Flood Damage Detection in Satellite Imagery Using Machine Learning

Margarita Popova | margeorita@gmail.com

TABLE OF CONTENTS

INTRODUCTION

SUMMARY

Natural disasters such as floods and hurricanes require operational monitoring of the situation and quick decisions from emergency managers and the government. It is crucial to estimate the devastating consequences of natural threats quickly and effectively to provide help for affected people as quickly as possible. With satellite images, we can analyze huge territories at once and Machine Learning techniques help us process these big data and detect all kinds of devastations on the images.

THE GOAL OF THE PROJECT

The goal of the following work is to detect devastation after Hurricane Harvey on satellite images with Machine Learning methods.

SOURCES

The project used an imagery dataset that came from the IEEE data port. The images came from MAXAR – a satellite images distributor.

Elevation models from the open source USGS portal were also used as an additional data source.

DATA DESCRIPTION

The Imagery Dataset contain 21057 images divided on two classes – Damage and No Damage.

The number of images with damage:

- **14021**

The number of images without damage:

- **7036**

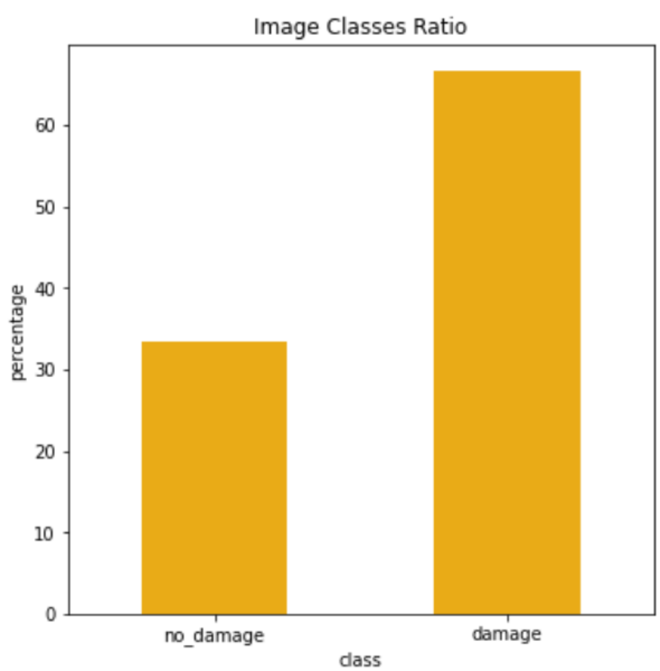


Figure 1. The image classes ratio in the dataset

The filenames of images contain the coordinates of the images' central points. Using that data, additional features were added to the dataset to try to predict damages based on only geographical metrics such as elevation, slopes, latitude and longitude.

DATA EXPLORATION

Terrain metrics was extracted from the Digital Elevation Model (DEM) from the open-source resource USGS (United States Geological Survey).

Terrain metrics:

- Elevation
- Slope degree
- Slope aspects (the orientation of slope)
- Distance to the closest water body

Analysis of these metrics showed that the majority of data points were located on a flat surface with a slope degree less than 2^o (see fig. 2).

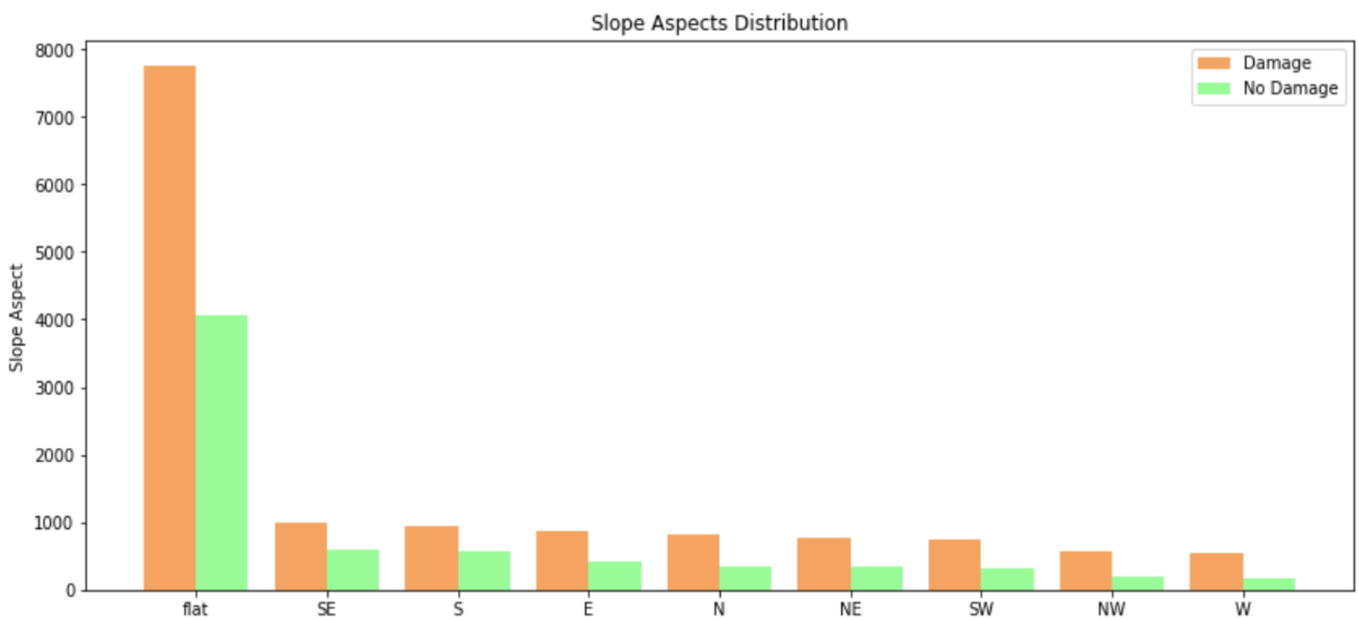


Figure 2. The Slope aspects distribution in the dataset

LOGISTIC REGRESSION

Logistic Regression model was used to predict the damages with coordinates and terrain metrics. As predictors in the models were used next features:

- Slope degree
- Elevation
- Longitude
- Western slopes
- Southern slopes
- Southeast slopes

Unfortunately, the accuracy of that model did not greater than 66% and we can not trust such a model because we have 66% data points with damages and we could get the same accuracy even without modelling just guessing that all data points contain the damage.

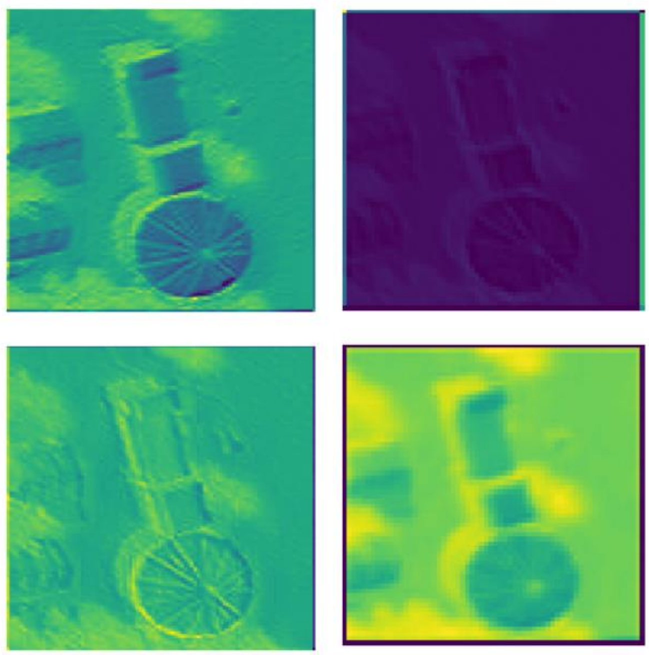
Simple models do not handle predict flood damage with a limited set of features in the dataset and that is the reason for using of more complex models for image classification.

DEEP LEARNING

The Convolutional Neural Network (CNN) with 16 hidden layers was used for image classification to detect flood damages after hurricane.

The accuracy of the CNN is about **97%**.

To understand how convolutional neural networks learn dependencies of an image, different features captured at each layer can be visualized and analyzed (see fig. 3).



a) b)

Figure 3. Representations of Activations of CNN layers.

a) - Outputs after the first layer convolution applied

b) - Outputs after the second layer convolution applied

The CNN detects the lighter structures surrounded by darker background. In our case it is flooded houses surrrounded by dirty brown water.

PREDICTIONS

The CNN model also was examined in detecting damages on the satellite images from a different dataset. These new images came from the Louisiana Swipe Map powered by ESRI and they have a size and format differ from the images in our main dataset.

CNN predicted damages for the 10 images and made only 2 mistakes which is quite decent result for the CNN.

Figure 5. Incorrect CNN Prediction: 0.12% - Damage, 99.88% - No Damage

Figure 5. Correct CNN Prediction: 99.54% - Damage, 0.46% - No Damage

FUTURE DIRECTIONS

Future research will be focused on testing the data from different sources to find and fix the weaknesses of the model through various approaches such as adding new hidden layers in the model, using pre-trained models, expanding the image dataset and others.

Weaknesses of the current CNN model that should be corrected:

- The CNN may be confused by shadows from buildings
- The CNN may be confused by clouds on the images
- The CNN may classify the image with a bright house in the middle of the plowed field as an image with damage with high probability