Deep Learning Approach for Disaster & Damage Classification Using Satellite Imagery

Project Member Names: Jesse Maki (3040751652), Jeremy Millford (3040748480), Shawn Schulz (3033824499)

Video Presentation link: https://youtu.be/88MZr2LDnoI

3040751652_3040748480_3033824499

ABSTRACT

Natural Disasters increasingly threaten communities worldwide, resulting in a crucial need for rapid damage assessment and classification. This paper presents an efficient deep-learning approach using Convolution Neural Networks (CNN) for classifying both disaster types and damage classification from satellite imagery. Our model analyzes a dataset of approximately 26,000 satellite images across three disaster types: wildfires, flooding, and post-disaster hurricane images. The CNN architecture employs two convolution layers with 3x3 kernels, followed by MaxPooling and dense layers, achieving effective classification despite the challenges that follow using satellite imagery resized to lower resolutions. The results demonstrate the model's capability to assist emergency management teams in rapid disaster assessment, potentially improving response times and resource allocation during critical periods.

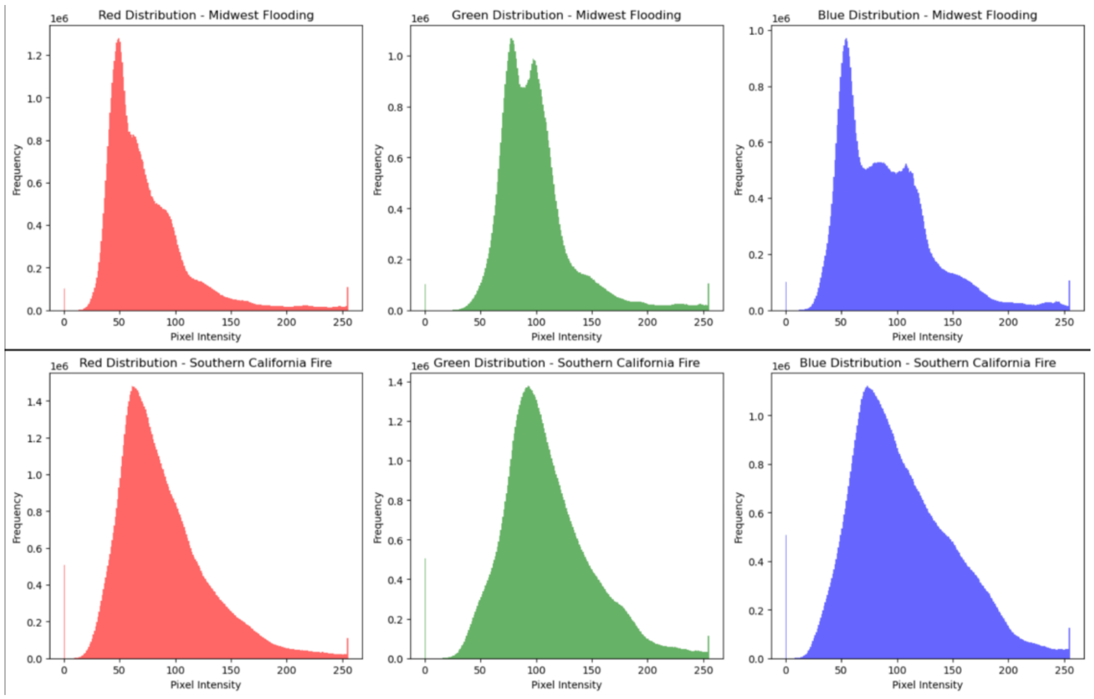
INTRODUCTION

Natural disasters are becoming a more frequent and devastating occurrence in many countries. Effective prediction of future events can help emergency management teams advise and help those affected. Low orbit satellite imagery data is a great data source for large scale categorization because there are so many weather and imagery satellites recording information at any moment. However for recovery efforts and planning the on the ground details of building damage, flood vs fire damage, and classification of the degree of damage to infrastructure and the environment are more important. The challenge is that satellite imagery data has a low resolution for features at ground level scale and can be difficult for a human observer to categorize or classify. These challenges make this project a great candidate for a machine learning project that can make the most of low resolution features, recognize patterns, and process many images quickly. Our team recognized this use case application and we have utilized a convolution neural network with deep learning layers to make predictions of disaster type and damage level. We chose a CNN because the convolution layers give the model more information to work from which it can compare to find patterns that we thought would be especially effective for edge detection which would be relevant when classifying building damage. Not surprisingly, some models in literature reports with similar goals have similar models to this that utilize CNN and deep learning (Kim et al, 2022). Where we think our model is different is it utilizes a very accessible and efficient architecture and is still able to make predictions to a high degree of accuracy, precision, and recall. With this accessibility more researchers could hopefully use the model and help more people affected by natural disasters.

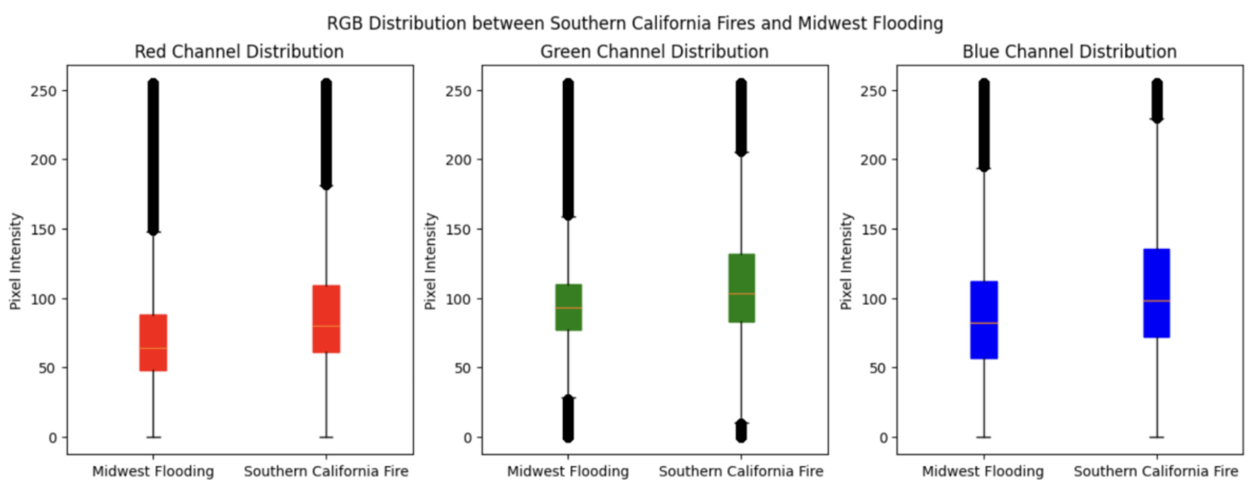
DESCRIPTION OF DATA

We knew we were going to be working with image data with classifications so we made a data frame that included the type of disaster label, the flattened image data, and the disaster damage label. From there we could separate the image data in the RGB values to give us more data and information to work with. To better understand the RGB distributions we plotted them according to disaster type as histograms and box charts in figures 1-3 below. During exploratory data analysis, we found good separation of color channels by disaster type but less strong separation by hurricane damage level.We generally found less strong differences in distributions for grayscale pixel intensities, image size distributions, and sobel edge filter distributions. In figure 4 you can see our chart of image sizes which we used to decide on our image resize normalization. We also plotted with information using a log transformation in Fig 5. To account for stronger distribution differences between RGB pixel intensities but not in filters or grayscale pixel intensities, we chose to design a convolutional neural network model to help take advantage of good features in full color channel images, the large size of all datasets, and to utilize deep learning to apply better filters for classification. Such a model should simultaneously be able to make use of the better RGB distribution separation between disaster types while also being able to adaptively learn filters to help distinguish damage level for the hurricane Matthew dataset. We also explored the distribution of damage classification labels for the Hurricane Mathew data set to better inform ourselves of training data strengths and weaknesses (Fig. 6). Data was prepared for modeling through a minimal set of preprocessing steps. After loading the data, image labels were encoded either with binary labels for task A or one hot-encoded categorical labels for task B. All images were resized to 64 x 64 pixels and pixel intensity values were normalized to a scale between 0 and 1. Image resizing was chosen to balance model training speed with an acceptable retention of resolution of features. The data was randomly shuffled, 20% of the data was withheld for validation, and the pixel intensity data was converted to the tensorflow tensor data format with a batch size of 32 before being inputted into one of two convolutional neural network models depending on the relevant task

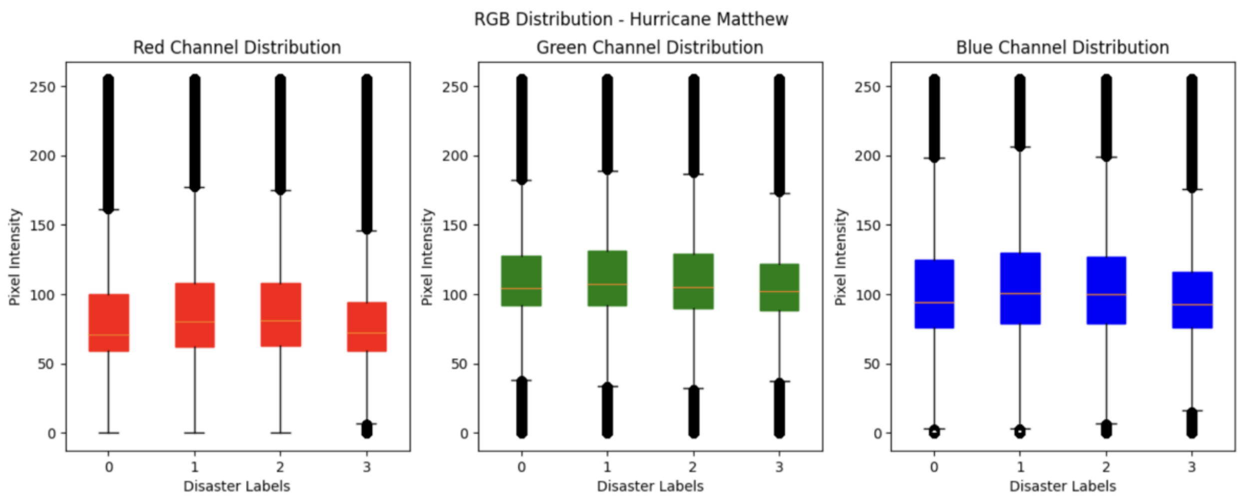
Description fig 1.



Description fig 2.



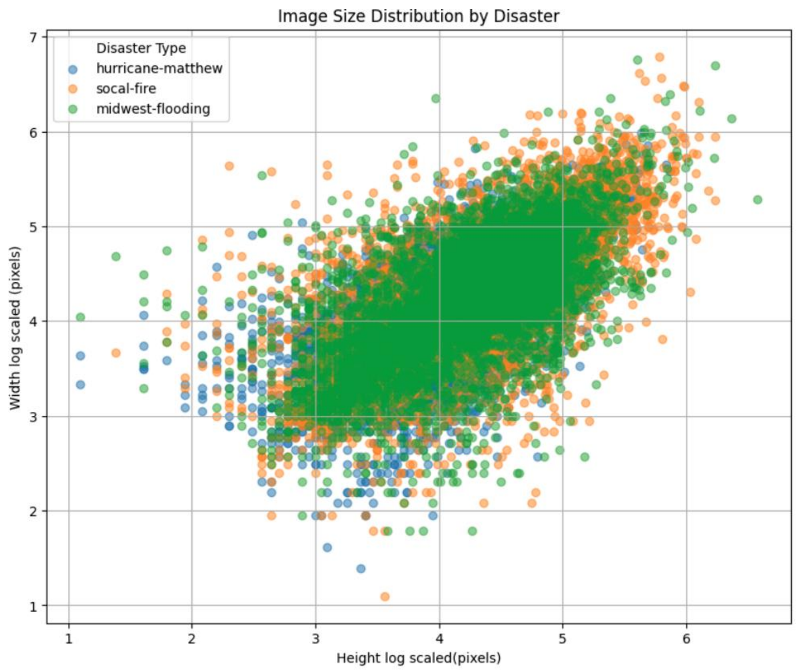
Description fig 3.



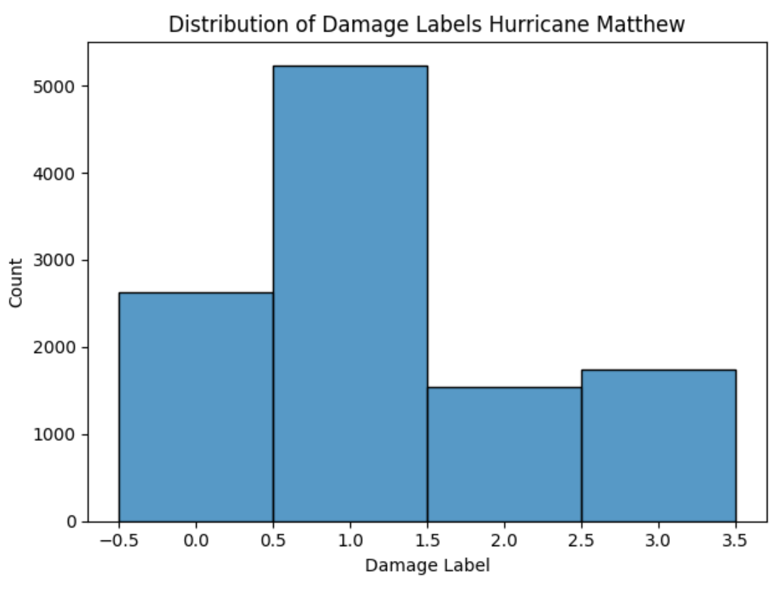
Description fig 4.

Disaster	Number of pictures	Image size (L/W mean) pixels	Damage labels
SoCal fire	8380	99.8/101.4	8380
midwest-flooding	7004	82.4/84.2	7004
hurricane-matthew	11151	60.1/61.5	11151

Description fig 5.



Description Fig 6.



METHODOLOGY

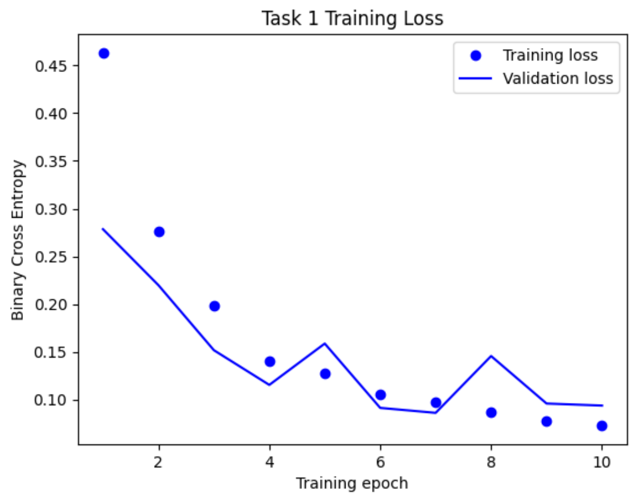
For Task A, the current implementation of the CNN model architecture consists of two convolutional filters with a 3x3 kernel using ReLU activation, followed by a MaxPooling layer. This output is then passed into a Flatten layer to convert the 2D feature maps into a 1D vector. Next, a dense layer with 12 units and ReLU activation processes this flattened vector, followed by a final dense layer with softmax activation for binary classification. A 50% neuron dropout is applied between the two dense layers. The model uses Adam optimizer, binary cross entropy for loss calculations during backpropagation and percentage accuracy from training and validation labels as the final performance metric. The model was trained for 10 epochs.

For Task B, the CNN model architecture consists of 4 convolutional filter layers with 3x3 kernels, a max pooling layer, a Flatten layer, then a Dense layer with 36 units, a 20% neuron dropout layer and a 4 unit dense layer for the 4 levels of damage classification. The relu activation function was used for all convolution layers and the second to last Dense layer, and the softmax function was used for the output Dense layer. The AdamW optimizer was used. Categorical cross entropy was used as the loss function during backpropagation with F1 score as the metric used for performance assessment. The model was trained for 30 epochs.

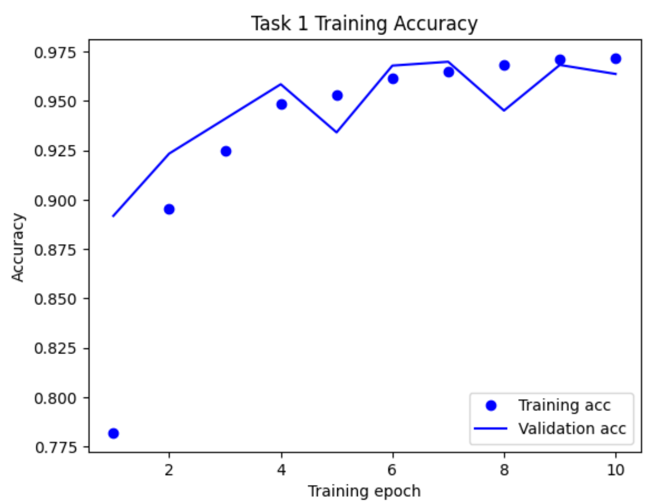
RESULTS

For task A, our CNN model had very high accuracy for the training and validation data, achieving greater than 80% accuracy early on. We performed hyperparameter tuning by tweaking the number of filters applied in convolutional layers and the number of max pooling layers which helped to achieve higher accuracy for both models. After tuning hyperparameters for task A, the addition of max pooling layers, higher filter numbers, greater epochs for training and a 50% neuron dropout improved accuracy to a very high value of ~93%. Interestingly, validation accuracy and validation loss fluctuates to higher and lower values respectively in epochs prior to the final training epoch, suggesting that there is still additional model accuracy that can be squeezed out via hyper parameter tuning

Results fig 1.

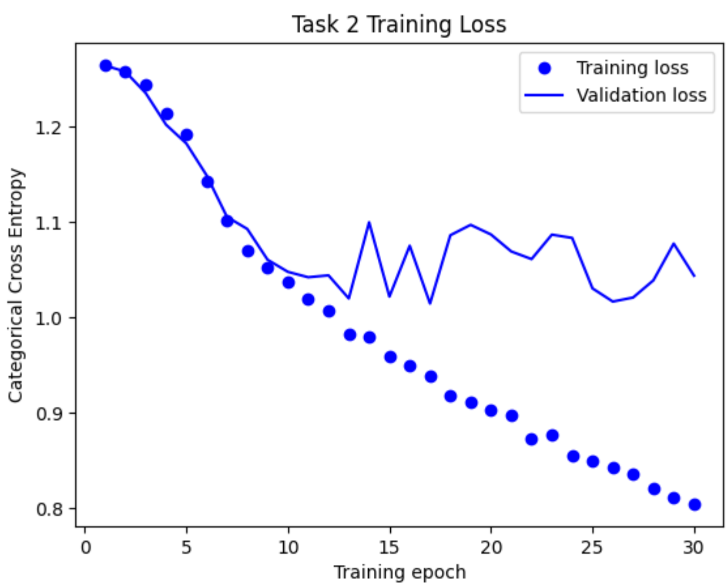


Results fig 2.

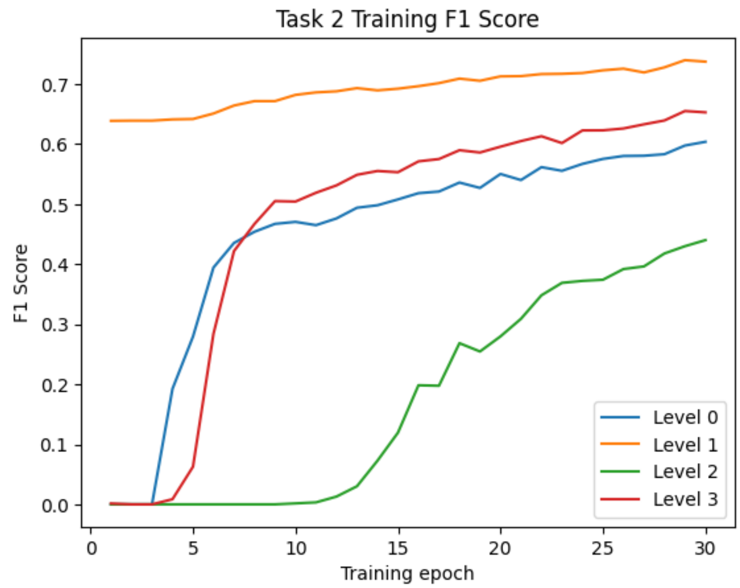


For Task B, our CNN model shows acceptable F1 Scores for training and validation data, achieving ~45.6 for validation F1 score. Interestingly, our validation loss remained constant past 10 epochs while our training loss continued to decrease, suggesting overfitting and a limitation to the generalizability of our model for task B.Surprisingly, it seems that damage level 0 and damage level 3 have similar f1 scores, damage level 2 has the lowest f1 score and damage level 1 has the highest f1 score. We initially would have predicted that level 0 or level 3 might have the highest f1 score, since no damage or very high hurricane damage would likely look more different than an intermediate amount of damage, but perhaps the model has trouble distinguishing normal pictures of landscapes in level 0 from actually damaged areas, or conversely pictures depicting damage might be noisy and seem undamaged from our somewhat compressed resized images used to train the model.

Results fig 3.



Results fig 4.



Overall our model performed quite well for task A and performed acceptably for task B. Our results show a potential to be used during the aftermath of natural disasters to help quickly assess areas in danger and help guide response teams. Our first model for task A could potentially be used to narrow searches in areas impacted by hurricanes. However, since the F1 scores are not as high for certain damage levels, this should be paired with human oversight and expert analysis to ensure that damaged areas aren't missed and undamaged areas aren't erroneously searched. Additional work could be done to optimize model training and prediction to use less compute resources, and to resolve some idiosyncrasies in model performance for Task B to improve F1 scores for all damage level assessments. Future work could also be done to attempt alternative model architectures besides 2D convolutional layers, such as using a diffusion model architecture and/or a variational autoencoder model to detect damaged areas from satellite images. Additionally, further analysis could be done to ensure the construction of an unbiased model. For example, our dataset only included satellite images in particular regions in the US, which could limit the generalizability of the model. Our model also could be prone to discriminating against certain neighborhoods or regions based on wealth or other social inequalities. To help investigate this, further data could be collected on the areas imaged and used to help examine potential bias in our models.

Discussion:

With the successful implementation and testing of our model our hope is that it could help more emergency management teams support those affected by natural disasters. Societal impact of having a model infrastructure in place to assess damage for quicker repair, monitor in real time, and provide an early warning to high risk areas would be very beneficial. Economical analysis and aid projection for areas would be an application with a high impact.

With satellite information there can be associated ethical concerns about privacy that should be noted but since the resolution is not too revealing at the ground level the potential benefits of the model usage are so high privacy concerns aren't a paramount concern. Other pressing ethical concerns could be the use of this information and model by a corporation to enact predator practices like higher insurance rates spikes before a large natural disaster for instance. Finally a common concern with some machine learning models is lack of understanding and the model being a "black-box" or the model making unintended conclusions for example classifying based on the economic level of infrastructure in an area. This could be ameliorated with a human in loop to help others understand and adjust the model as well as add and classify new data appropriately.

Citations

Kim J, Lee M, Han H, Kim D, Bae Y, Kim HS. Case Study: Development of the CNN Model Considering Teleconnection for Spatial Downscaling of Precipitation in a Climate Change Scenario. *Sustainability*. 2022; 14(8):4719. https://doi.org/10.3390/su14084719

Smith, A. B., & Matthews, J. L. (2015, March 11). *Quantifying uncertainty and variable sensitivity within the US billion-dollar weather and climate disaster cost estimates - natural hazards*. SpringerLink. https://link.springer.com/article/10.1007/s11069-015-1678-x

USAFacts. (2024, September 6). *Are major natural disasters increasing?* https://usafacts.org/articles/are-the-number-of-major-natural-disasters-increasing/