



white paper series: machine learning

# CUTTING THE CODE

*Accessible machine learning*





# EXECUTIVE SUMMARY

This paper investigates the use of low-code machine learning tools for determining the severity of structural damage following natural disasters. The proposed model determines the relevance for damage assessment of images posted to Twitter and then provides damage severity levels to images taken in the aftermath of 2021’s Hurricane Ida in Louisiana. A low-code framework for damage assessment through geo-located social media imagery provides viable training data points for machine learning algorithms. It lowers the barriers to entry into the field of artificial intelligence for disaster preparedness professionals and has the potential to accelerate recovery in regions with limited resources.

## 3.1 Introduction

Two anthropogenic forces, the expanding volume of data generated through social media and the increasing severity and frequency of natural disasters drive the convergence of the fields of artificial intelligence and natural disaster recovery.<sup>1</sup> On average, Twitter users post half a billion Tweets every day.<sup>2</sup> A growing global population and increasing anthropogenic atmospheric warming accelerate the number, severity, and frequency of extreme weather events and “natural” disasters.<sup>3</sup> However, after natural disasters occur, the generation of social media posts can spike. On August 30th, 2017, after Hurricane Harvey made landfall in South Texas, 2 million tweets were generated containing the keywords “Hurricane Harvey,” “Harvey,” or “HurricaneHarvey.”<sup>4</sup> Increases in global populations and adoption of smartphone technologies will result in a higher volume of disaster-related crowd-sourced data. This data has the potential to be an important resource to further the understanding of natural disaster recovery.

Consequently, data generated from social media is perishable; the contextual details that accompany an image posted to Twitter become less relevant as time progresses. Therefore, it is necessary to rapidly analyze this data to ensure its useability in disaster recovery.

Field-tested research suggests integrating hazard characteristics, community exposure and vulnerability, and social media information into rapid damage assessment processes.<sup>5</sup> For emergency managers, community organization leaders, and damage assessors, identifying areas requiring higher support levels and properly assessing damaged structures helps speed up informed decision-making and recovery.<sup>6</sup>

Regardless of location, anyone can post images to social media after a natural disaster. A significant challenge when using images sourced from social media for decision-making purposes is validating if the datapoint is accurate and authentic.<sup>7</sup> Geolocated Tweets within an affected region in post-disaster communities offer a higher level of validity and reliability, allowing researchers to use imagery data for accelerating recovery efforts. However, an estimated

1% of Twitter users have precise location tracking turned on, which is necessary to identify locations of damaged structures and accelerate recovery efforts.<sup>8</sup> Further, of the 6.7 million tweets collected by Alam et al. from Hurricane Harvey, 115,525 (1.7%) contained an image, with only 1,155 (0.17%) of them having precise geographic location necessary to identify areas in need.<sup>9</sup>

Analyzing large image datasets requires powerful algorithms to automate the analytical process. Artificial intelligence and machine learning techniques can be applied to images taken in post-disaster communities to help prioritize areas in need of higher levels of assistance.<sup>10</sup> Preliminary damage assessment through image classification of damaged buildings can help prioritize distribution of resources following disasters.

Typically, innovative uses of AI are spearheaded by computer science experts with programming knowledge capable of creating models. However, for planners, emergency managers, and first responders, taking advantage of state-of-the-art AI and machine

learning techniques often has a steep learning curve. More recently, machine learning tools are becoming accessible to broader audiences due to tech start-ups specializing in low-code modeling interfaces. Lowering the barriers to machine learning increases opportunities for the use of models.

One example of a low-code machine learning platform is Lobe.ai. Through Lobe.ai, users can annotate whole images by filling out a text box and training machine learning models with the click of a button. This study uses Lobe.ai and open-source imagery datasets to train machine learning models for damage assessment. Additionally, a field experiment in Louisiana, where participants generate imagery data of structural damages in the aftermath of Hurricane Ida, provides a testing dataset for the model. This study aims to provide disaster recovery professionals with an accessible model of how the use of artificial intelligence for damage assessment through social media imagery can accelerate the recovery of affected communities.

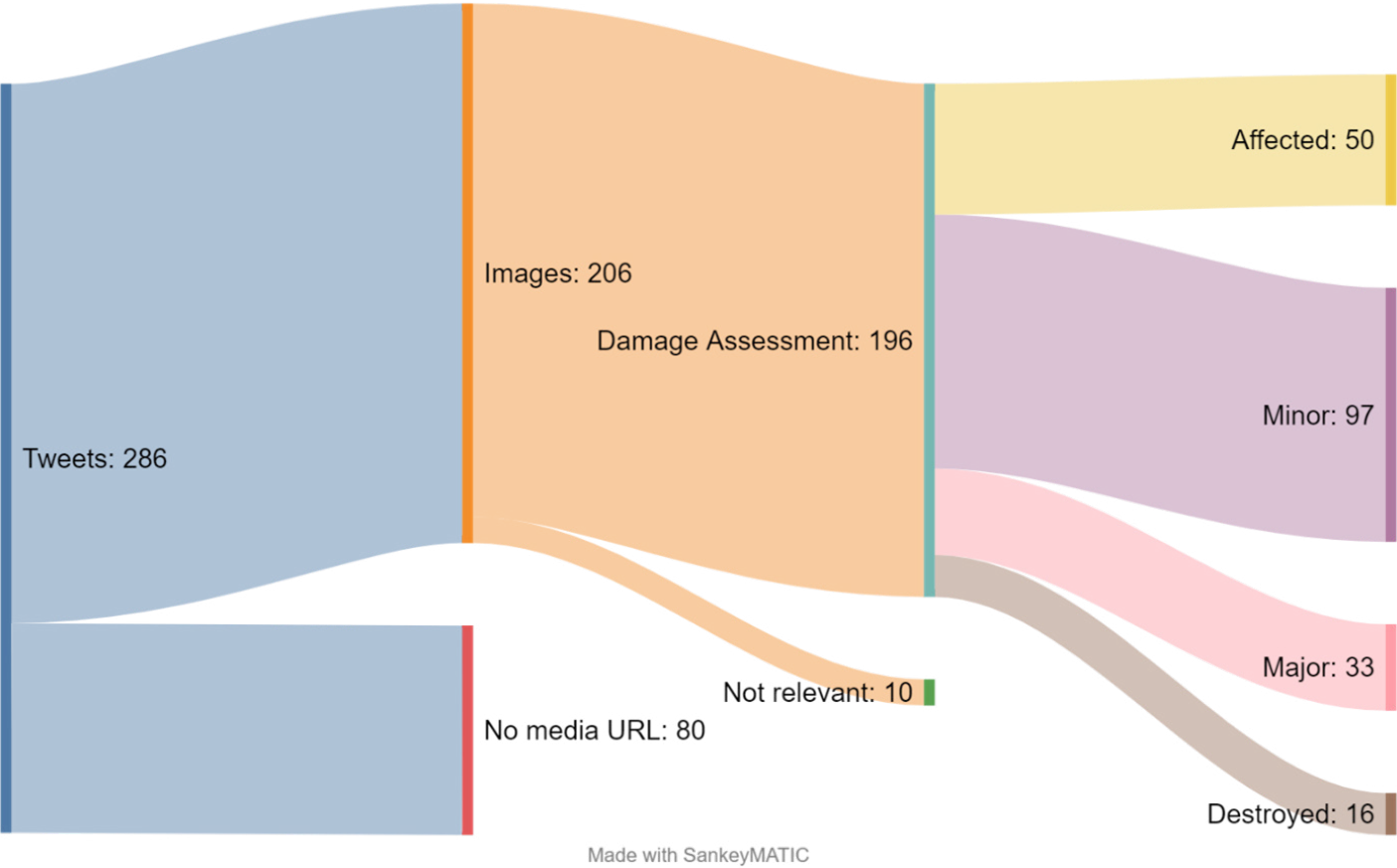


Figure 3.1 – Image classification of sequential machine learning models

3.2 Datasets

This model uses three datasets from recent natural disasters.

**1. Hurricane Matthew (2016):** 407 images from Hurricane Matthew that struck Haiti in October of 2016. Researchers from the Qatar Computing Research Institute collected images from Twitter and host disaster related datasets for public use on their Crisis Natural Language Processing (CrisisNLP) data-portal.

**2. Typhoon Ruby (2014), Nepal Earthquake (2015), Ecuador Earthquake (2016), and Hurricane Matthew (2016):** 662 images selected as being relevant to machine learning for damage assessment from CrisisNLP datasets.

**3. Hurricane Ida (2021):** 216 images collected via Twitter by the University of Michigan “Rising Above the Deluge” Urban Planning Masters Capstone team.

3.3 Classification

This method for categorizing damage assessment of imagery from social media posts requires two sequential machine learning models.

Filtering Images for Damage Assessment

The first model categorizes images based on their relevancy to damage assessment (Assessment, Non-relevant). If a building or partial building is included in the image, it is deemed relevant for damage assessment. *See Image 3.1 for examples.*

Assessing Damages

The second model classifies damage severity according to FEMA’s Preliminary Damage Assessment Guide (Affected, Minor, Major, Destroyed).<sup>11</sup>

- Affected:** no damage, aesthetic damage
- Minor:** non-structural damage, loose siding/roofing
- Major:** structural damage to roof or walls
- Destroyed:** no structure remains, imminent threat of collapse

Selecting FEMA’s damage classification scale in this

machine learning model provides annotations with rigid guidelines for annotation protocol. Additionally, results from this machine learning model can be given to damage assessment professionals without the need for translating scales of damage.

3.4 Methods

In this section, we describe the collection of relevant data and categorical assessment of damages from natural disasters through minimal-code machine learning models.

Generate and Retrieve Twitter Data

Our team, consisting of graduate students from the University of Michigan, collected data related to structural damage and conducted community outreach interviews in New Orleans, Louisiana, and surrounding parishes from February 16th to February



Image 3.1: Images classified by Lobe.ai as Assessment and NR (not relevant) for machine learning.



19th, 2022. Although it was a limited purposeful sample, it allowed us to understand the salience of using this technology under the most ideal of circumstances.

Ten student participants split into two study groups with instructions to capture images on smartphones and post images to Twitter of damages to infrastructure caused by Hurricane Ida. Instructions for one group include turning on precise location tracking on their smartphones and through the mobile Twitter app. Other participants were instructed to turn off locational data on Twitter. Participants were provided with disaster-related keywords from CrisisLex to assist with crafting tweet texts. Still, they were given free rein to post Tweets as if they were pursuing damage assessment following a natural disaster event.<sup>12</sup> Unknown to the reseachers for this project, participants deliberately posted tweets with misleading information and unrelated imagery to better mimic social media during natural disasters.

Additionally, participants were asked to post 50% of their tweets with the hashtag #MURP\_Deluge. The inclusion of specific hashtags in disaster recovery social media posts significantly improves researchers’ ability to identify and collect relevant imagery.

Using Twitter’s API platform and filtering by participants’ unique usernames, we identified 286 tweets posted to Twitter during the experiment, and 216 contained images (Figure 3.1). To better simulate an actual disaster event, researchers must assume no knowledge of an individual’s username. By filtering for the hashtag #MURP\_Deluge, researchers were able to identify all 169 tweets containing images. However, only 4 of the remaining 117 tweets (3.4%) were able to be identified using selected disaster verbiage (“Assessing Damage,” “Buildings damaged,” “Nothing Left,” “In bad shape”).

Use of Low-Code Machine Learning

Lobe.ai simultaneously trains with two machine learning algorithms to improve the model’s speed and accuracy (MobileNetV2 and Resnet-50V2, respectively). Developing a model begins with uploading a training dataset and labeling images via image classification (Image 3.2). Image augmentation includes adjustments to brightness, contrast, saturation, hue, rotation, zoom, and noise of images. Image augmentation alters existing data, providing additional inputs for training the model and increasing the

Crisis NLP Dataset: Social Media Images for Damage Assessment			
Event	Total Images	Assessment Images	%
Ecuador	1438	204	14%
Matthew	596	278	47%
Nepal	18456	108	1%
Ruby	833	90	11%

Table 3.1: Percentage of social media images relevant to damage assessment under FEMA’s PDA. Dataset: 1 – Hurricane Harvey

Lobe.ai Damage Assessment: Hurrican Ida (Testing Data)			
Damage Type	Count	Correct	%
Affected	50	33	66%
Minor	97	72	74%
Major	33	11	33%
Destroyed	16	9	56%
Total	196	125	64%

Table 3.3: Accuracy from manual calculations as testing images are uploaded. Dataset: 3 – University of Michigan, Hurricane Ida

This study tracks image classification accuracy manually in a spreadsheet to understand how accurately new imagery is classified when introduced into a model. Manual tracking is necessary because Lobe.ai provides the accuracy of the entire dataset and individual classes after the model trains through several iterations. This step analyzes how low-code machine learning models perform under real-world testing conditions.

3.5 Results

Of the 216 images collected from the Twitter experiment, 196 contained relevant images for damage assessment (91%). The percentage of relevant images is exceptionally high compared to real-world datasets (Table 3.1). A higher rate of relevant images is likely due to the idealistic nature of the field visit.

If location is turned on when posting tweets, the geo coordinates associated with the tweet can be retrieved by Twitter API. During the Louisiana field experiment locational data is available through Twitter’s API on 37 of the 216 total images (17%). This result shows a 100-fold increase in social media posts with images and precise location coordinates included in the tweet data compared to the estimate for Hurricane Harvey (0.17%). While this experiment is likely an idealized scenario for locational data collection, it provides strong evidence for increasing two-way communication on social media platforms.

Increasing two-way communication through social media platforms can increase public understanding of the importance of data in disaster recovery. If in

likelihood that a new image will be classified correctly.

Lobe.ai, currently in its beta phase of development, only allows testing of one image at a time. After the model guesses the classification of the uploaded image, the user confirms if the model made a correct assumption. After confirmation, the image is placed in the training dataset. Lobe.ai continuously runs and updates the model throughout the annotation process as more images are added.

Manual Tracking of Testing Images

Lobe.ai Damage Assessment: Model Output		
Damage Type	Images	Correct
Affected	337	96%
Minor	303	95%
Major	142	89%
Total	858	93%

Table 3.2: Accuracy from Lobe.ai Damage Assessment Machine Learning Model. Datasets: 2 and 3 – CrisisNLP and University of Michigan

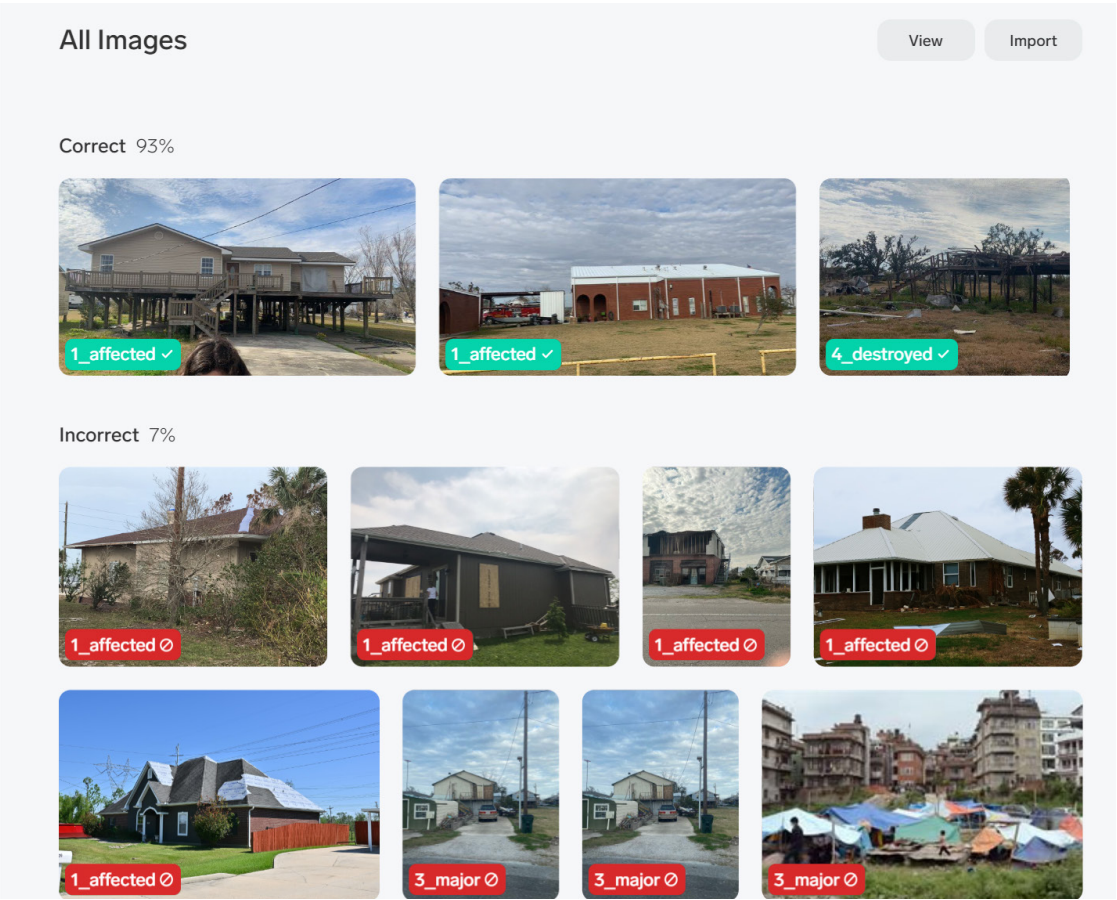


Image 3.2: Lobe.ai’s training interface for the damage assessment model.

the days leading up to Hurricane Harvey, people were given instructions on how to turn on precise location data and use event-specific hashtags for requesting aid or assessment, a significant quantity of images would be available to researchers of damage assessment. Generation of data at scale provides the necessary data inputs to train machine learning algorithms accurately for extremely low upfront costs. Machine learning algorithms become more accurate as more training data is made available.

The damage assessment model on Lobe.ai states that 93% of the images in the entire dataset (858) are predicted correctly (Table 3.2). However, manual tracking of model performance during individual image uploads from testing data shows a lower overall model accuracy at 64% (Table 3.3). The ~30% difference in accuracy results is likely due to Lobe.ai and its ability to retrain the model in real-time. However, real-world use mimics the results shown by the manual output accuracy in Table 3.3. Field use of image classification and machine learning in damage assessment would not be able to retrain a model until well after the preliminary damage assessment is conducted.

The accuracy of the current Lobe.ai damage assessment model is currently too low for utilization in a pilot study on preliminary damage assessment. The risk for miss-classification of damaged structures assessed as “Major” (33%) and “Destroyed” (56%) is high, to the point where recovery operations would be further inhibited using this tool. Higher quantities of image data are required to increase the accuracy of multi-class machine learning models.

### 3.6 Discussion

Findings from this study indicate that the current state of no-code machine learning platforms is achieving their goals of increasing accessibility. However, scalable use-phase implementation of no-code models remains uncertain. Further iterations of this study should export Lobe.ai models to no-code apps, such as Microsoft’s Power Platform, to collect and test larger quantities of images. One-thousand images from each class should be included in the training dataset to better address the accuracy concerns with

damage assessment models.

Collecting images from site visits is a massive time commitment by community organizations and disaster recovery agencies. The use of social media platforms as a means for data collection might provide the quantity of data required to more accurately train machine learning algorithms and speed up disaster recovery.

### 3.7 Conclusion

The current frameworks in place for disaster recovery and damage assessment cannot adapt to increasingly volatile global weather events. As a higher share of the global population gains access to smartphone technology and interconnectivity through social media platforms, government agencies must change how they collect valid data points. Machine learning and disaster recovery are on a collision course for implementation in real-world scenarios. How we construct new tools for expediting preliminary damage assessment processes and which communities are included in their creation will affect survivors of natural disasters over the coming decades.

Using social media as a platform for two-way communication for disaster preparedness can drastically increase the amount of valid and relevant data points for image collection. Images from a single natural disaster event can assist in decision-making for the prioritization of aid and can be made available to researchers for improving machine learning models.

Low code tools have a significant role in the future of disaster recovery operations. However, two impediments to adoption remain; first, higher levels of public understanding of machine learning will be required to gain acceptance. Secondly, increasing the accuracy of low code models to an acceptable level requires additional data points. Low code tools provide the means to solve complex problems intuitively by

lowering the technical barriers to entry into machine learning. Without the use of artificial intelligence in disaster recovery, the billions of data points generated through social media remain inaccessible.

### 3.8 Resources for Next Steps

Research teams looking to replicate this paper’s image collection method can use Twitter’s API and the Tweepy Python library to parse image URLs from tweets.

[Twitter API documentation](#)  
[Tweepy documentation](#)

A researcher with introductory-level programming knowledge will be able to search tweets by hashtags, sort by location, and determine if images are present. Upon creating a dataframe of individual tweets, image URLs can be parsed from the media entity of each tweet. If the quantity of images is low, saving unique images from a web browser is also a viable option.

[Tutorial for gathering images from Twitter](#)

Machine learning models created on Lobe.ai can be exported to no-code apps, such as Microsoft’s Power Platform, or as Python-based notebooks as Tensorflow.

[Integrating Lobe.ai and Microsoft Build](#)

# ENDNOTES

---

1 Kontokosta, C. E. (2021). Urban Informatics in the Science and Practice of Planning. Journal of Planning Education and Research, 41(4), 382–395. <https://doi.org/10.1177/0739456X18793716>

2 Omnicore. (2022, February 22). Twitter by the Numbers (2022): Stats, Demographics & Fun Facts. <https://www.omnicoreagency.com/twitter-statistics/>

3 Bhaskara, G. I., Filimonau, V., Wijaya, N. M. S., & Suryasih, I. A. (2021). The future of tourism in light of increasing natural disasters. Journal of Tourism Futures, 7(2), 174–178. <https://doi.org/10.1108/JTF-10-2019-0107>

4 Alam, F., Ofli, F., Imran, M., & Aupetit, M. (2018). A Twitter Tale of Three Hurricanes: Harvey, Irma, and Maria. ArXiv:1805.05144 [Cs]. <http://arxiv.org/abs/1805.05144>

5 Chen, Y., & Ji, W. (2021). Rapid Damage Assessment Following Natural Disasters through Information Integration. Natural Hazards Review, 22(4), 04021043. [https://doi.org/10.1061/\(ASCE\)NH.1527-6996.0000504](https://doi.org/10.1061/(ASCE)NH.1527-6996.0000504)

6 Chen, Y., & Ji, W. (2021). Rapid Damage Assessment Following Natural Disasters through Information Integration. Natural Hazards Review, 22(4), 04021043. [https://doi.org/10.1061/\(ASCE\)NH.1527-6996.0000504](https://doi.org/10.1061/(ASCE)NH.1527-6996.0000504)

7 Tapia, A. H., Bajpai, K., Jansen, B. J., & Yen, J. (2011). Seeking the Trustworthy Tweet: Can Microblogged Data Fit the Information Needs of Disaster Response and Humanitarian Relief Organizations. 10.

8 Rowe, W. (2019, February 15). How To Track Tweets by Geographic Location. BMC Blogs. <https://www.bmc.com/blogs/track-tweets-location/>

9 Alam, F., Ofli, F., Imran, M., & Aupetit, M. (2018). A Twitter Tale of Three Hurricanes: Harvey, Irma, and Maria. ArXiv:1805.05144 [Cs]. <http://arxiv.org/abs/1805.05144>

10 Nguyen, D. T., Ofli, F., Imran, M., & Mitra, P. (2017). Damage Assessment from Social Media Imagery Data During Disasters. Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, 569–576. <https://doi.org/10.1145/3110025.3110109>  
11FEMA. (2020). Preliminary Damage Assessment Guide (p. 127)

12 Temnikova, I., & Castillo, C. (2015). EMTerms 1.0: A Terminological Resource for Crisis Tweets. 13.



---

## CUTTING THE CODE: ACCESSIBLE MACHINE LEARNING

Authors: Jeffrey Pritchard, Danielle Stewart, Kiley Fitzgerald

---

### about this project

This project is a joint effort by students and faculty within the Master of Urban and Regional Planning program at the University of Michigan and the National Disaster Preparedness Training Center (NDPTC) as a Capstone project for the Winter 2022 semester.

A key focus of the University of Michigan team is to work in a manner that promotes the values of equity, uplifting local voices, transparency and honesty. As a result, the outcomes of this capstone aim to speak to both our collaborators at the NDPTC and the local communities impacted by disasters across the United States. Our responsibilities as researchers will also include the implementation and/or recommendation of innovative solutions to issues surrounding machine learning, damage assessments, prioritization determinations, and social infrastructure networks.

