fig. 1 Typical post-hazard damage detection pipeline from satellite imagery; the “Measure” boxes are the additional modules added in this research for possible generalizability enhancement.

Antakya image blocks are used to train the “source” damage detection model and must be labeled accordingly. The segmentation mechanism is chosen here over object detection, as building footprints are available from various public sources (e.g., FEMA² and Microsoft³). The associated uncertainty can thus be removed from the study on the proposed models’ generalizability. The segmentation model analyzes the visual features through Convolutional Neural Networks (CNNs) or other Deep Learning architectures and decides the label for each pixel. Hence, inside the training data, each pixel must be labeled (herein damage status) to train segmentation models. During inference, the model processes new images and generates segmentation masks, *i.e.*, pixel-wise probabilities for label categories, highlighting the areas likely to be damaged in the case of damage assessment problems (see Figure 1). These masks can then be used for further analysis or decision-making processes. For instance, they can be compared with building footprint boundaries (see Figure 1) to count damaged structures. Considering the pre-hazard image as the reference condition, any structural visual alterations can be interpreted as damages while labeling the Antakya image blocks.

Accordingly, two groups of labels are defined. For buildings with visible changes yet no collapsed parts, we use the label “D”. For buildings that collapse entirely or partially, we use the label “C”. Although C labels are relatively more straightforward, the D labels may require the pre-hazard images to be compared with post-hazard images. C and D aid in distinguishing damaged and collapsed buildings, which convey distinct consequences in loss estimation and affected population. Across 5,679 block images, 1,975 blocks include damaged buildings, with 879 C and 1,096 D segmentation patches labeled to train

²FEMA, Fema USA structures, <https://gis-fema.hub.arcgis.com/pages/usa-structure>, accessed: 2023-09-10.

³Microsoft US building footprints, <https://github.com/microsoft/USBuildingFootprints>, accessed:2023-09-10.

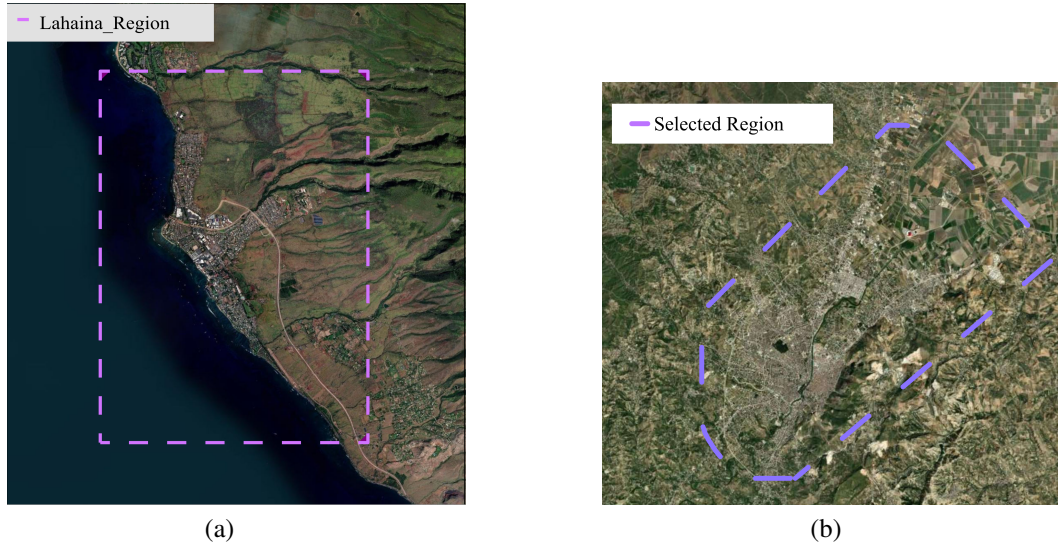


Fig. 2 Selected regions: (a) Lahaina, Maui County, Hawaii, USA and (b) the city of Antakya, Turkiye.

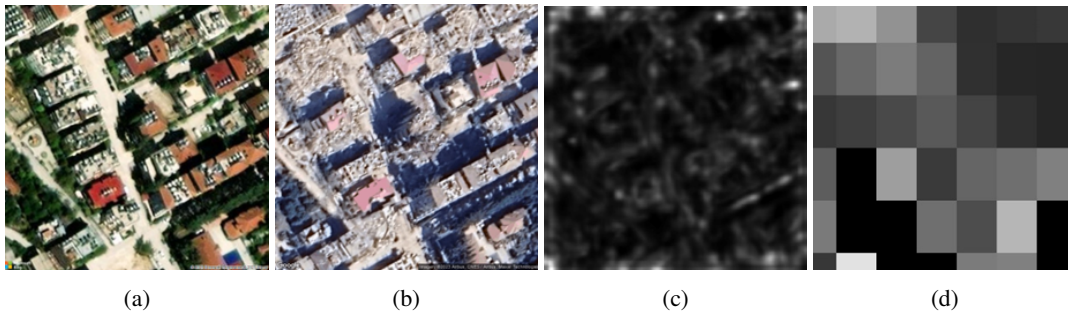


Fig. 3 Antakya block No. 1458 (a) pre-hazard satellite image, (b) post-hazard satellite image, (c) change detection (CD) maps between pre- and post-hazard images, (d) NASA ARIA map.

deep segmentation models. It is worth mentioning that even though two labels, C and D, are defined, the ultimate damage assessment will have three labels, D, C, and non-damaged (undamaged), which are the building footprints that did not match any damage masks (see Figure 1).

As noted earlier, regular post-hazard damage assessment techniques employ satellite images as inputs (only post-hazard images or combined pre- and post-hazard images). The first measure we implemented here is to augment the RGB satellite images with new channels (see Figure 1). The goal is to provide the network with salient visual features while minimizing visual intricacies caused by the varied urban textures between source and target regions, such as the predominant roof styles and building geometries, surrounding land cover, and differences in visual characteristics of training and testing images, *e.g.*, angle or luminosity. Two sets of such features are used, one by employing a **change detection (CD)** model and the other by employing a publicly available "damage proxy map" by NASA, namely the NASA ARIA⁴ project. NASA ARIA publishes these maps after major incidents and thus can be included seamlessly in the rapid post-hazard damage assessment pipelines. In this study, the "general-purpose" CD model proposed by El Amin et al. [18] is used, with an example shown in Figure 3d. These ARIA and resulting CD maps are stored for all blocks in Antakya and Lahaina, with the former used to train the Deep-Ensemble models, explained in the following section, and the latter used to test the proposed strategies in promoting generalizability.

Deep Learning Mechanisms

The image segmentation task involves learning the mapping between satellite images, CD, NASA ARIA maps, and the C and D masks. C and D are segmentation mask labels, and later footprints with over-threshold overlapping with them are

⁴NASA, Nasa Aria program, <https://aria.jpl.nasa.gov/>, accessed: 2023-09-10.

termed as C and D buildings. Various architectures exist for this task, from which the U-Net [19] and U-Net++ [20] are employed. For the U-Net's encoder layers, two models are employed: EfficientNet-b7 (CNN-based [21]) and Segformer mit_b5 (Visual Transformer-based [22]). The input data includes satellite images with and without CD and ARIA proxy map augmentations. This combination leads to 6 models, summarized in Table 1. The models are optimized for 20 epochs, and training time augmentation strategies are included to enhance generalizability. The pixel accuracy of the selected models is calculated and found to be similar, indicating that variations in channel augmentations or transitioning from CNN-based to Transformer-based models do not consistently affect accuracy. The minor fluctuations in accuracy are expected due to the similarity between hazard damage features and urban textures in the training and validation data.

Table 1 Semantic segmentation models and input layers.

Name	Model	Input channels	Architecture	Train pixel accuracy
Eff-Net	Efficientnet-b7	3:(RGB)	U-Net++	95.7
Eff-CD-Net	Efficientnet-b7	4:3(RGB)+1(CD)	U-Net++	95.9
Eff-ARIA-Net	Efficientnet-b7	4:3(RGB)+1(ARIA)	U-Net++	95.6
Tra-Net	Segformer-mit-b5	3 (RGB)	U-Net	95.6
Tra-CD-Net	Segformer-mit-b5	4:3(RGB)+1(CD)	U-Net	95.7
Tra-ARIA-Net	Segformer-mit-b5	4:3(RGB)+1(ARIA)	U-Net	95.6

Some models may overfit the data and focus on image intricacies rather than damaged features during optimization. To mitigate this, an ensemble strategy is devised across different models. The damage masks produced by each model (total of six) are combined to prioritize blocks with high agreement on damage presence and reduce masks in areas with low agreement. The segmentation mask for each damage category is represented as a 640-by-640 matrix mask with elements of 0 (non-damaged) or 1 (damaged). The following ruleset constructs the pixel-wise ensemble strategy: a) a block, *i.e.*, 640-by-640 pixel mask, is termed as damaged for a given model (total of six) if more than 1% of its pixels are labeled as damaged inside it, b) if a block is termed as damaged by all six models, the masks are then overlaid, c) if 4 to 5 models designate a block as damaged, then the majority vote ensemble is carried out for each pixel, and d) if less than four models agree on a block being damaged, then the whole block pixels' are labeled as non-damaged. The resulting "inter-model" ensemble is termed the Ensemble model henceforward.

Aside from using CD and ARIA maps, which are designed to provide the networks with salient damage features for both training and testing, the RGB satellite images are still used as the main input features, which may suffer from visual intricacies (*i.e.*, non-damage-related features, to which models may overfit). Accordingly, an "intra-model" ensemble strategy is employed using Test Time Augmentation (TTA) techniques. Although computer vision tools are mostly trained with image augmentation inside their training process, models may still lack generalizability to unseen areas. This problem, for instance, is shown in Figure 4 for Lahaina Block No. 408's C mask output. This figure reveals that the trained (with Antakya data) Eff-Net model suffers from generalizability issues for the Lahaina region, and the positioning of the damaged features inside the input image impacts its outcome. Accordingly, for each of the six models in Table 1, the intra-model ensemble is carried out with the main input image, 90°, 180°, and 270° clockwise (CW) rotated images, and the resulting masks are overlaid. The inter-model ensemble model with the masks produced by the intra-model ensemble is termed as the Ensemble-R model hereafter.

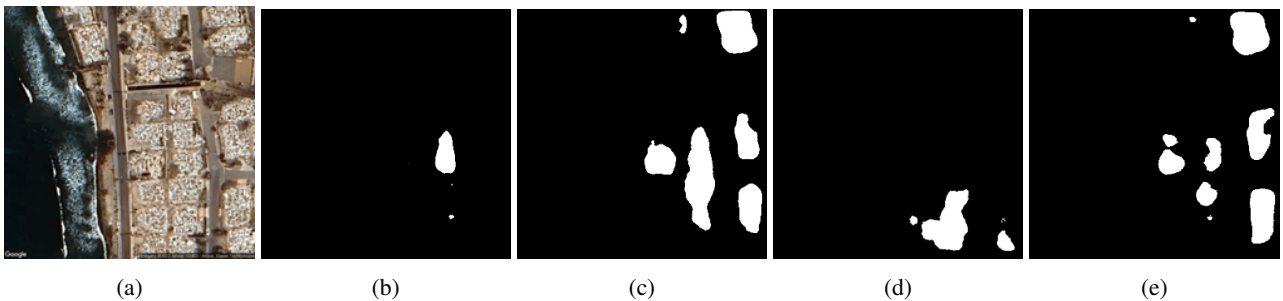


Fig. 4 Lahaina block No. 408's Eff-Net C mask output (a) regular image input, (b) 90° (CW) rotated image, (c) 180° (CW) rotated image, (d) 270° (CW) rotated image.

Results and Discussion

This section evaluates the damage estimation results using each deep segmentation model and assesses the proposed design's effectiveness in promoting generalizability. The results of the damage detection model are evaluated in two ways. First, damage maps produced by the proposed framework are compared to the Maui damage assessment by ESRI⁵. ESRI (Environmental Systems Research Institute) is a geographic information system (GIS) software providing tools and platforms for mapping, spatial analysis, and location intelligence. The Maui damage assessment results are produced using a Deep Learning model that was trained with UAV images from the Woolsey fire⁶, and then fine-tuned for the Lahaina incident. The second evaluation is done by first labeling each building in categories D, C, and non-damaged (undamaged) to compare different models and strategies in terms of performance. Multi-class accuracy metrics, including precision, recall, and the F1 score, are used to assess damage detection outcomes and determine how different measures affect them.

For two individual Eff-Net and Tra-Net and the Ensemble and Ensemble-R model, the resulting damage map with the overlap threshold of 25% is plotted against the Maui damage assessment by ESRI in Figure 5 (if the building footprint overlapping with the damage mask is 25% or more of its area, then that structure is labeled as “damaged”, see Figure 1 for mask/footprint matching). Eff-Net missed most of the damaged buildings compared to the damage map obtained by ESRI and the Ensemble approach. On the other hand, Tra-Net offers a better distribution and quantity of damaged buildings, yet it has numerous misclassifications between C and D masks. The Ensemble model tends to improve both the quantity and distribution of damaged buildings and offer a better damage severity assessment than the individual models. Ensemble-R reports a comparable ratio of damaged buildings, 61.73% compared to 64.68% of ESRI, with the damage distribution also in good agreement with that of ESRI. Furthermore, Ensemble-R can also classify the damage's severity into two classes of D and C damage, a capability that the ESRI model lacks. According to accuracy metrics.

Table 2 Lahaina wildfire damage detection performance with and without damage severity classification.

Model	Damage detection task (two categories)			Damage detection + Severity tasks (three categories)		
	Precision (macro)	Recall (macro)	F1 Score (macro)	Precision (macro)	Recall (macro)	F1 Score (macro)
Eff-Net	0.68	0.61	0.45	0.53	0.42	0.33
Eff-CD-Net	0.70	0.69	0.58	0.50	0.48	0.41
Eff-ARIA-Net	0.69	0.65	0.52	0.56	0.45	0.37
Tra-Net	0.73	0.77	0.71	0.55	0.51	0.40
Tra-CD-Net	0.76	0.80	0.73	0.59	0.58	0.55
Tra-ARIA-Net	0.67	0.66	0.54	0.61	0.50	0.45
Ensemble	0.76	0.80	0.73	0.63	0.58	0.55
Ensemble-R	0.85	0.89	0.87	0.80	0.85	0.82

To compare the six individual models and evaluate the effectiveness of the proposed measures, Table 2 recaps their damage detection performance. Table 2 includes two classification categories: the first refers to the binary damage/non-damage classification (*i.e.*, two-label damage detection) to evaluate overall damage detection performance, while in the second assessment, damage severity is also measured by having three labels of non-damage, and D and C damages. According to the earlier inventory labeling presented at the beginning of the Results section, 59.6% of the buildings were labeled as C, 9.8% as D, and 30.6% as no-damage buildings. Consequently, damage detection data has 69.4% damaged buildings, while the D class constitutes only 9.8% of data in the damage severity task, implying a severely imbalanced classification task. Due to the imbalance in both problems, macro-estimations are employed in Table 2. In the damage detection task, Eff-Net, a CNN-based model, falls short of even a random guess, achieving an F1 score below 50%. This poor performance is evident in Figure 5b also, where Eff-Net misses a significant number of damaged buildings. In contrast, Tra-Net achieves a respectable F1 score of 0.71. However, in the more challenging task of damage detection with severity measurement, Eff-Net again performs no better than chance, while Tra-Net, despite its strong performance in the damage detection task, struggles with misclassifying between (**fully or partially collapsed**) C and (**damaged but not collapsed**) D categories, as seen in Figure 5c. The incorporation of CD and ARIA maps as channel augmentations has been shown to improve damage detection

⁵ESRI, Maui damage assessment, <https://www.arcgis.com/apps/dashboards/c6ba275c2762470bbaf0c660a89a46e7>, accessed: 2023-09-10.

⁶USGS, Woolsey fire, <https://www.usgs.gov/media/before-after/woolsey-fire>, accessed: 2023-09-10.

and severity classifications. However, the RGB channels still carry environmental features that can affect model decision-making. This is exemplified by the generalizability issues observed in individual models, such as the Eff-Net C-mask model, which fails to capture damaged structures in certain images. Rotating the images by 90, 180, and 270 degrees counterclockwise and re-feeding them into the model improves segmentation outcomes, highlighting the impact of urban textures and building placement on model detection ability. The application of Test-Time Augmentation (TTA) methods, such as rotation, can enhance model performance and provide more accurate segmentation results. Moreover, the ensemble strategy, which combines the outputs of six models, can be further refined by applying TTA to each model, resulting in the Ensemble-R model. The performance of the Ensemble and Ensemble-R models is compared, demonstrating improved damage detection and severity measurement capabilities. The resulting masks offer greater confidence in detecting damage and measuring severity.

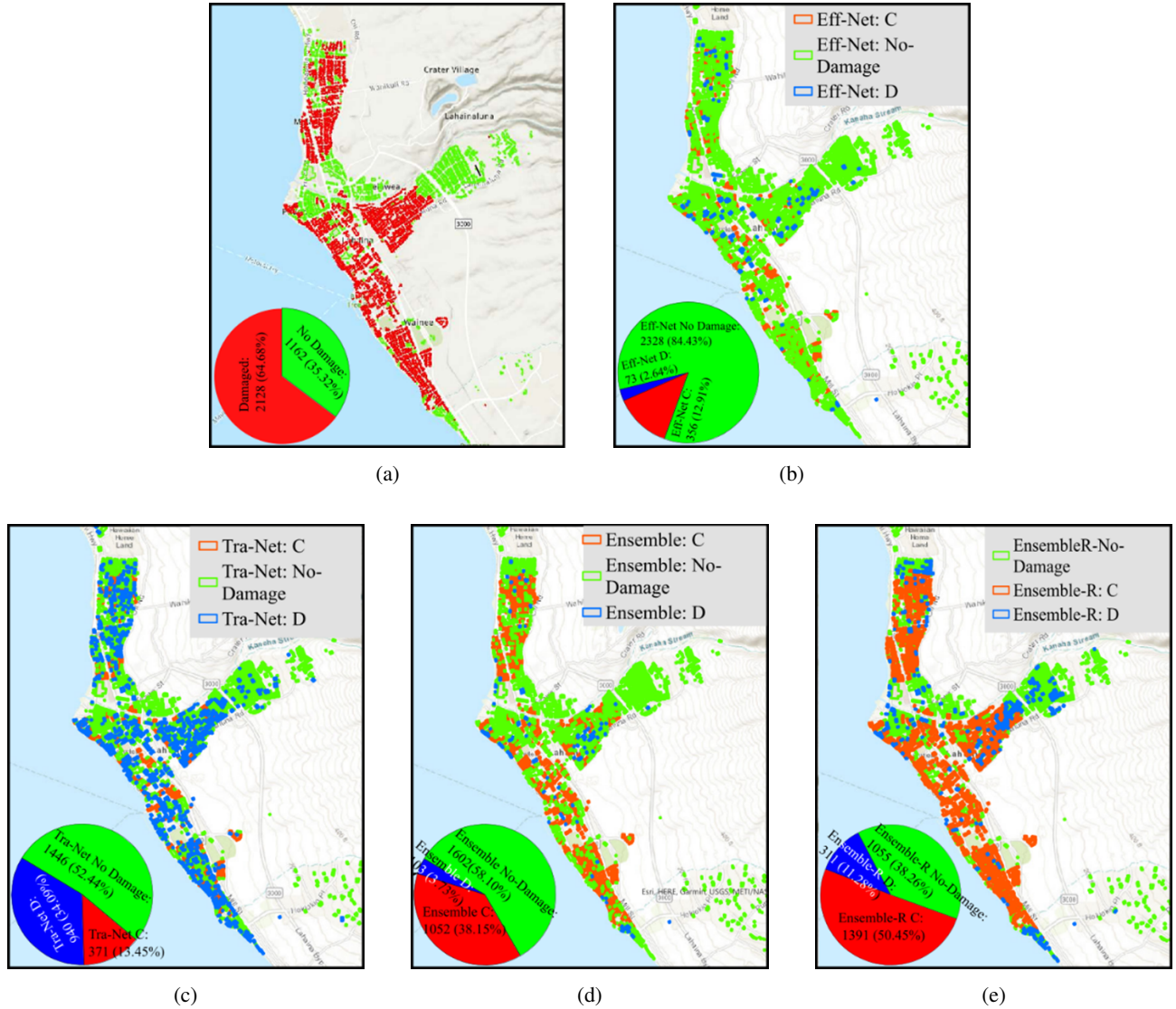


Fig. 5 Lahaina damage assessment: (a) ESRI estimation (ESRI, 2023), (b) Eff-Net estimation, (c) Tra-Net estimation, (d) Ensemble estimation, (e) Ensemble estimation.

Conclusions

Rapid first-hand damage assessments after natural incidents substantially aid response measures. The generalizability concerns in such assessments are vital as pre-trained models must be applied to the impacted region without any fine-tuning (due

to lack of time due to the time-consuming labeling procedures) or verified inspection reports from the field. Thus, data and Machine Learning solutions must be devised to promote the generalizability of a rapid post-hazard assessment tool training performed in one or more "source" areas to the desired target regions. Based on this, various measures are developed and introduced in this manuscript that entail no fine-tuning on the target site and rely solely on public data sources. The aerial data from the 2023 Türkiye earthquake (source) and the Maui wildfire incident (target) are used first to investigate the generalizability concerns and then evaluate how well the developed Machine Learning strategies overcome those. To promote generalizability, various measures were designed and evaluated. Satellite images were fused with two damage proxy maps, a change detection model and NASA ARIA maps, to alleviate differences in urban texture and visual artifacts. Six deep segmentation models (CNN- and Visual Transformer-based) were trained with normal and augmented satellite images. A two-level ensemble strategy was developed: (1) an intra-model ensemble using Test Time Augmentation (TTA) strategies and (2) a pixel-based deep ensemble strategy across the six models. These measures were applied to the test area of Lahaina city, resulting in improved generalizability and high accuracies: macro F1 scores of 0.87 for damage detection and 0.91 for damage severity categorization, outperforming a single Visual Transformer-based model without interventions (0.71 and 0.50, respectively).

References

1. United Nations. Department of Economic and Social Affairs, *The World's Cities in 2016*. UN, 2016.
2. N. N. Ridder, A. M. Ukkola, A. J. Pitman, and S. E. Perkins-Kirkpatrick, "Increased occurrence of high impact compound events under climate change", *NPJ Clim Atmos Sci*, vol. 5, no. 1, p. 3, 2022.
3. K. Do, M. Mahish, A. K. Yeganeh, Z. Gao, C. L. Blanchard, and C. E. Ivey, "Emerging investigator series: a machine learning approach to quantify the impact of meteorology on tropospheric ozone in the inland southern California", *Environmental Science: Atmospheres*, 2023.
4. D. Nelson, Y. Choi, B. Sadeghi, A. K. Yeganeh, M. Ghahremanloo, and J. Park, "A comprehensive approach combining positive matrix factorization modeling, meteorology, and machine learning for source apportionment of surface ozone precursors: Underlying factors contributing to ozone formation in Houston, Texas", *Environmental Pollution*, vol. 334, p. 122223, 2023.
5. W. Sun, P. Bocchini, and B. D. Davison, "Applications of artificial intelligence for disaster management", *Natural Hazards*, vol. 103, no. 3, pp. 2631–2689, 2020.
6. T. W. Juliano et al., "Brief communication: The Lahaina Fire disaster: How models can be used to understand and predict wildfires", *Natural Hazards and Earth System Sciences Discussions*, vol. 2023, pp. 1–7, 2023.
7. M. M. L. Ramadhan, G. Jati, M. R. Alhamidi, M. H. Hilman, W. Jatmiko, and others, "Collapsed Building Detection Using Residual Siamese Neural Network On LiDAR Data", in *2021 6th International Workshop on Big Data and Information Security (IWBI)*, 2021, pp. 29–34.
8. D. Duarte, F. Nex, N. Kerle, and G. Vosselman, "Multi-resolution feature fusion for image classification of building damages with convolutional neural networks", *Remote Sens (Basel)*, vol. 10, no. 10, p. 1636, 2018.
9. F. Yamazaki, W. Liu, and K. Horie, "Use of Multi-Temporal LiDAR Data to Extract Collapsed Buildings and to Monitor Their Removal Process after the 2016 Kumamoto Earthquake", *Remote Sens (Basel)*, vol. 14, no. 23, 2022, doi: 10.3390/rs14235970.
10. Y. Wang, A. W. Z. Chew, and L. Zhang, "Building damage detection from satellite images after natural disasters on extremely imbalanced datasets", *Autom Constr*, vol. 140, p. 104328, 2022.
11. Z. Liu et al., "Remote sensing-enhanced transfer learning approach for agricultural damage and change detection: A deep learning perspective", *Big Data Research*, vol. 36, p. 100449, 2024.
12. W. Yang, X. Zhang, and P. Luo, "Transferability of convolutional neural network models for identifying damaged buildings due to earthquake", *Remote Sens (Basel)*, vol. 13, no. 3, p. 504, 2021.
13. J. Ge, H. Tang, N. Yang, and Y. Hu, "Rapid identification of damaged buildings using incremental learning with transferred data from historical natural disaster cases", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 195, pp. 105–128, 2023, doi: <https://doi.org/10.1016/j.isprsjprs.2022.11.010>.
14. M. Janalipour and A. Mohammadzadeh, "A fuzzy-ga based decision making system for detecting damaged buildings from high-spatial resolution optical images", *Remote Sens (Basel)*, vol. 9, no. 4, p. 349, 2017.
15. USGS, "M 7.8 - Pazarcik earthquake, Kahramanmaraş earthquake sequence", 2023.
16. M. Ozturk, M. H. Arslan, and H. H. Korkmaz, "Effect on RC buildings of 6 February 2023 Turkey earthquake doublets and new doctrines for seismic design", *Eng Fail Anal*, vol. 153, p. 107521, 2023.
17. T. government, "2023 Kahramanmaraş damage assessment (in Turkish)", 2023.
18. A. M. El Amin, Q. Liu, and Y. Wang, "Convolutional neural network features based change detection in satellite images", in *First International Workshop on Pattern Recognition*, 2016, pp. 181–186.
19. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation", in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III* 18, 2015, pp. 234–241.
20. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation", *IEEE Trans Med Imaging*, vol. 39, no. 6, pp. 1856–1867, 2019.
21. M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks", in *International conference on machine learning*, 2019, pp. 6105–6114.
22. E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with transformers", *Adv Neural Inf Process Syst*, vol. 34, pp. 12077–12090, 2021.