Identifying Damaged Roofs in Baltimore City for Inspection

Data Science For Social Good
Summer Fellowship

Carnegie Mellon University

Justin Clark
Harvard University

Jonas Coelho
FGV - EBAPE

Chae Won Lee
University of Washington

Rayid Ghani
Carnegie Mellon University

Kit Rodolfa
Carnegie Mellon University

Introduction

In Baltimore City, there are more than 15k abandoned residential buildings. The lack of maintenance on some of these buildings leads to severe structural damage and negative impacts to residents. However, the city has limited capacity to inspect properties, especially for damaged roofs, which prevents timely and effective intervention.

More than 1 in 10 residences in Baltimore are known vacancies and potentially damaged

We partnered with Baltimore City's Department of Housing and Community Development (DHCD), a government organization that works to improve the quality of life for all Baltimore City residents by revitalizing and redeveloping communities and promoting access to quality affordable housing.

Approach

Use high resolution aerial images of Baltimore City and other city government provided data to apply machine learning methods to detect damaged roofs in residential row houses in Baltimore City.

Data

The DHCD provided aerial imagery of the entire city and tabular data for all of its the buildings including the following:

- Housing inspection notes
- Construction and demolition permits
- 311 calls complaints
- Real estate data, including sales price
- Tax parcel address with geo references for each building

In addition, we also used Census Data (ACS5 2020) that provided information about race, hh median income, rent status, and historical redlining class at a census Block Group level for auditing the model for bias.

Method

Models

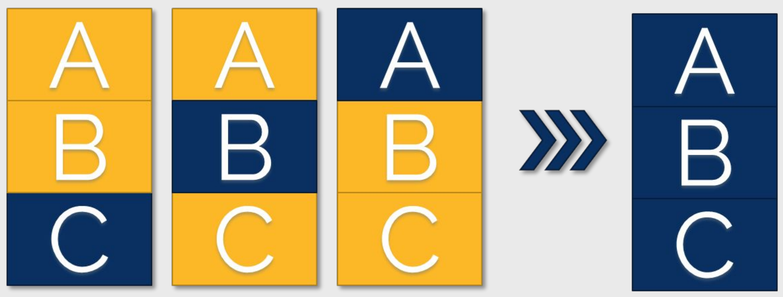
We tested the performance of three different models and model combinations. Here were our best performing models:

Type	Data used
Proportion of dark pixels	Aerial imagery
Random forest	Inspection notes, time knowingly vacant, number of 311 calls within radius and more
Neural network	Aerial imagery

In the first model, we use the proportion of dark pixels in the image of each block lot as a proxy for roof damage. The second Random Forest model only uses tabular data. Finally, we classify aerial images using a pre-trained CNN (Resnet-18).

Labelling and validation

To train and validate our models, we used labels manually created by interns that looked at roof pictures to visually assess its damage level. Performing a k-fold splitting, we were able to train on different subsets of the labeled set and validate the prediction on the remaining subset. This way, we were able to get predictions for the entire set without including them on the training process, as shown on the figure below. Yellow subsets are used for training on each round with blue subsets being the ones where predictions are made



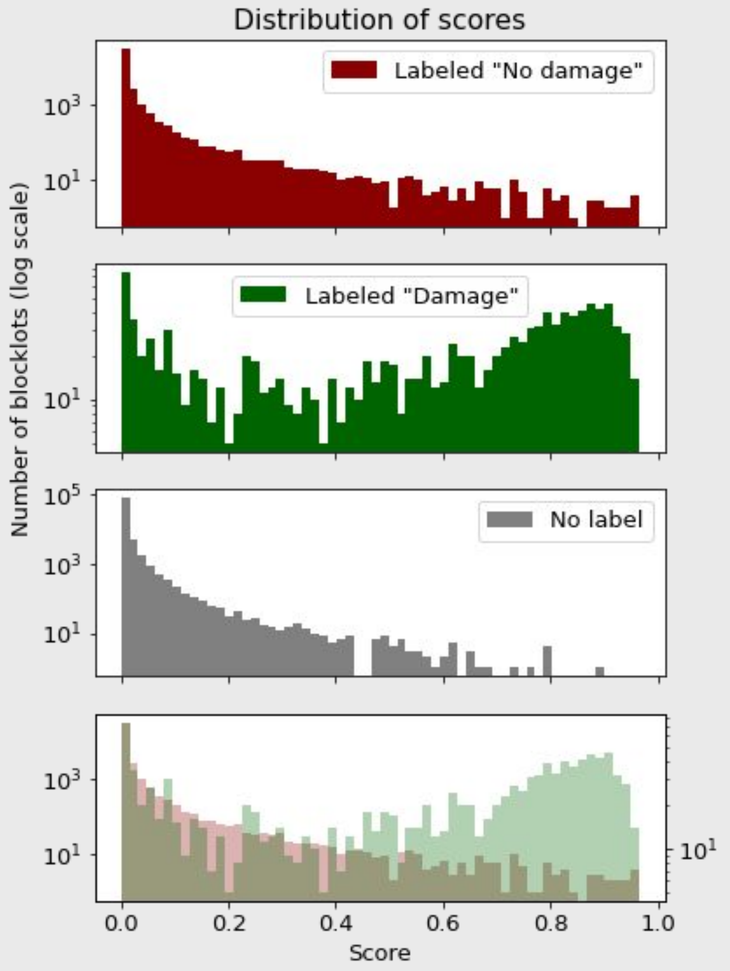
Metrics

The city has a capacity of inspection around 1000 units a year, so we used this number as a threshold on what units to prioritize. After doing predictions for the entire labeled set, we ranked each residential unit according to their predicted damage score and selected the top thousand highest scores. We then considered everything in this list as positive prediction and everything below it as a negative prediction. Precision and recall for each model are the following:

Top 1k N Labeled	Top 1k N Positive Labels	Tok 1k N Negative Labels	Top 1k Precision	Top 1k Recall
966	824	142	0.853	0.662

Results

Our best model captures 824 known damaged roofs and 80 previously unknown damaged roofs in the top 1000. We focus on **precision** and **false negative rate** to measure efficiency to avoid disproportionately missing damaged roofs. The figures below show the distribution of each label on the top thousand list.



Impact

The model will have the potential to improve the lives of 5000 households who live on blocks with damaged roofs in Baltimore City. Our model also obtained results that did not demonstrated measurable bias against historically disadvantaged areas. Thus, the following potential impacts are expected:

- Increased safety by intervening on collapsing buildings
- Reduced blight
- Accelerating redevelopment in Baltimore City
- More informed decision making by the DHCD

BALTIMORE CITY
DEPARTMENT OF HOUSING & COMMUNITY DEVELOPMENT

This work was done during the Data Science for Social Good Fellowship at Carnegie Mellon University.