For locations in the United States at random dates, to complete our non-fire dataset we also obtain data for the non-fire class in two ways.

- We use a simple CNN model in three ways:
  - Model 1: CNN + 3 channel data: RGB
  - Model 2: CNN + 8 channel data: RGB + NDVI + BAI + NBR_RAW
  - Model 3: CNN + Regularization (dropout) + 8 channel data (RGB + NDVI + BAI + NBR_RAW)

- We also use a transfer learning model: ResNet-50 with pretrained weights + refining tail 2 layers

Models

- Our best model was the CNN with dropout layers and 8 channel data (Model 3).
  - We exhibit very low variance, and some bias.
  - We estimated Bayes Optimal Error with non-expert human labeling to be ~15%. We believe that we were able to beat this accuracy by synthesizing multiple channels of data, which is not comprehensible to the human eye. Expert human labels may have also yielded a smaller Bayes Optimal Error.

Results

- Our best model was the CNN with dropout layers and 8 channel data (Model 3).
  - We exhibit very low variance, and some bias.
  - We estimated Bayes Optimal Error with non-expert human labeling to be ~15%. We believe that we were able to beat this accuracy by synthesizing multiple channels of data, which is not comprehensible to the human eye. Expert human labels may have also yielded a smaller Bayes Optimal Error.

Discussion

- Simple CNN + More (3) channels works best (bias-free)
- Very deep models like ResNet-50 can lead to overfitting on the data

Future Work

- As expected, adding additional channels of data significantly improved our results, increasing test accuracy by 8 percentage points. We believe that this is because significant information is contained in these supplemental satellite images.
- Dropout Layers also significantly improved our results, increasing test accuracy by 17 percentage points. Although dropout is a regularization technique and thus is primarily intended to reduce variance, this drastic reduction in our model bias was also an interesting and welcome result. Our interpretation of this is that since we were only able to train for limited (~10) epochs, in either case, i.e. with and without dropout, our model without dropout was not near its true convergence. With dropout, however, we believe that the model had sufficient computational resources to train for longer epochs. However, the dropout may have made speed-up training such that the accuracy obtained after only 10 epochs may be near-converged training accuracy with dropout.
- This ineffective use of transfer learning with the application of ResNet-50 we attribute to overfitting, since we see a large widening of the variance as well as a drop in performance overall with this method.

Reference