# Comparative Assessment of Building Damage from High-Resolution Satellite-Based Images

Erica Scaduto (escaduto@ucdavis.edu), Yuhan Huang (yuhhuang@ucdavis.edu)

## I. Introduction

The automatic extraction and evaluation of building damage by natural hazards can aid in assessing risk management and mapping distributions of urban vulnerability [1, 2, 3]. Although wildfire is a common and critical natural disaster posing significant threats, constrained by the methods and data quality, most previous studies only focused on large-scale disasters. Deriving a reliable and efficient building extraction and damage classification method has presented challenges due to regional differences in development type (e.g. rural vs metropolitan), as well as the sheer varieties in building characteristics (i.e. color, shape, materials). In this project, we intend to compare different machine learning algorithms for image classification to develop a general framework for fire-induced building damage evaluation from high-resolution remotely sensed imagery in California.

Our results show promising utility for machine learning models, based on spectral, texture, and convolutional features, in the applications for post-fire building-damage monitoring. For the binary classification scheme, the Random Forest (RF) classifier performed the best (Table II) with an overall accuracy of 93% and a Cohen's kappa of 0.73. For the multiclass scheme (i.e without vegetation mask), XGBoost performed better than the 5-layer neural networks. It is able to detect building areas but is less accurate in predicting damage when compared with the binary case. Feature engineering also proved to be an essential step in model building. Particularly, the addition of SNIC segmentation which greatly aided in the improvement of overall model performance for both RF and XGBoost classifiers.

## II. Methodology

### 2.1 Algorithms

Unsupervised clusters were extracted through k-means and Learning Vector Quantization (LVQ) algorithms. K-means clustering partitions the dataset into k groups by iteratively refining the clusters based on proximity and cluster mean. Due to its local search procedure, the k-means algorithm can have drawbacks mainly affected by the initial starting conditions. Developed by Kohonen, LVQ applied the winner-take-all Hebbian learning-based approach and consists of the competitive layer (i.e. sub-classes), weights of each neuron, and learning rate [4]. Two other cluster-based segmentation algorithms were also applied to the images, the Simple Linear Iterative Clustering (SLIC) and Simple Non-Iterative Clustering (SNIC). The SLIC segmentation algorithm performs a localized k-means optimization that clusters pixels based on color and a weighted Euclidean distance [5]. SNIC is a non-iterative version that aims to address some of the underlying limitations of SLIC. The simplification allows for explicit connectivity, cheaper computation, and less memory [5].

Supervised classification models in this project include SVM, tree-based models, and neural networks. SVM with radial basis function (rbf) kernel is isotropic, infinitely smooth, and invariant to translation of input data. Tree-based machine learning algorithms, especially ensembled methods such as random forest and gradient boosting, integrate weak learners to form a strong model to promote diversity of model simulations and flexibly handle nonlinear relationships. We also built multi-layer neural networks which can learn rich representation from large datasets. Activation functions used to establish the network structure include sigmoid (similar to smoothed Heaviside and returns small value close to 0), relu (returns the max between 0 and the input), and selu (Scaled Exponential Linear Unit). The selu function uses two predefined parameters to scale the loss and returns internal normalized values [6].

### 2.2 Dataset

*2.2.1 xView 2 Building Damage Assessment*

The xView 2 Building Damage Assessment dataset provides annotated building footprint polygons with 5 damage classes (minor-damage, major-damage, destroyed, no-damage, un-classified) with corresponding RGB images at a resolution below 0.8 meter ground sample distance (GSD). The original dataset provides 850,000 buildings over six different natural disasters across 15 countries. The imagery for the xView dataset was sourced from Maxar/DigitalGlobe Open Data Program [7]. Building footprints were extracted and annotated via a web-based annotation tool coupled with several review and validation stages. For the purposes of this study, we will focus on the geographic annotations from fire damages within California. Specifically, the building damages from Carr fire in Santa Rosa from Jul 23 to Aug 30, 2018 and Woolsey fire in SoCal from Nov 9 to 28, 2018 (A1). There were 27,254 no-damage and 4,604 damaged/destroyed buildings for a total of 31,858 buildings footprints throughout the study region.

*2.2.2 NAIP Imagery*

The National Agriculture Imagery Program (NAIP) satellite images are acquired at 0.6-1m resolution GSD with Red, Green, Blue (RGB) and near-infrared band (NIR). Since 2009, NAIP imagery has been acquired every three years. Based on fire events instances focused in this study, NAIP images were obtained for pre-fire condition in 2015 and post-fire conditions in 2018. Acquisition and processing of NAIP imagery were completed using the GEE python API via the Colab notebook (A1).

## III. Implementation

The project (A2) was conducted on Google Colab using GEE API, sklearn, and tensorflow. Image acquisition, unsupervised features extraction, and supervised classification with decision tree, SVM, and random forest were run at a server-end through GEE API. XGBoost and neural network were run locally on Tesla p100 GPU.

### 3.1 Geo-processing

The annotated xView JSON data attributes include: building footprint geolocation as WKT, disaster type, location name, unique event ID, damage ranking, and image metadata. The original datasets were converted to geographical features as geojson and shapefile. These formats with referenced spatial attributes allow readability in GEE. The original classes ‘minor-damage’, and ‘major-damage’, were combined into the ‘destroyed’ category due to unequal representation. Files were further hosted in a Google Cloud bucket to increase accessibility and loaded as assets into GEE as `XVIEW_training` and `XVIEW_testing`. Finally, using the extents from the dataset, a bounding box was created to extract the NAIP imagery for pre- and post-fire conditions.

### 3.2 Feature Engineering

To inform the final classification outputs, a number of indices were derived taking advantage of NAIP’s RGB & NIR bands, along with object-based layers including segmentation and geometric feature extractions (e.g. canny edge detection, compactness, etc.). Both pre- and post-fire data were selected as inputs in the model to inform image changes of damaged areas.

*3.2.1 Spectral*

Unique spectral properties of fire induced reflectance can be observed in the visible, near-infrared, and mid-infrared ranges [8]. Although channels in the visible spectral range (0.4μm - 0.7μm) are not particularly best at discriminating burned surfaces, recent burns and charred surfaces generally have lower reflectances. Likewise, in contrast to healthy vegetative cover that has high reflexivity in the NIR (0.8μm–0.9μm), there will be lower reflectance in this region. The Normalized Burn Ratio (NBR) detection algorithm makes use of the differential response between NIR and the higher reflectance seen in the SWIR (1.5μm–1.6μm).

*3.2.2 Indices*

Indices are widely used in remote sensing applications for spectral enhancement and quantitative evaluation of environmental conditions. There are a broad range of applications, from identifying soils to assessing the health of vegetation. For example, the Normalized Difference Vegetation index (NDVI) algorithm takes advantage of the differential properties between NIR and Red bands as a proxy to measure

vegetative health. As such, values close to 1 are indicative of healthy, dense vegetation related to high photosynthetic capacity (Table I). For the purposes of this study, a vegetation mask was derived using thresholds of NDVI values (>0.1) to further extract land cover with only urban and bare soil. A water mask was also applied using Hansen's Global Forest Change (2015) product [9].

*3.2.3 Texture*

Image texture can be represented as the spatial relationship of pixels. The gray-level co-occurrence matrix (GLCM) calculates frequency of pixel pairs and can be used to extract second-order statistical measurements from each band of the image. In this study, through visual examination, variance (spread-out of grayscale values), inverse difference moment (image homogeneity), dissimilarity, and cluster shade were calculated for each NAIP band [10]. Local contrast and correlation between pixel pairs were also calculated for red and NIR bands. Moreover, we also calculated local associations of neighborhoods as quantified by Geary's C for the NIR band.

*3.2.4 Convolution Filters*

Convolutions can remove white noises and enhance emboss or edges. We calculated a set of different convolutional filters including Gaussian low-pass filter (radius=7, sigma=2, blur the image), shape-sensitive filter (octagon, radius=5), Manhattan filter based on rectilinear (city-block) distance (radius=4), and Canny edge detection filters (gradient magnitude=50, sigma=1) to highlight different characteristics of buildings. All convolutional kernel sizes were tuned with best visual quality.

*3.2.5 Clustering*

Unsupervised clustering algorithms including k-means and LVQ were used as additional layers to inform pixel similarity. We trained the wekaKMeans clusterer with sampled 5000 pixels where k=15 number of clusters. The LVQ clustering algorithm was also executed with similar criteria as k-means.

*3.2.6 Segmentation*

To further extract meaningful representation of the objects in the image, two segmentation techniques were implemented including the Simple Linear Iterative Clustering (SLIC) and Simple Non-Iterative Clustering (SNIC). Both methods generate superpixels starting from seeds on equally spaced grids. The iterative SLIC segmentation algorithm is based on the KMeans clustering where we designated numClusters=15, numIterations=1000, neighborhoodSize==20, with forced convergence and unique labels set to False. The SNIC layer was created with the following parameters including: size of the superpixel seed spacing=30 pixels, compactness=15, connectivity=8, neighborhood=200, and seeds generated by the Segmentation.seedGrid package. Note that the compactness factor allows for spatial distance weighting where larger values cause clusters to be more compact.

**3.3 Pixel-Based Classification**

Several pixel-based classification algorithms including CART, Support Vector Machine (SVM), and Random Forest (RF) were used to classify the NAIP imagery into two categories: 'No-Damage' and 'Destroyed'. We also operated multiclass (non-building, no-damage, and destroyed) classification at a scale of 1 meter pixel resolution on the whole image without making through XGBoost and deep neural networks (both 2-layer and 5-layer). Hyperparameters of XGBoost were tuned through grid search with 5-fold validation. The best performing model is a 500-tree ensemble with the depth of 10, a learning rate of 1, and a weight of 0.5 for trees being boosted (i.e., min_child_weight). Regularization terms (lambda and alpha) and subsampling ratios (subsample and column sample by tree) were left with default settings of 0 and 1, respectively. The simple neural network for multiclass was established with ADAM optimizer and tuned on mini-batches with 201 epochs. The 2-layer network includes a 15-node sigmoid layer and a 5-node relu layer. The 5-layer neural network has the first two and the fourth relu layers with 1024, 512, 10 hidden units respectively, the third selu layer with 32 units, and the fifth sigmoid layer with 5 units.

**IV. Results**

For the binary classification with mask, random forest classifier performs the best (Table II) with an overall accuracy of 93% and a Cohen's kappa of 0.73. Producer and user accuracies of each class show

higher accuracy for no-damage areas relative to destroyed building footprints (Table III). Overall, a higher percentage of false negatives are seen with omission errors for destroyed class at 31% , while only 2% were observed in the no-damage class. By manual inspection, we observe that this is particularly the case for instances of building footprints with varying degrees of minor to moderate damage, i.e. structures that are not completely destroyed and still intact. Examples of classification results can be seen for three main areas that were greatly affected by fire-induced destruction in the Santa Rosa region (A3). In order to consider the effects of scaling during the validation step, we also ran accuracy assessments across varying scales (5 to 120 pixels) (A5). Overall, important feature layers that contributed to the RF classifications include SNIC, bare soil index, shadow (B), gaussian low-pass filter (B), and local correlation (N) (A4).

For the multiclass problem without vegetation mask, XGBoost performs better than 5-layer neural networks although the latter has relatively higher overall accuracy (90.4% vs. 90.9%). The problem of the neural networks was that it has the tendency of labeling most pixels as non-building due to the unbalance of samples for each class. For XGBoost predictions (A6), it is able to detect building pixels from the background but might not be able to preserve the shape and the completeness of them. The F1 score is 0.88 but Cohen's kappa is 0.19 and the balanced accuracy is 40.62%. Besides, there is also some confusion between no-damage and destroyed. The ROC curve (A7) indicated that the destroyed class gives relatively more accurate predictions. Important features for XGBoost are SNIC segmentation and convolutions (shape and manhattan) of red and NIR channels. As for the neural networks, the 2-layer network has an accuracy of 80.3% but achieved better predictions than the 5-layer network when visually evaluated. However, the 2-layer network has the same problem as the XGBoost and omits more buildings.

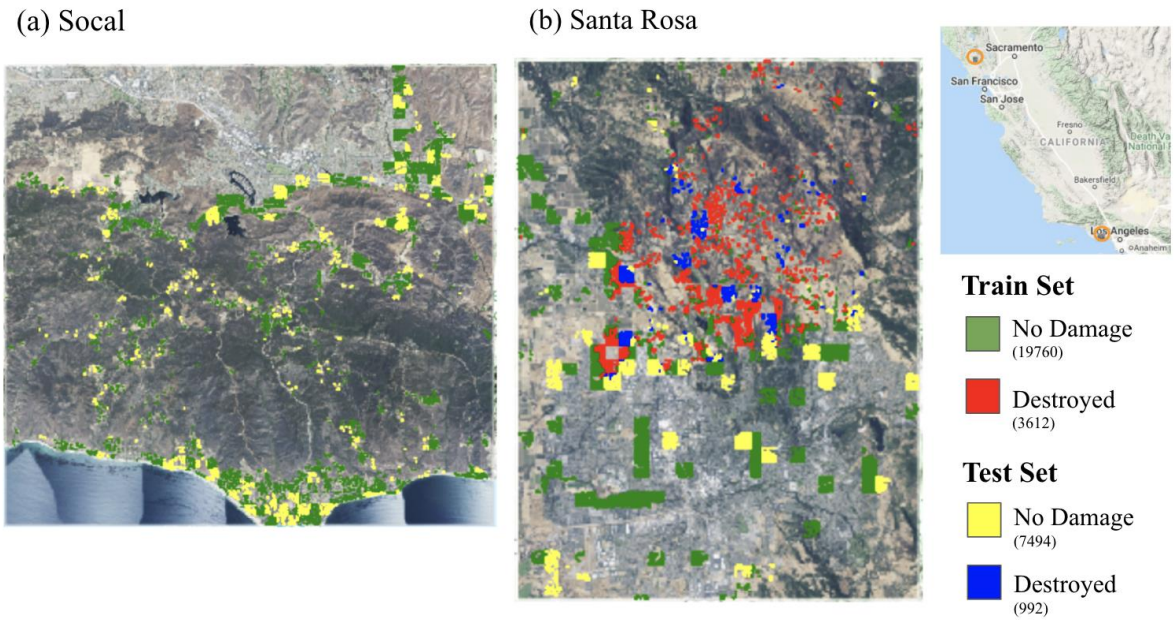
## V. Discussion

Several insights and major take-aways can be made particularly from the issues dealing with imbalances between inter-class training samples, data handling, and memory allocation limitations, among others. For the case of whole image multi-label classification, without applying a vegetation mask, the training data is dominated by non-building pixels which lead to model overfit and bias towards this label. In addition, when modifying a neural network, simply increasing the number of layers might achieve better global accuracy but may impair the prediction performance of some classes, especially for unbalanced sample size. Furthermore, due to RAM usage restrictions on Google Colab notebook, there were several compromises made during the training process. For example, XGBoost and neural networks were only built on a subset of 4 million pixels from the whole training dataset, while the SVM and RF classifications were trained on resampled imagery. Finally, although simple neural networks failed to detect buildings and classify damages accurately, sophisticated visual recognition networks, such as Fast R-CNN or Mask R-CNN might largely improve the results, especially for semantic segmentations. Additional possibilities include implementing a change detection workflow to incorporate temporal differences of post-disaster classifications within pre-disaster building footprints. This will mainly limit confusion and noise associated with the non-building classes and improve overall results.

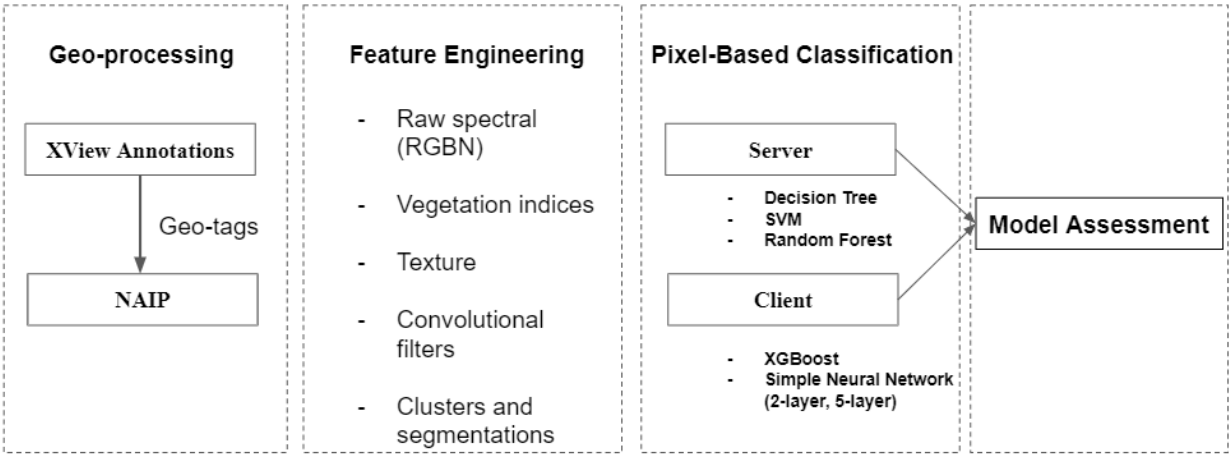
**REFERENCES**


1. Belgiu, M., & Drăguţ, L. (2014). Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *96*, 67-75.
2. Chen, R.; Li, X.; Li, J. Object-based features for house detection from RGB high-resolution images. *Remote Sens.* 2018, *10*, 451.
3. McNamara, D., Mell, W., & Maranghides, A. (2020). Object-based post-fire aerial image classification for building damage, destruction and defensive actions at the 2012 Colorado Waldo Canyon Fire. International Journal of Wildland Fire, 29(2), 174-189.
4. T. Kohonen, "Learning Vector Quantization", The Handbook of Brain Theory and Neural Networks, 2nd Edition, MIT Press, 2003.
5. Achanta, R., & Süsstrunk, S. (2017). *Superpixels and Polygons Using Simple Non-Iterative Clustering*.
6. Klambauer, G., Unterthiner, T., Hochreiter, S. (2017). Self-normalizing neural networks. *Advances in neural information processing systems*.
7. Gupta, R., Hosfelt, R., Sajeev, S., Patel, N., Gaston, M. (2019). XBD: A Dataset for Assessing Building Damage from Satellite Imagery.
8. Pereira J.M.C., Sá A.C.L., Sousa A.M.O., Silva J.M.N., Santos T.N., Carreiras J.M.B. (1999) Spectral characterisation and discrimination of burnt areas. In: Chuvieco E. (eds) Remote Sensing of Large Wildfires. Springer, Berlin, Heidelberg
9. Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. 2013. “High-Resolution Global Maps of 21st-Century Forest Cover Change.” Science 342 (15 November): 850–53.
10. Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6), 610-621.

# APPENDIX




**A1.** XView dataset overlaid on top of NAIP imagery in study region (a) Southern California (i.e. Woolsey fire) and (b) Santa Rosa (i.e. Carr fire).

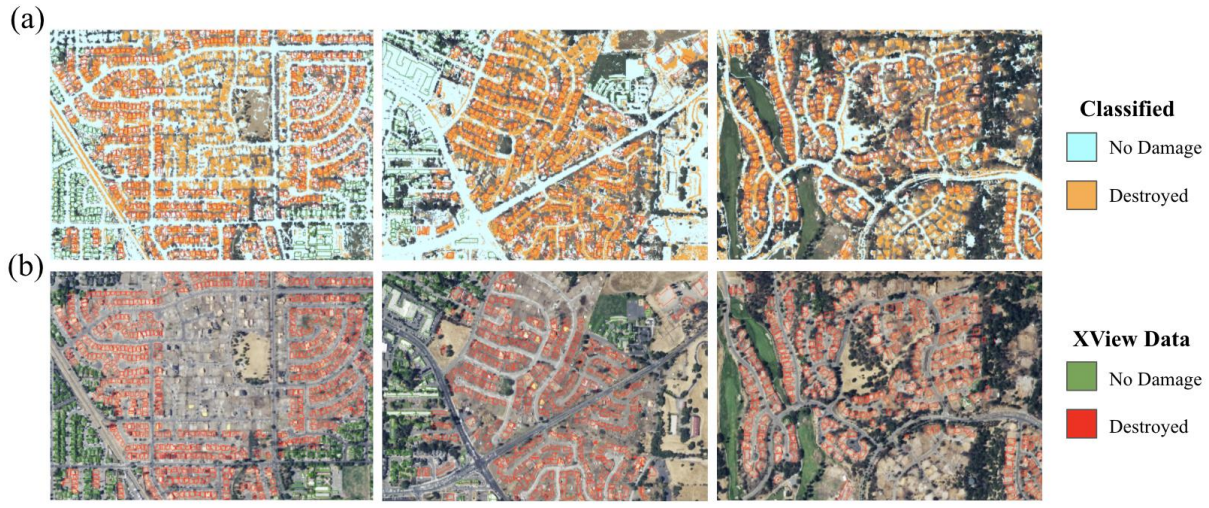



**A2.** Project pipeline

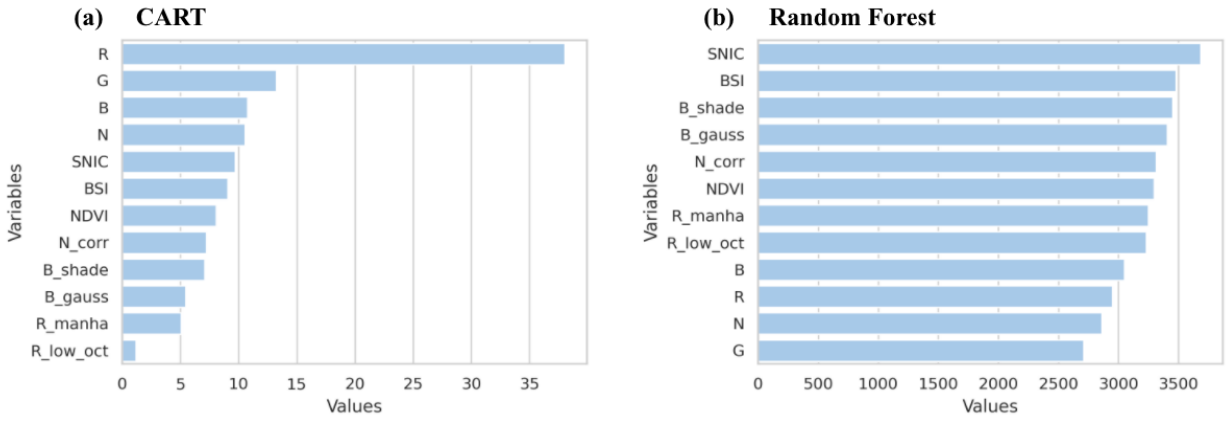
**Table I. Information Regarding Remote Sensing Indices**

| Index | Abbreviation | Equation | Description |
|---|---|---|---|
| Normalized Difference Vegetation Index | NDVI | (NIR - Red) / (NIR + Red) | Indicator of healthy vegetation based on properties of photosynthetic activity. |
| Bare Soil Index | BSI | (R+B-G)/(R+G+B) | Enhance identification of bare soils and fallow lands. |
| Shadow Index | SI | sqrt((256-B)*(256-G)) | Informative for identifying building shadows i.e. height |
| Gray-Level Co-Occurrence Matrix | GLCM Texture | - | Computes texture properties including entropy, shade, and contrast based on the |

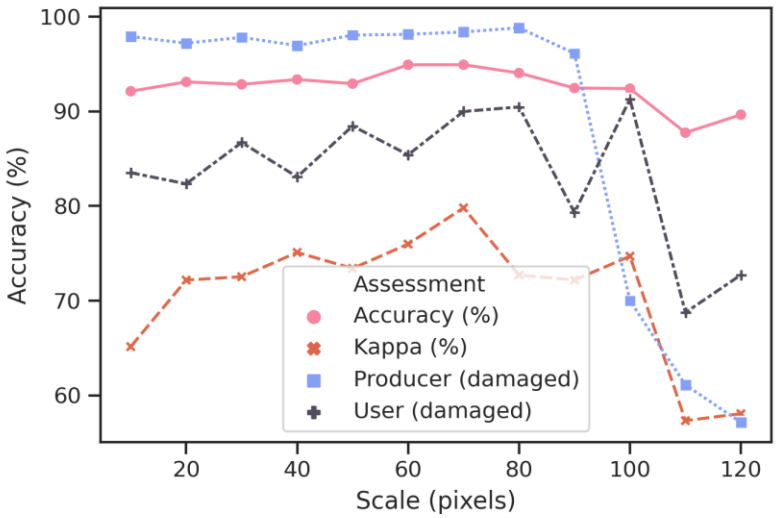
| | | | |
|---|---|---|---|
| Canny Edge Detection | Edge | - | spatial relationship of pixels in the GLCM. Extracts structural information from different vision objects. |



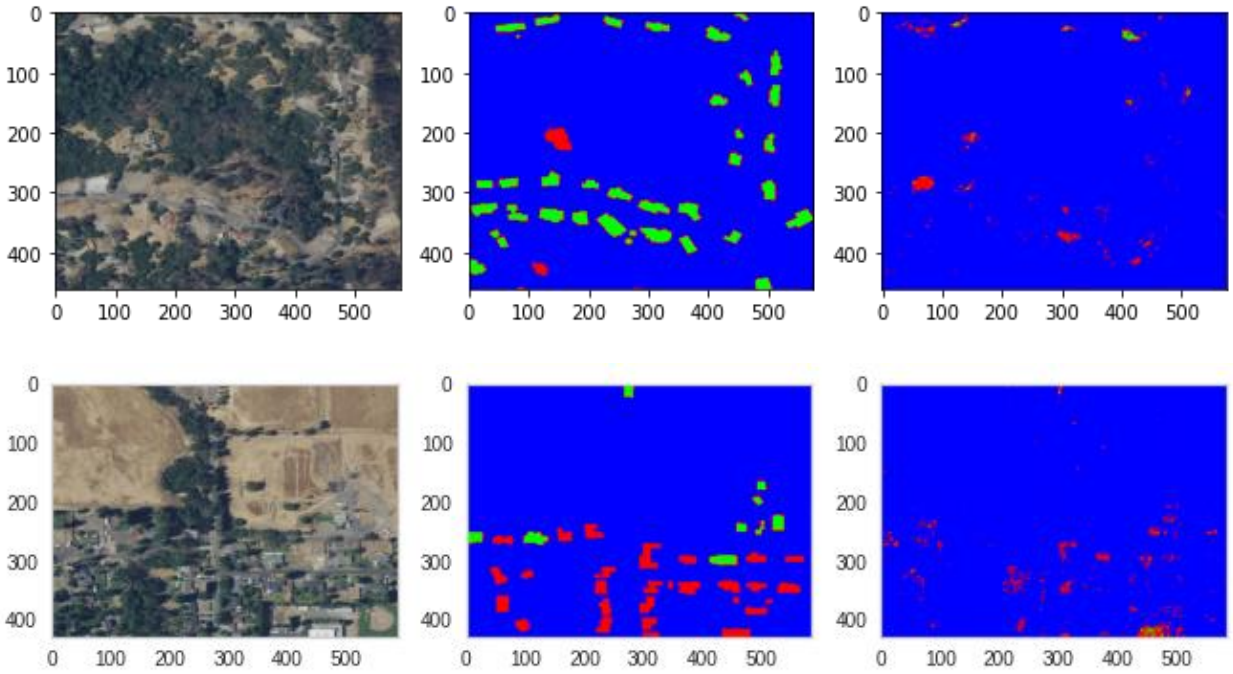

**A3.** (a) Visually illustrates the Random Forest classification output (b) Annotated dataset from XView for comparative purposes.



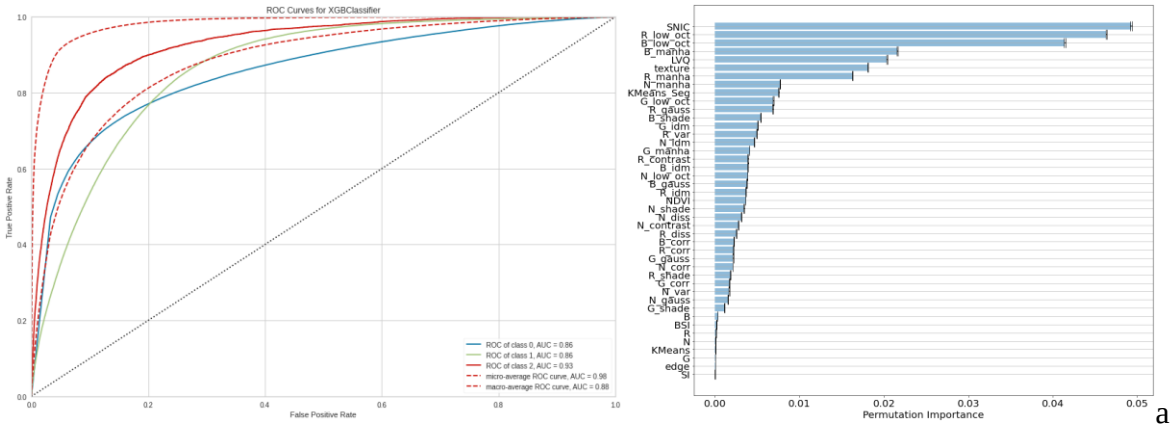

**A4.** Feature importance from (a) CART and (b) Random Forest




**A5.** The effects of scaling during the validation step to assess final accuracy of building damage of random forest classifier.



**A6.** Predictions from XGBoost models. Left: NAIP images in RGB; Middle: ground truth annotations; Right: predicted buildings (red: no damage, green: destroyed buildings, blue: non-building).




a

**A7.** Feature Importance and ROC Curve for the XGBoost model.

**Table II. Model performance metrics for CART, SVM, and RF classifiers including the overall accuracy, Kappa, Producer, and Consumer accuracies.**

| Model | Overall Accuracy | Kappa | Producer Accuracy | | Consumer Accuracy | |
|---|---|---|---|---|---|---|
| | | | *No Damage* | *Destroyed* | *No Damage* | *Destroyed* |
| CART | 0.86 | 0.51 | 0.91 | 0.60 | 0.92 | 0.59 |
| Support Vector Machine | 0.91 | 0.69 | 0.99 | 0.6 | 0.91 | 0.95 |
| Random Forest | 0.93 | 0.73 | 0.98 | 0.69 | 0.94 | 0.88 |

**Table III. Confusion matrix comparing Random Forest classified results with xView ground-based assessment results.**

| | | Reference | | | Commission Errors | User's Accuracy |
|---|---|---|---|---|---|---|
| | | No-Damage | Destroyed | Totals | | |
| **Classified** | No-Damage | 1534 | 34 | 1568 | 6% | 94% |
| | Destroyed | 101 | 222 | 323 | 13% | 87% |
| | Totals | 1635 | 256 | 1891 | - | - |
| | Omission errors | 2% | 31% | - | - | - |
| | Producer's accuracy | 98% | 69% | - | - | - |