Applied Machine Learning Methods to Predict the Air Quality Impact of Wildfires in Northern California

Undergraduate Research Symposium

April 8th, 2021

Gautam Agarwal, Shreyans Saraogi, Eliot Kim

ProjectX

UofT AI

Massachusetts Institute of Technology

University of Toronto

Stanford University

Berkeley University of California

Carnegie Mellon University

Johns Hopkins University

Illinois University of Illinois at Urbana-Champaign

McGill University

University of Michigan

Caltech

Université de Montréal

UBC

Cornell University

Wisconsin University of Wisconsin–Madison

Queen's University

University of Massachusetts Amherst

USC University of Southern California

University of Alberta

Universidad de Chile

Duquesne University

USP Universidade de São Paulo

Technische Universität Darmstadt

Politechnika Warszawska


X

Wildfires

- 7,948 fires occurred in California in 2018
- 1,975,086 acres of land were destroyed
- Resources worth $635M were diverted
- 22,905 structures were destroyed

Air Quality

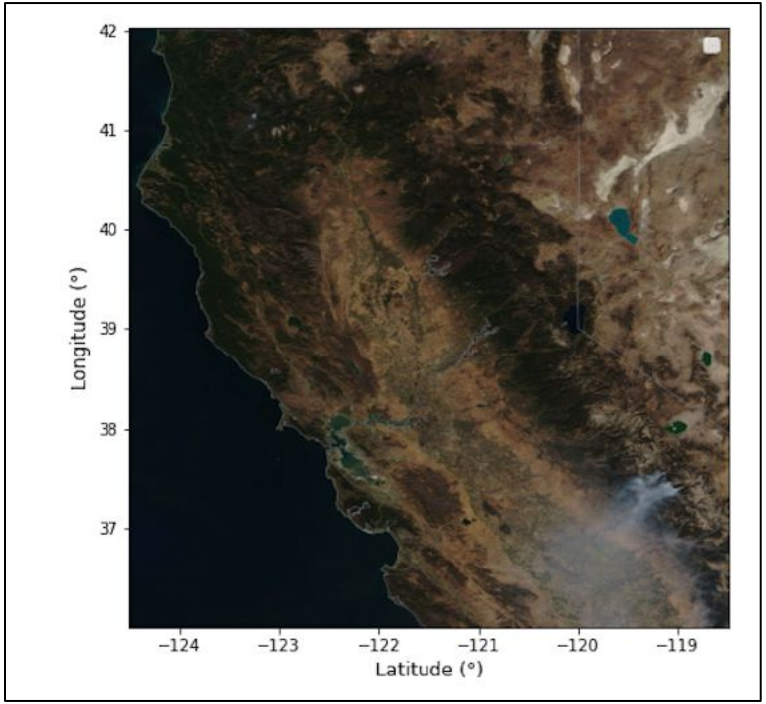
- **Fine Particulate Matter** (**$PM_{2.5}$**) : Small breathable pollutants harmful to human health

- **$PM_{2.5}$** levels reached 453µg/m^3 in Northern California during October 2021

- The World Health Organisation recommended safe limit for **$PM_{2.5}$** is 10µg/m^3

Domain: Northern California

- **36°N to 42°N, 118.5°W to 124.5°W**
- Varied land uses and topographies
- Prone to high-impact wildfires
- Frequent and severe droughts induced by climate change



Datasets

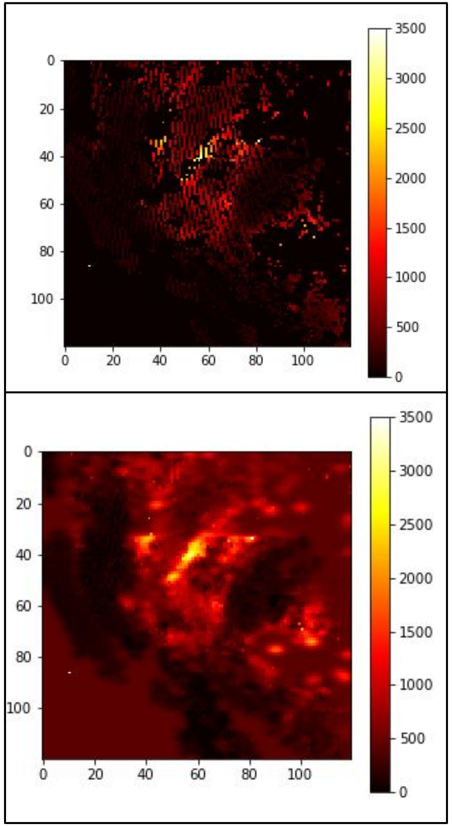
- Moderate Resolution Imaging Spectroradiometer **(MODIS)**
  - Collected from the Terra and Aqua Satellites
- NCEP North American Regional Reanalysis **(NARR)**
  - Satellite and deterministic model generated data
- Date Range: January 1st, 2010-December 31st, 2019

Datasets

- Four Categories
  - Fire: MODIS Fire
  - Air Quality: MODIS Aerosol Optical Depth (AOD)
  - Meteorological: NARR
  - Land Cover: MODIS Vegetation Index product
- Custom grid: 120X120 grid with 0.05^{o} by 0.05^{o} grid cells

Interpolation

- Spatial kriging applied to NARR and AOD datasets to estimate missing values
    - **Top:** Un-interpolated AOD
      Bottom: Interpolated AOD
    - Determines missing values based on distance-dependent variance



Data Transformation

- Yeo-Johnson Power Transformation: Mean = 0, Standard Deviation = 1
- Dimensionality Reduction
  - Principal Component Analysis
  - Autoencoding

Baseline Model

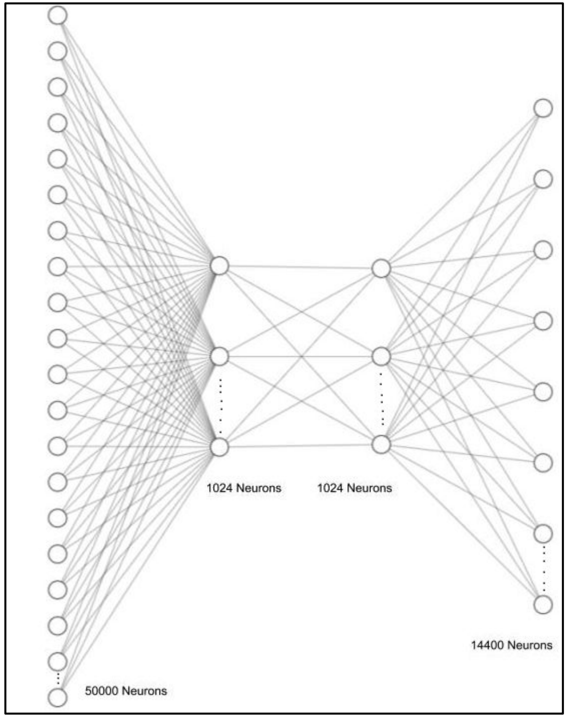
Persistence: Next day = Current day

Deep Learning Models

Basic: Artificial Neural Network
Time Series: Recurrent Neural Network

Neural Networks Overview

- Network of nodes connected by weights
- Capable of learning complex nonlinear functions
- Computationally expensive to train
- Rapidly developing field



Experiments

- **Inputs:** Fire, vegetation, weather data from previous day
  Output: AOD prediction of next day

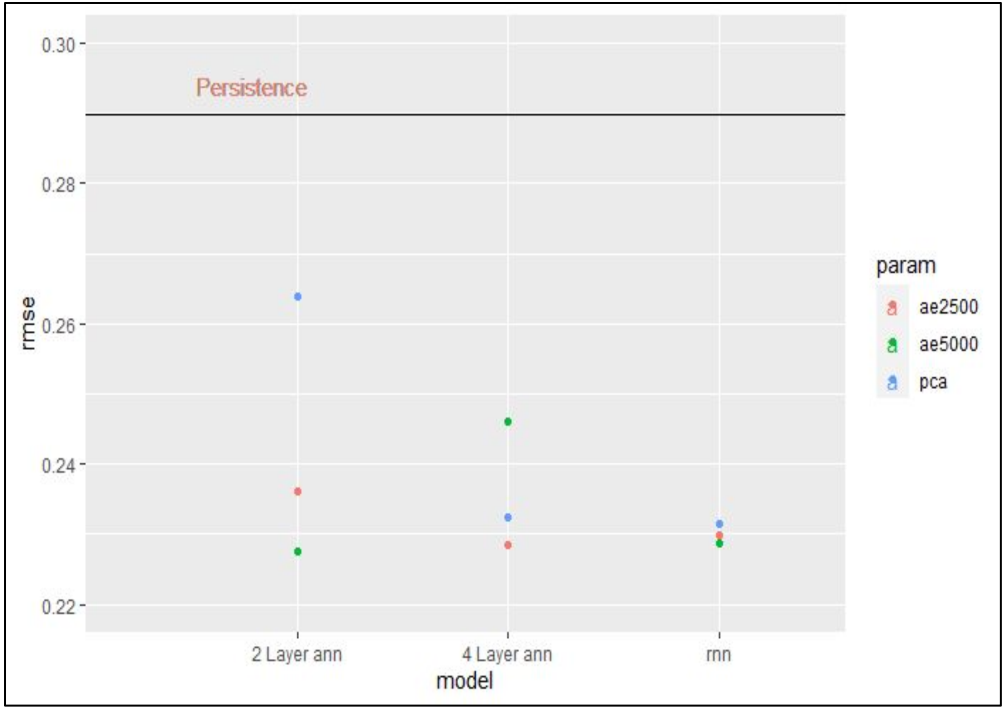
- Code in Python and R

- Autoencoding and model training on Google Cloud AI Platform

- Several hours to train each model

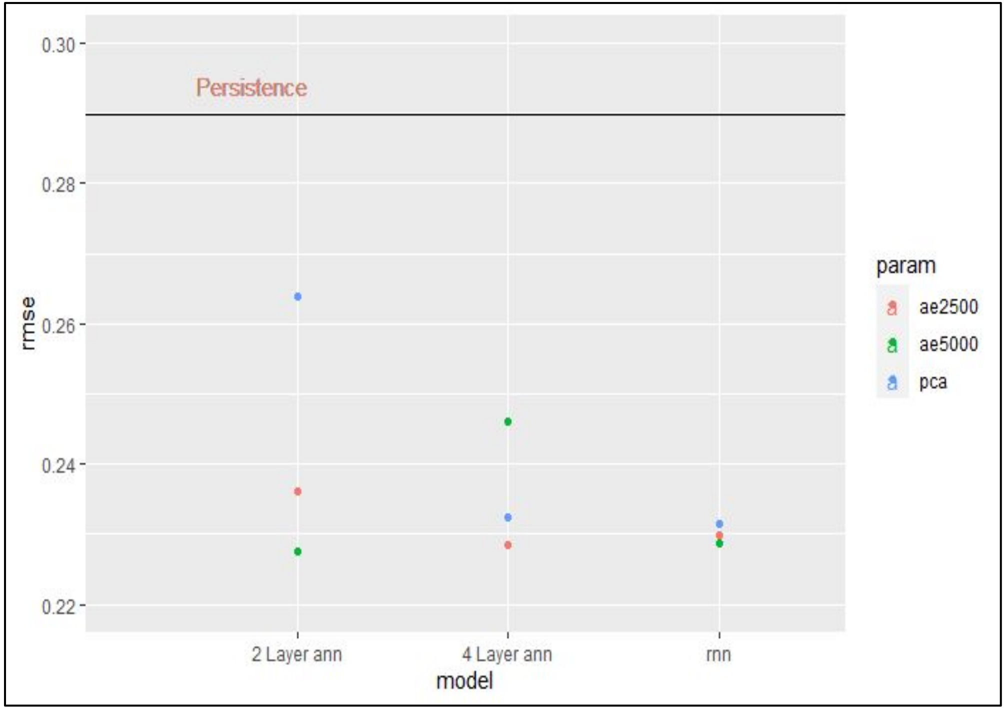
Results

- Neural networks provided improved performance compared to baseline
- Decreased variability in performance with more parameters
- Dimensionality reduction shows no clear trend



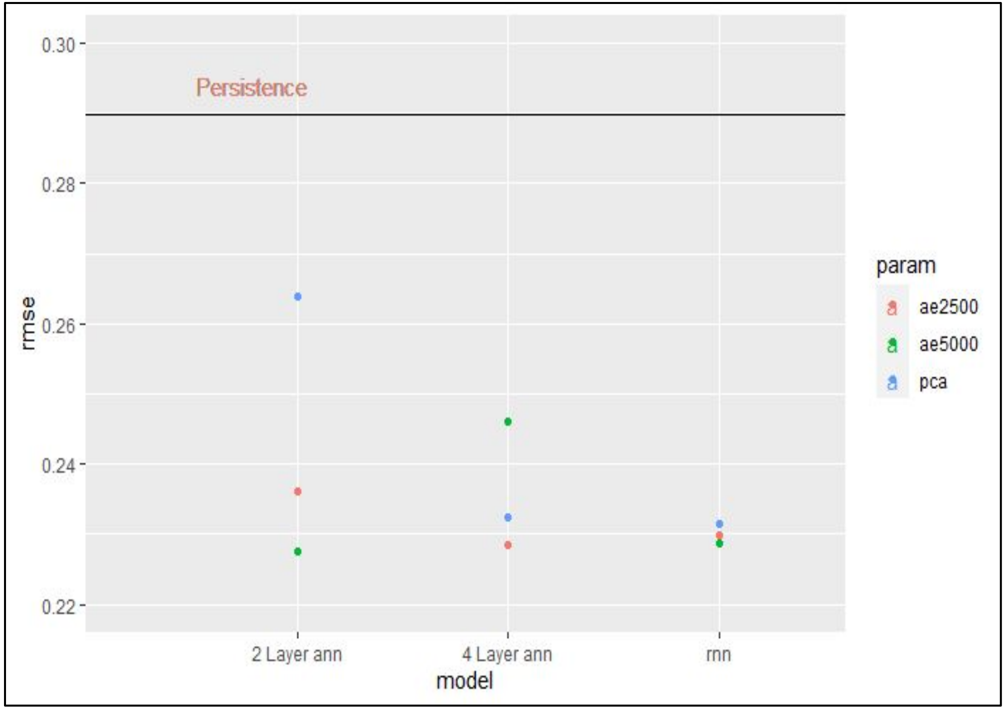
Results

- Decreased variability in performance with more parameters



Results

- Dimensionality reduction shows no clear trend

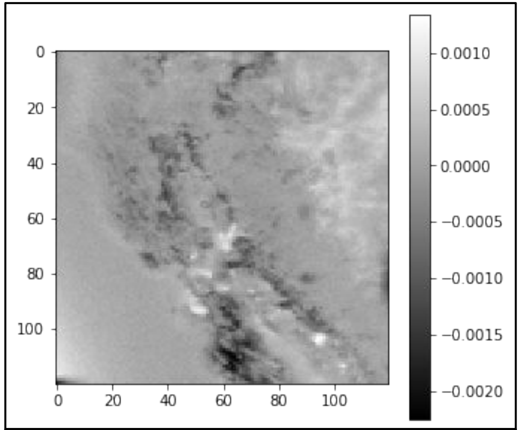


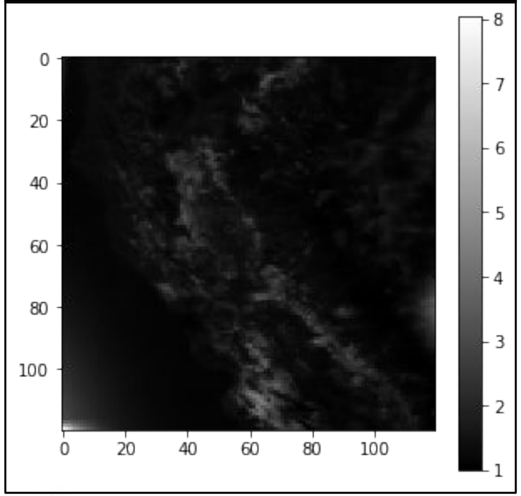
Analysis of Results

- **Top:** 2-layer ANN AOD predictions
  Bottom: AOD training labels

- Results are not state-of-the art and need refinement

- Not yet feasible to implement our models in real-world scenario

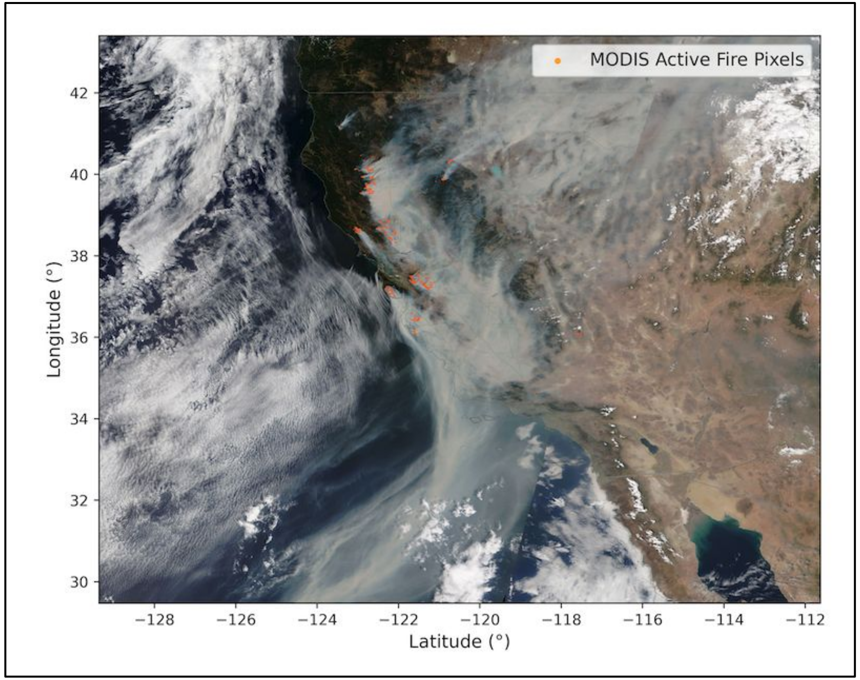




Relevance

- Predicting trends of air quality during disasters essential for damage control

- Possible use in analyzing changing trends of air quality



Future Work

- Refinement of scope and models needed for real-world use

- Further interpolation and transformation methods to improve input features

- Improve model precision and work towards publishable results

Conclusion

- Environmental data well-suited for ML applications

- Deep learning models provide improved predictions

- ML has potential for high-resolution, timely, and relevant predictions of wildfire air quality impacts

Thank you!

Any questions?

GitHub Repo: https://github.com/GAInTheHouse/BadgerX

Contacts
Gautam Agarwal, gagarwal8@wisc.edu
Eliot Kim, ejkim23@wisc.edu
Shreyans Saraogi, saraogi@wisc.edu