2022 ACWE
14th Americas Conference on Wind Engineering
May 17-19, 2022 – Lubbock, TX
National Wind Institute – Texas Tech University

Hybrid Framework for Post-Hazard Building Performance Assessments with Application to Hurricanes

David B. Roueche [a], Jordan O. Nakayama [b], Barbaros M. Cetiner[c], Sabarethinam Kameshwar [d], Tracy L. Kijewski-Correa [e]

[a] *Auburn University, Auburn, AL, dbr0011@auburn.edu*
[b] *Auburn University, Auburn, AL, jon0003@auburn.edu*
[c] *University of California, Los Angeles, CA, bacetiner@ucla.edu*
[d] *Louisiana State University, Baton Rouge, LA, skameshwar1@lsu.edu*
[e] *University of Notre Dame, Notre Dame, IN, Tracy.L.Correa.5@nd.edu*

ABSTRACT: This study describes a hybrid framework for post-hazard building performance assessments. The framework relies upon rapid imaging data collected by regional scout teams being integrated into broader data platforms that are parsed by virtual teams of hazards engineers to efficiently create robust performance assessment datasets. The study also pilots a machine-in-the-loop approach whereby deep learning and computer vision-based models are used to automatically define common building attributes, enabling hazard engineers to focus more of their efforts on precise damage quantification and other more nuanced elements of performance assessments. The framework shows promise, but to achieve optimal accuracy of the automated methods requires regional tuning.

KEYWORDS: Reconnaissance; hurricane; deep learning; performance assessment.

1 INTRODUCTION

Post-event reconnaissance has long played a critical role in natural hazards engineering, spurring advancements in science, policy, and practice. Traditional approaches to reconnaissance have primarily utilized on-site assessments with high-resolution, GPS-enabled cameras and field notes capturing perishable information on select sample structures. Alternatively, for larger areas, coarse damage assessments have been conducted using remote sensing technologies such as medium and high-resolution aerial imagery (e.g., Massarra et al. 2019). Digital data collection and virtual assessment platforms (e.g., Gurley and Masters, 2011; Kijewski-Correa et al. 2014; Roueche et al. 2018) have also been developed and integrated into reconnaissance workflows. However, the reconnaissance community currently has much greater capability to capture perishable field data than it does to efficiently parse and transform it into useful engineering data. The Structural Extreme Events Reconnaissance (StEER) network was formed in 2018 to help address this challenge amongst others (Kijewski-Correa et al. 2021). The mission of StEER is to deepen the capacity of the Natural Hazards Engineering (NHE) community for coordinated and standardized assessments of the performance of the built environment following natural hazard events. To date, enhancement and standardization of the perishable data captured by field reconnaissance teams is largely a human effort. The need exists for hybridized human-in-the-loop or machine-in-the-loop approaches to effectively leverage automation in producing robust yet accurate reconnaissance datasets. The objective of this study is to pilot a hybrid framework for post-hazard building performance assessments used by the StEER network following Hurricane Ida (2021), and highlight advantages, challenges, and remaining gaps of this approach.

2 METHODS

2.1 *Hurricanes Laura (2020) and Ida (2021) Data Collection*

Hurricanes Laura (Category 4, landfall on 27 August 2020) and Ida (Category 4, landfall on 29 August 2021) were two of four hurricane landfalls in Louisiana in 2020 and 2021, causing over $100B in economic losses between them. Hurricane Laura caused the heaviest damage in Lake Charles, LA, where peak 3-second gusts (at 10 m height in open terrain) were measured at 59 m/s (132 mph), near design levels (Roueche et al. 2020). Hurricane Ida made landfall near Port Fourchon, LA, before impacting Houma, New Orleans, and other communities in southeast LA. Peak gust wind speeds were estimated to be 57 m/s (128 mph) near landfall (Prevatt et al. 2021). In part by limitations caused by the ongoing COVID-19 pandemic, the StEER network responded to both events by activating a network of virtual assessment team members, while simultaneously deploying small scout teams led by the first and second authors equipped with rapid imaging techniques including vehicle-mounted street-level panoramic imaging platforms and Unmanned Aerial Systems (UAS). For Hurricane Laura, the scout teams also performed targeted forensic load path assessments within select building clusters of interest. Additional street-level panoramas were captured by the University of Hawaii National Disaster Preparedness and Resilience Center and the NHERI RAPID Experimental Facility (Berman et al. 2020) and used in this study. All panoramic images were collected using NC Tech iStar Pulsar cameras, which captured a 360° by 160° field of view within an 11k stitched panorama (11000 by 5500 pixels) geotagged with ~2.5 m horizontal accuracy. A summary of the panoramic imagery collected by StEER following Hurricanes Laura (2020) and Ida (2021) is provided in Table 1.

Table 1. Surface-level panoramas captured by StEER during Hurricanes Laura and Ida.

	Hurricane Laura (2020)	Hurricane Ida (2021)
Total Length of Routes (km)	842	541
Total Number of Panoramas	842,496	135,250

2.2 *Virtual Reconnaissance*

Following the deployments, imagery was archived on the DesignSafe-CI platform (Rathje et al. 2017) and uploaded to cloud-based viewing platforms to facilitate easy access for remote assessments. Surface-level panoramas were uploaded to the Mapillary imaging platform. UAS imagery was processed into 2D orthomosaics and 3D point clouds and hosted on a cloud-based platform using proprietary software with licenses available through the NHERI RAPID EF. A database of supplemental resources was also generated, including aerial imagery from NOAA, property appraisal websites for affected parishes, and realtor websites. Building performance datasets were then constructed via remote assessments based on the various pre-storm and post-storm data sources. Samples were generated in clusters across the hazard gradient using pre-storm imagery to remove potential damage biases. Clusters were primarily selected in regions where overlapping terrestrial and aerial imagery was present, and, for Hurricane Laura (2020), where forensic load path assessments had been conducted that could be used to infer structural characteristics for similar buildings within the cluster.

For Hurricane Laura (2020), 439 virtual assessments were conducted manually by a team of undergraduate students from Auburn University and the University of Notre Dame. For Hurricane Ida (2021), virtual assessments were conducted manually on a targeted dataset of 540 buildings. Targets were selected without prior knowledge of their damage levels, and were stratified based on occupancy, location with the hurricane wind field, and approximate year of construction. As a feasibility study, assessments focused on a limited subset (Table 2) of the full suite of data fields typically documented in StEER building performance datasets (Roueche et al. 2021).

The virtual assessments were conducted by manually reviewing the available data sources, which included pre- and post-event aerial and surface-level panoramic imagery, oblique imagery from the Civil Air Patrol, and realtor websites. Review of multiple sources was often needed because each source had different geospatial and temporal extents, and resolution. For example, the NOAA post-event aerial imagery was typically collected the soonest after landfall (typically within 24-48 hours), before repairs and most cleanup began, but coverage was not universal, and resolution was better in certain areas than others. Post-storm surface-level panoramas provided necessary views of the elevations of each target, and resolution was typically sufficient for precise damage quantification, but some of the data was collected after initial cleanup and roof tarping began, obscuring some of the key damage details, and all four elevations of the buildings were not always visible.

For Hurricane Laura (2020), each virtual assessment took approximately 16 minutes to complete on average. Much of the time required was in assembling the various sources for a given record and comparing between them. To help facilitate more rapid reconnaissance for Hurricane Ida, pre-populated fields were added to the Fulcrum app that automatically generated web links to pre- and post-event imagery wherever possible using the location of the target building and modified, location-specific URLs. The scope of the virtual assessments were also reduced slightly (ignoring roof slope, structural system, foundation type, and wall cladding). On average, the manual reconnaissance process took just over 9 minutes for each assessment for Hurricane Ida (2021).

Table 2. Data fields prioritized for virtual reconnaissance common to both storms.

Name	Data Type	Description
Latitude, Longitude	Numeric	GPS coordinates for the centroid of the building footprint
Investigator	Text	Name of the virtual investigator
Address	Text	Physical address of the building
Occupancy	Single Choice	Building occupancy class per the International Code Council building code
Roof Shape	Multi-Choice	Gable, Hip, Flat or combinations of the above
Number of Stories	Numeric	Number of stories, ignoring ground floors of elevated buildings
First Floor Elevation	Numeric	Hight above ground to the lowest horizontal structural member
Roof Structure Damage	Percentage	Roof framing members damaged
Roof Substrate Damage	Percentage	Roof substrate (i.e., decking) damage (if present)
Roof Cover Damage	Percentage	Roof cover damaged or missing
Wall Structure Damage	Percentage	Wall framing members damaged or missing
Wall Substrate Damage	Percentage	Wall substrate (i.e., sheathing) damaged or missing
Wall Cladding Damage	Percentage	Wall cladding damaged or missing
Foundation Failure	Binary	Indicator for whether failure occurred at the foundation level
Wind Damage Rating	Categorical	Discrete damage levels based on component-level wind-induced damage

Surge Damage Rating	Categorical	Discrete damage levels based on evidence of surge impacts

2.3 *Automated Attribute Assignment*

The Building Recognition using Artificial Intelligence at Large Scale (BRAILS) tool developed by the NHERI SimCenter (NHERI SimCenter 2022a) was employed to demonstrate the potential for automating attribute assignments within a hybridized, machine-in-the-loop reconnaissance framework. BRAILS was developed as a tool for building out exposure datasets to facilitate regional simulations of the response of the built environment to natural hazards at the building scale. In alignment with its original purpose, BRAILS contains, amongst others, modules for detecting the roof shape and roof cover type, using pre-event aerial imagery as the input, and modules for classifying building occupancy, detecting number of floors, presence of raised foundations, garages, chimneys, and predicting building first-floor elevation, height, roof pitch and façade-to-window ratios using pre-event street-level imagery as the input. BRAILS utilizes machine learning, deep learning, and computer vision techniques for these modules, and is written in Python as an open-source tool for the research community.

The baseline occupancy classifier in BRAILS (Wang 2021) is built on ResNet50, a residual deep learning neural network model with 50 layers, and is trained on 15,743 labeled images. The occupancy labels for this dataset were obtained from two sources: OpenStreetMaps and New Jersey Department of Environmental Protection Building Footprints dataset, with 7,942, and 7,801 images drawn from each data source respectively. The baseline roof shape classifier in BRAILS (Wang 2020) was also trained using ResNet50 architecture, with training data consisting of 6,000 labeled images, 2,000 each of the three classes – flat, gable, and hip. The labels were assigned from OpenStreetMap and were manually reviewed to ensure they areaccurate and paired with images free from obstructions with clear views of the entire roof. The Number of Floors detector (Cetiner 2020) was trained on the EfficientDet-D4 architecture using a dataset of 80,000 building images from SimCenter testbeds in New Jersey and Louisiana. Finally, the Elevated Foundation classifier (Hornauer 2020) uses the ResNet-50 architecture and was trained on a dataset of 1,200 building images manually labeled from an early version of StEER. For additional implementation details and information on rest of the BRAILS modules please see official BRAILS Documentation (NHERI SimCenter 2022b).

For this pilot study, each of these four BRAILS modules were used on the Hurricane Ida dataset only, because (1) it contained a more diverse set of buildings relative to Hurricane Laura, which was primarily composed of single-story, single-family residential buildings; and (2) the input imagery from Google Satellite and Street View for the target buildings were extracted prior to the imagery being updated with post-storm conditions.

The initial predictions for roof shape, occupancy, number of floors and elevated foundation using BRAILS involved the following steps:

1) Pre-event aerial and ground-based imagery was downloaded for each building using the location of the building and the Google Maps and Google Street View APIs.
2) Images were manually reviewed to discard any unsuitable images, such as those with blocked views of the target building due to trees, and blank images due to no Street View data being available. Predictions were only made on the remaining buildings.
3) The BRAILS modules were executed on the input images using a local PC workstation.
4) The manually tagged StEER data was mapped to the BRAILS classification options to evaluate the differences between the manual and predicted attribute classes.

Regarding the final step, StEER protocols specify a more refined class list than what is currently implemented in BRAILS, and subsequently the more refined class list had to be mapped to the

BRAILS classification lists. The BRAILS Roof Type Classifier has a class for flat, gabled, and hipped roofs. Any StEER roof shape labeled hip, complex, or gable/hip combo, was labeled as hip for the BRAILS retraining. The remainder of the StEER classifications corresponded to gable or flat and were matched accordingly to the BRAILS classification. The Occupancy Classifier label COM1 was mapped to any non-residential structure in the dataset (i.e., school, other institutional, utilities buildings), RES1 was mapped to all single-family homes, and RES2 was mapped to all multi-family homes including duplexes. Number of stories did not require any re-mapping. For elevated foundations, StEER quantifies the first-floor elevation above ground level, while the BRAILS Elevated Foundation module is binary, indicating whether the building is elevated or not. It was assumed that any building in the StEER dataset with a first-floor elevation greater than 4 ft has an elevated foundation.

2.4 *Retraining BRAILS*

The BRAILS classifiers were originally trained on diverse datasets that may not be well-tuned to regional construction practices for a given target region. Therefore, a secondary goal of the study was to evaluate the effect on the prediction accuracy of BRAILS by retraining the modules on a portion of the Hurricane Ida dataset and using the retrained model to predict the attributes of the remaining portion of the dataset.

As a pilot study, two of the four modules used in this analysis were therefore retrained using a subset of images and tested against the remaining images. The retraining splits for each module are summarized in Table 3. A weighted average was used to distribute the percentage of each module's retraining images. An additional retraining split was used where an equal number of images from each class was used to train the module, as unbalanced training data has been shown to negatively impact the predictive capabilities of such classifiers (Buda et al. 2018).

Table 3. Number of images used to retrain the BRAILS roof type and occupancy classifiers by retraining version. Version 1.0 refers to the pretrained classifiers.

Module	Classification	Total No. of Images	Version 2 (10%)	Version 3 (25%)	Version 4 (50%)	Version 5 (75%)	Version 6 (Equal)
Roof Type Classifier	Flat	14	1	4	7	11	14
	Gable	224	22	56	112	168	14
	Hip	200	20	50	100	150	14
Occupancy Classifier	COM1	29	3	7	15	22	20
	RES1	225	23	56	113	169	20
	RES2	20	2	5	10	15	20

3 RESULTS

Using the pretrained BRAILS classifiers on the screened Hurricane Ida (2021) dataset produced mixed results when comparing to the manually-assigned labels. The occupancy and elevated foundation classifier predictions matched 68% and 64% of the manual labels respectively, while the roof type and number of floors predictions only matched 57% and 56% of the manual labels. The F1-scores (calculated as the weighted average F1 score across all labels) are also provided in Table 4. Disagreements between occupancy labels primarily centered around false positives, with the BRAILS occupancy classifier being overly aggressive in assigning COM labels to what were actually RES1 occupancies (see Figure 4(a)). For Elevated Foundations, a review of the disagreements between predicted and manual labels illustrates the confounding factor of breakaway walls. As illustrated in Figure xx, it can be challenging even for a human to classify whether a building is an elevated one-story building versus a two-story. Data librarians assigning manual tags are instructed to look for evidence such as exterior stairs, perimeter walls not completely surrounding the footprint of the ground floor, and location of the home (i.e., coastal buildings are typically elevated, with the ground floor consisting of breakaway walls that are not constructed to code). To a computer trained on identifying a structure constructed atop piers, it is therefore no surprise that elevated one-story homes with breakaway walls are classified as two-story, and indeed, for some applications, this may be an appropriate classification. For roof type, the pretrained BRAILS classifier matches hipped roof labels nearly 80% of the time (see Figure 4(a)) but struggled with gable and flat roofs. After reviewing the images, some of the challenges in roof shape classification may be related to the frequent use of metal roofs in coastal areas (the reflectivity of the surfaces and lack of patterns can wash out gable ridgelines), and the frequent presence of secondary roof surfaces (e.g., a roof over the front porch that ties into the main roof). Finally, for the number of floors, elevated homes again were a challenge, along with architectural dormers in one-story homes (Figure 1). The classifiers were more accurate with more regular construction.

Figure 1. Two examples of homes with a manual label of one-story but predicted labels of two story, due to (left) an elevated foundation, and (right) the presence of architectural dormers.

Given the apparent difficulties in recognizing some of the patterns in regional construction, particularly coastal regions, two of the BRAILS classifiers – specifically occupancy and roof shape – were retrained at different levels as described in Section 2.4. These two classifiers were chosen because they were the simpler of the two, and each represented one of the two input image types (aerial imagery vs street-level imagery).

Regenerating the predictions using the retrained BRAILS roof shape and occupancy classifiers showed steady improvement for occupancy with additional training data (Figure 2, Table 5), but inconsistent results for roof type (Figure 3, Table 5). Confusion matrices for both classifiers and all retraining regimens are given in Figure 4 and Figure 5. Retraining with a small proportion of the regional data resulted in fewer matches for occupancy, but a modest improvement for roof type, regardless of whether the retraining data was equally distributed between the labels (V6.0) or weighted based on the population distribution (V2.0). When 25% of the manually tagged dataset was used to retrain the classifiers, accuracy improved to over 70% for both classifiers. The best performance came with 75% of the data used for retraining. It should be noted that all of these results are deterministic, and given the relatively small sample sizes for this type of problem, a bootstrapping approach would be better suited for determining a more robust estimate of accuracy.

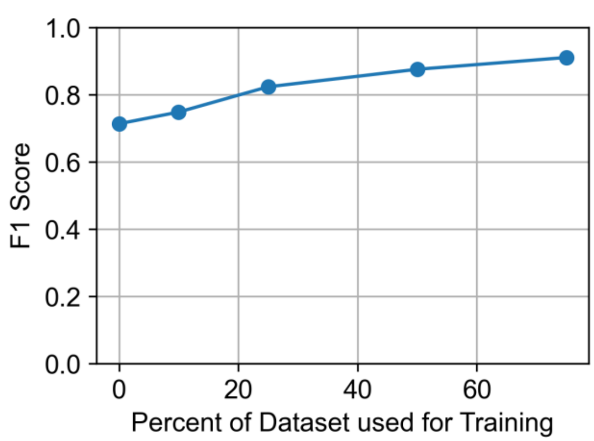
Table 4: Pretrained BRAILS classifiers accuracy and weighted average F1-Score.

	Occupancy Classifier	Roof Type Classifier	Number of Floors Detector	Foundation Classifier
N	274	438	270	262
% Correct[1]	0.67	0.56	0.52	0.695
F1-Score	0.713	0.664	0.543	0.695

[1] For purposes of this comparison, the manual labels are deemed correct, but the manual labels are also potentially in error, despite multiple checks, and some labels can be subjective.

Table 5: Effect of increasing the number of retraining images on the accuracy of the occupancy and roof type classifiers.

Version	Training Percentage	Occupancy Classifier		Roof Type Classifier	
		% Correct	F1 Score	% Correct	F1 Score
V1.0	0%	0.67	0.713	0.52	0.543
V2.0	10%	0.71	0.748	0.62	0.593
V3.0	25%	0.80	0.823	0.71	0.699
V4.0	50%	0.87	0.875	0.70	0.687
V5.0	75%	0.90	0.911	0.79	0.784
V6.0	Equal	0.7	0.789	0.63	0.638



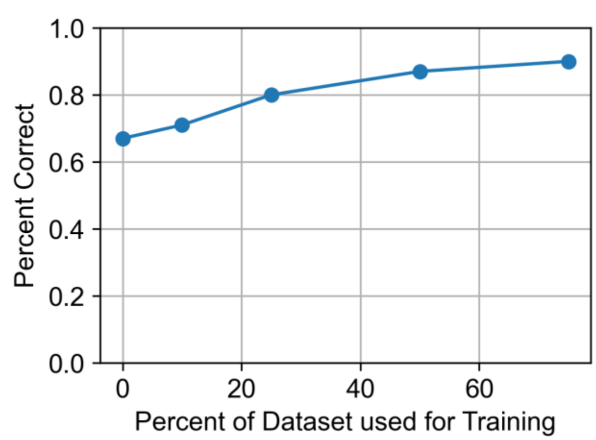


Figure 2: Changes in (left) accuracy and (right) F1-score of the Occupancy Classifier with increasing training data.

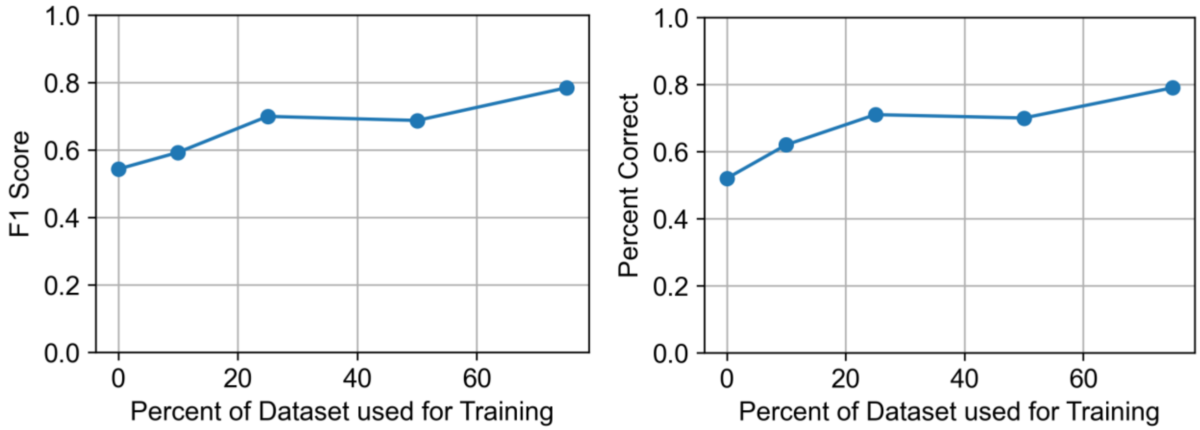


Figure 3: Changes in (left) accuracy and (right) F1-scor of the Roof Type Classifier with increasing training data.

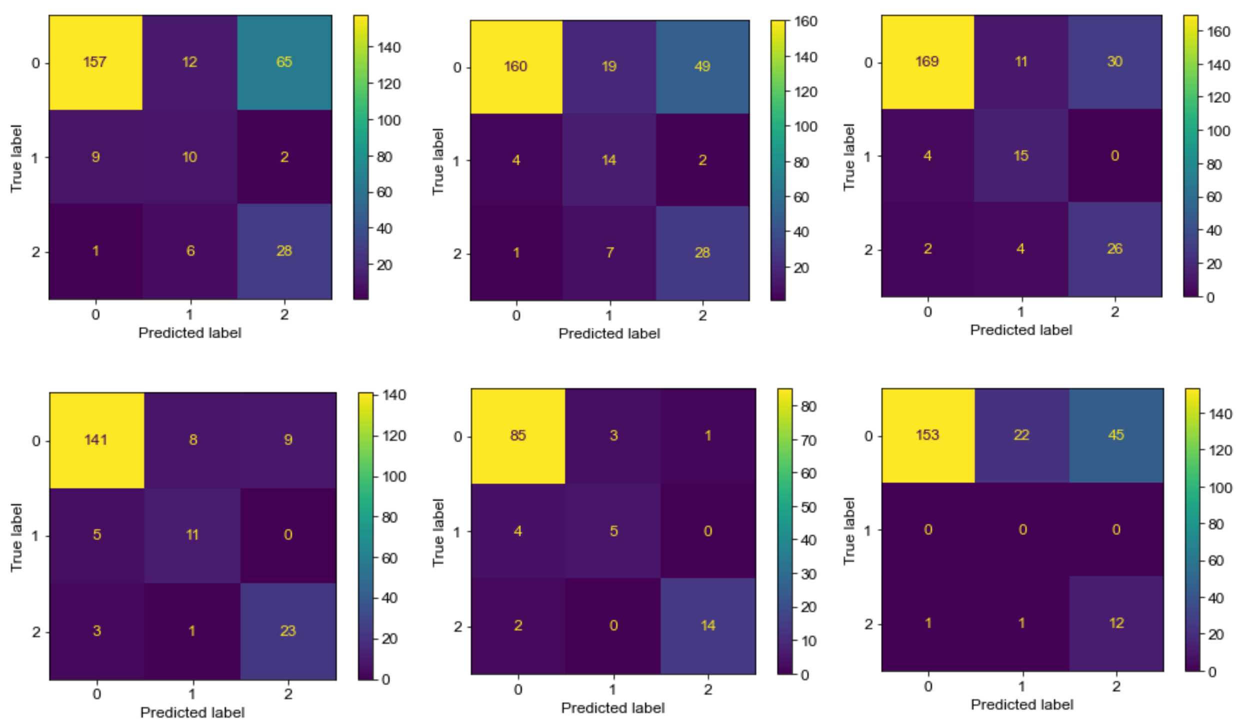


Figure 4: Confusion matrices for the Occupancy Classifier (a) Version 1, (b) Version 2, (c) Version 3, (d) Version 4, (e) Version 5, and (f) Version 6. In each confusion matrix, axis labels correspond to 0 = RES1, 1 = RES3, and 2 = COM.

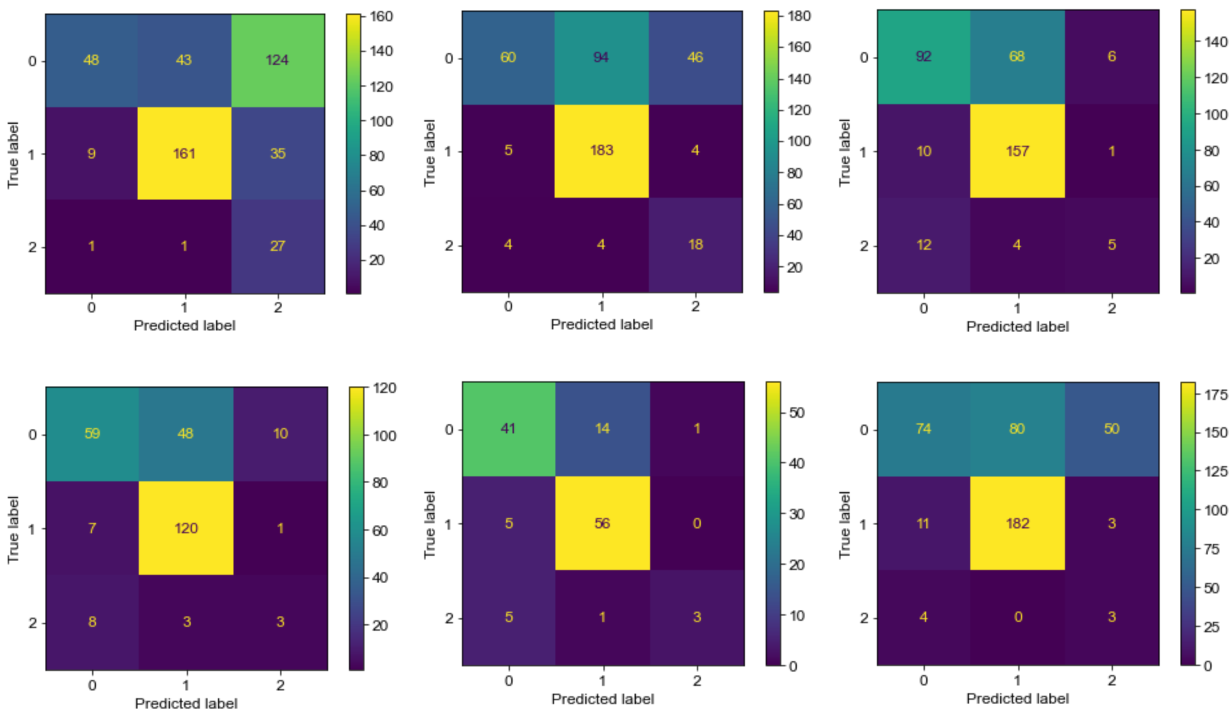


Figure 5: Confusion matrices for the Roof Type Classifier (a) Version 1, (b) Version 2, (c) Version 3, (d) Version 4, (e) Version 5, and (f) Version 6. In each confusion matrix, axis labels correspond to 0 = Gable, 1 = Hip, and 2 = Flat.

4 DISCUSSION AND CONCLUSIONS

In summary, this study describes a hybridized reconnaissance workflow used following Hurricanes Laura (2020) and Ida (2021) that integrated a variety of data sources within a cloud-based environment to conduct virtual performance assessments on individual buildings. Post-storm, surface-level imagery was essential to accurately quantifying component-level damage, although multiple data sources were typically relied upon to overcome limitations of each of the data sources related to resolution, coverage, and timing of data acquisition. On average, the virtual performance assessments took between 10 and 15 minutes to complete for each structure, depending on how many building attributes were quantified and the quality of the data available for each record.

There is potential to integrate machine-in-the-loop processes into the workflow, as demonstrated in this study by the use of the BRAILS tool developed by the NHERI SimCenter. However, the accuracy of the pretrained BRAILS classifiers on "in the wild" target structures for the Hurricane Laura and Hurricane Ida dataset were still significantly lower (on the order of 20%) than those found by Wang et al. (2021) using larger, more regular datasets. This suggests that tuning to local construction may be necessary for future applications, but the impacts of retraining to local construction in this study were mixed and showed high variability. The variability is possibly related to the relatively small sample sizes employed, but there are also systematic challenges that may not be resolvable by simply using more data. For example, proper classification of elevated one-story buildings with breakaway walls is likely to remain a challenge. There is a need for continued evaluations of ideal retraining datasets, and the optimal distribution of classes within these retraining datasets, to better tune pretrained models to local practices and in turn produce better accuracies.

5 REFERENCES

Berman, J. W., Wartman, J., Olsen, M., Irish, J. L., Miles, S. B., Tanner, T., Gurley, K., Lowes, L., Bostrom, A., and Dafni, J. (2020). "Natural hazards reconnaissance with the NHERI RAPID facility." *Frontiers in Built Environment*, 185.

Mateusz Buda, Atsuto Maki, and Maciej A Mazurowski. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106:249–259, 2018.

Gurley, K. R., and Masters, F. J. (2011). "Post-2004 Hurricane Field Survey of Residential Building Performance." *Natural Hazards Review*, 12(4), 177-183.

Kijewski-Correa, T., Smith, N., Taflanidis, A., Kennedy, A., Liu, C., Krusche, M., and Vardeman, C. (2014). "CyberEye: Development of integrated cyber-infrastructure to support rapid hurricane risk assessment." *Journal of Wind Engineering and Industrial Aerodynamics*, 133, 211-224.

Kijewski-Correa, T. L., Roueche, D. B., Mosalam, K. M., Prevatt, D. O., and Robertson, I. N. (2021). "StEER: A Community-centered Approach to Assessing the Performance of the Built Environment after Natural Hazard Events." *Frontiers of Built Environment*, 7.

Massarra, C. C., Friedland, C. J., Marx, B. D., and Dietrich, J. C. (2019). "Predictive multi-hazard hurricane data-based fragility model for residential homes." *Coastal Engineering*, 151, 10-21.

Rathje, E. M., Dawson, C., Padgett, J. E., Pinelli, J.-P., Stanzione, D., Adair, A., Arduino, P., Brandenberg, S. J., Cockerill, T., Dey, C., Esteva, M., Haan, F. L., Hanlon, M., Kareem, A., Lowes, L., Mock, S., and Mosqueda, G. (2017). "DesignSafe: New Cyberinfrastructure for Natural Hazards Engineering." 18(3), 06017001.

Roueche, D. B., Lombardo, F. T., Smith, D. J., and Krupar, R. J. (2018). "Fragility Assessment of Wind-Induced Residential Building Damage Caused by Hurricane Harvey, 2017." *Forensic Engineering 2018*, 1039-1048.

Cetiner, B. (2020). "Object Detection Model to Determine Number of Floors of Buildings (1.0)." Zenodo. https://doi.org/10.5281/zenodo.4421613

NHERI SimCenter. (2022a). "BRAILS." https://github.com/NHERI-SimCenter/BRAILS

NHERI SimCenter. (2022b). "BRAILS." https://nheri-simcenter.github.io/BRAILS-Documentation/

Sascha, H. (2020). "Foundation Detection Model (0.2). " Zenodo. https://doi.org/10.5281/zenodo.4145934

Prevatt, D. Kameshwar, S. Roueche, D. Rittelmeyer, B. Duarte, T. Heo, T. Ibrahim, H. Klepac, S. Lafontaine, O. Lin, T. Manuel, L. Pilkington, S. Pinyochotiwong, Y. Santiago-Hernandez, J. Strader, S. Gurley, K. Kijewski-Correa, T. Mosalam, K. Robertson, I. (2021) "StEER: Hurricane Ida Joint Preliminary Virtual Reconnaissance Report-Early Access Reconnaissance Report (PVRR-EARR)", in StEER - 29 August 2021, Hurricane Ida. DesignSafe-CI. https://doi.org/10.17603/ds2-w6km-fe51.

Wang, C. (2020). "Roof Type Image Classifier (v0.1)." Zenodo. https://doi.org/10.5281/zenodo.4567613

Wang, C. (2021). "Building Occupancy Type Classification Model (v0.2)." Zenodo. https://doi.org/10.5281/zenodo.4795548