Advancing Earth Science Predictive Modeling through Machine Learning and Deep Learning: A Multidisciplinary Approach

Michael Martin[1,2], Lei Yang[1], Hans Moosmüller[1,2], Farnaz Hosseinpour[1,2]

1. University of Nevada, Reno, NV, USA 2. Desert Research Institute, Reno, NV, USA

Correspondence: Michael.Martin@dri.edu

Introduction

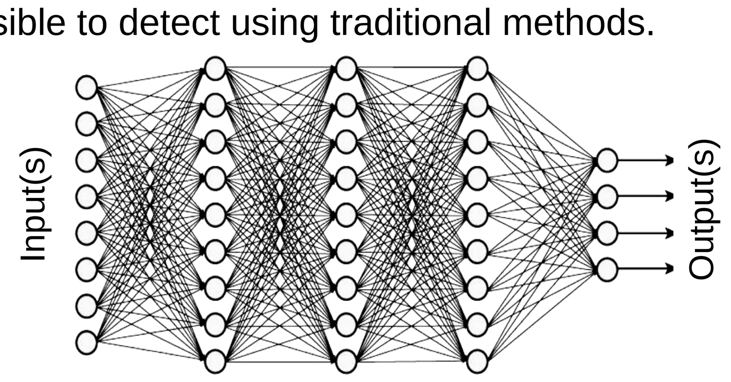
Wildfires pose an increasing threat to ecosystems, infrastructure, and public health, with smoke emissions significantly impacting air quality and climate. Fire-induced smoke emissions are influenced by numerous factors, including fire size, intensity, atmospheric conditions, and fuel bed composition. However, traditional environmental models for predicting smoke emissions face challenges such as high computational costs and inadequate parameterization of complex processes, limiting their predictive accuracy. This study leverages Artificial Intelligence to analyze vast datasets, uncover hidden patterns, and enhance the accuracy of wildfire smoke emission predictions. Using NASA's Big Data, including 24 years of historical environmental variables, we developed machine learning (ML) and deep learning models to improve wildfire smoke forecasting. Our results demonstrate that Long Short-Term Memory (LSTM)-based deep learning models outperform other ML-based approaches in predicting fire-induced smoke emissions during wildfire seasons. To further refine predictive performance, we aim to address spatial variability through the implementation of Convolutional Neural Networks. This approach will enhance the model's ability to capture spatial dependencies in wildfire-related data, ultimately improving predictive capabilities for wildfire management and environmental resilience.

Challenges

- Numerous factors and complex nonlinear interactions are involved in wildfire smoke emissions.
- High spatial/temporal variability exists in fire size/intensity and environmental conditions.
- Limitations in physical models for smoke prediction.
- High cost of computation and limited parameterization of complex processes in physical models which leads to reduced output accuracy due to above factors.

Hypothesis

Artificial Intelligence (AI) algorithms encompassing **Machine Learning** (**ML**) and **Deep Learning** (**DL**) techniques can process and analyze vast amounts of big data with numerous input variables, which allows for predicting future outcomes that may be challenging or impossible to detect using traditional methods.



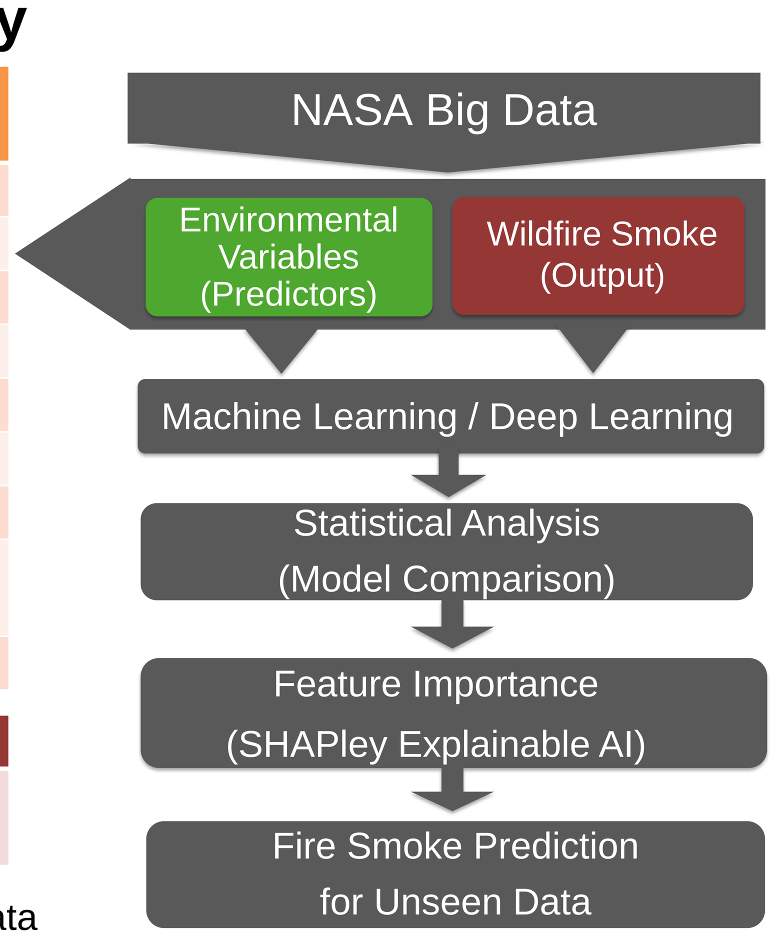
Goal/Objectives

- Develop cost-effective ML/DL-based predictive models as a high-performance alternative to traditional models
- Improve wildland fire smoke prediction in California & Nevada
- Enhance predictive accuracy and efficiency.

Data & Methodology

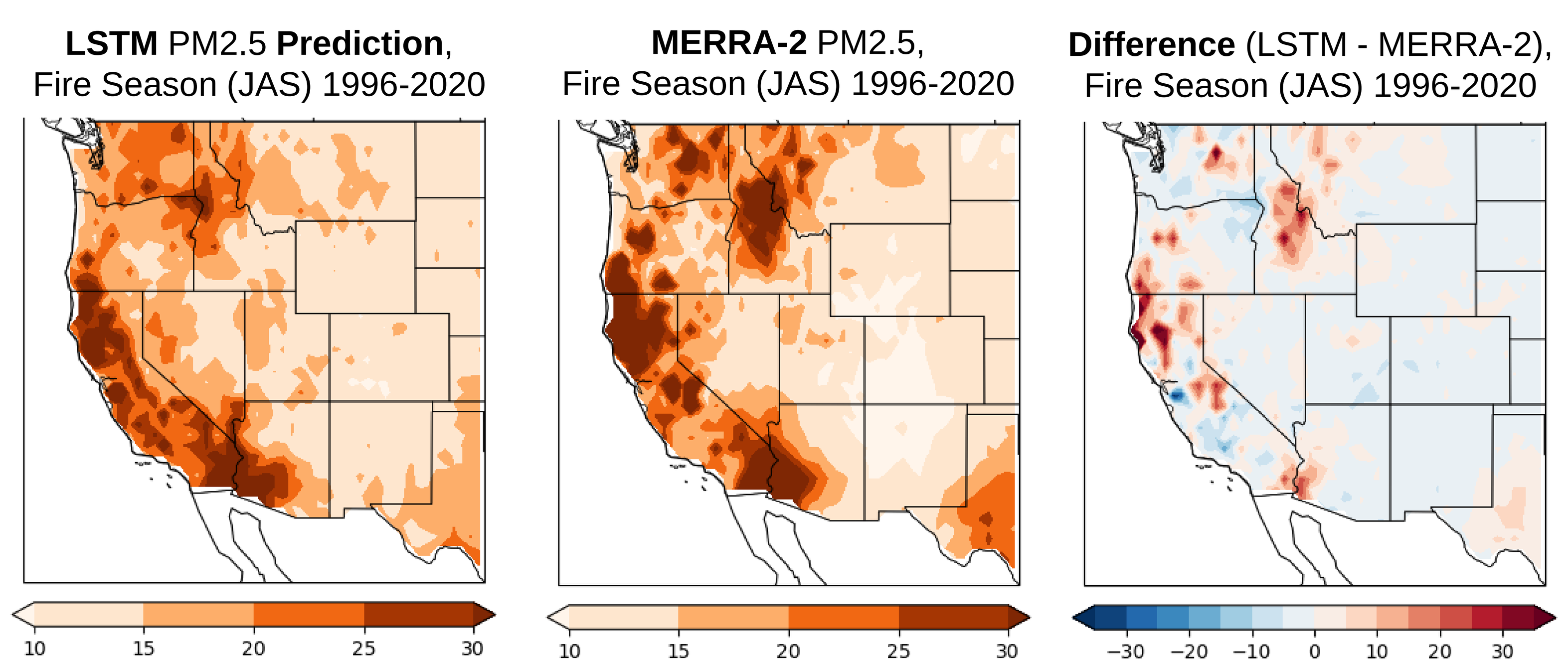
Environmental Variables (Predictors)
Vapor pressure deficit
Maximum Temperature
Potential evapotranspiration
Precipitation
Evapotranspiration
Palmer Drought Severity Index
Evaporative Drought Index
Soil moisture content (0-100 cm soil layer)
Snow water equivalent

Target Variable (Outcome)
Fire Smoke Emission: Particulate Matter (PM 2.5)



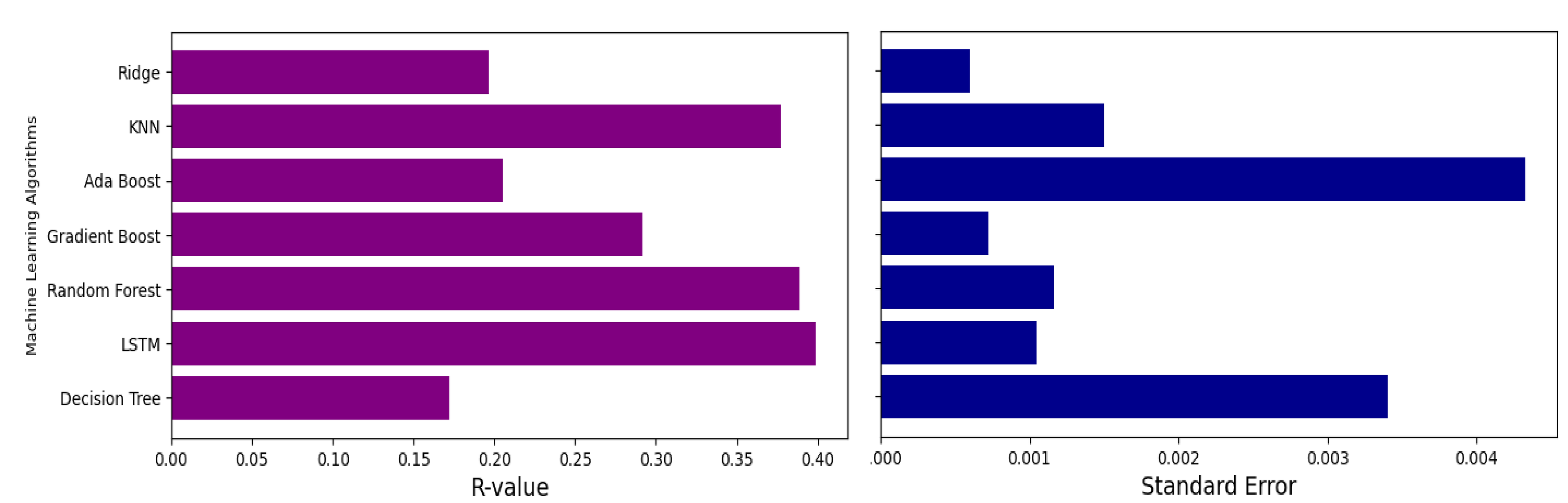
- 24 years of MERRA-2 daily data
- June-September, 1996-2020
- Western U.S., ~2x10^6 data points

LSTM Model Prediction vs. MERRA2 Reanalysis



- LSTM Predicts smoke well compared to observed/reanalysis data.
- Predictions may have positive/negative bias regarding spatial variability.

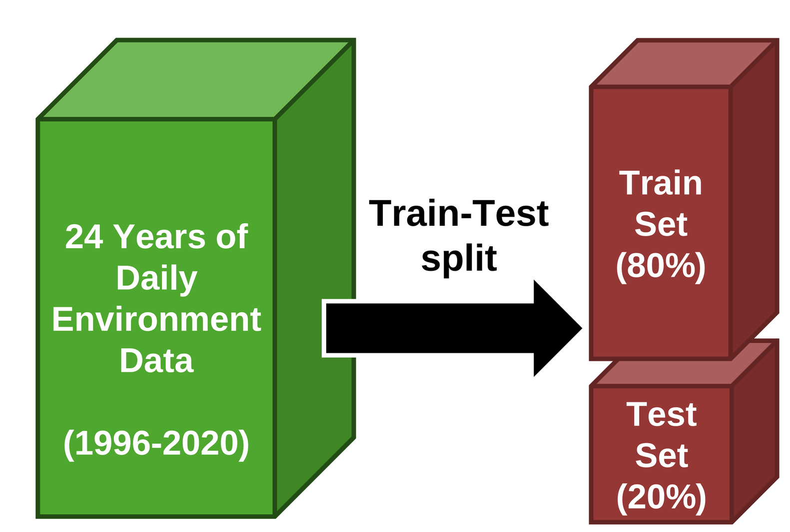
Performance Analysis for ML/DL Models



- Metrics for continuous outputs (R-value and Standard error).
- Various ML/DL to identify the ML with the highest metrics.
- Previous results showed **RF** performed better than other MLs (Martin et al., 2023).
- Recent results show **LSTM Deep Learning** outperformed RF**;** Designed for time-series.

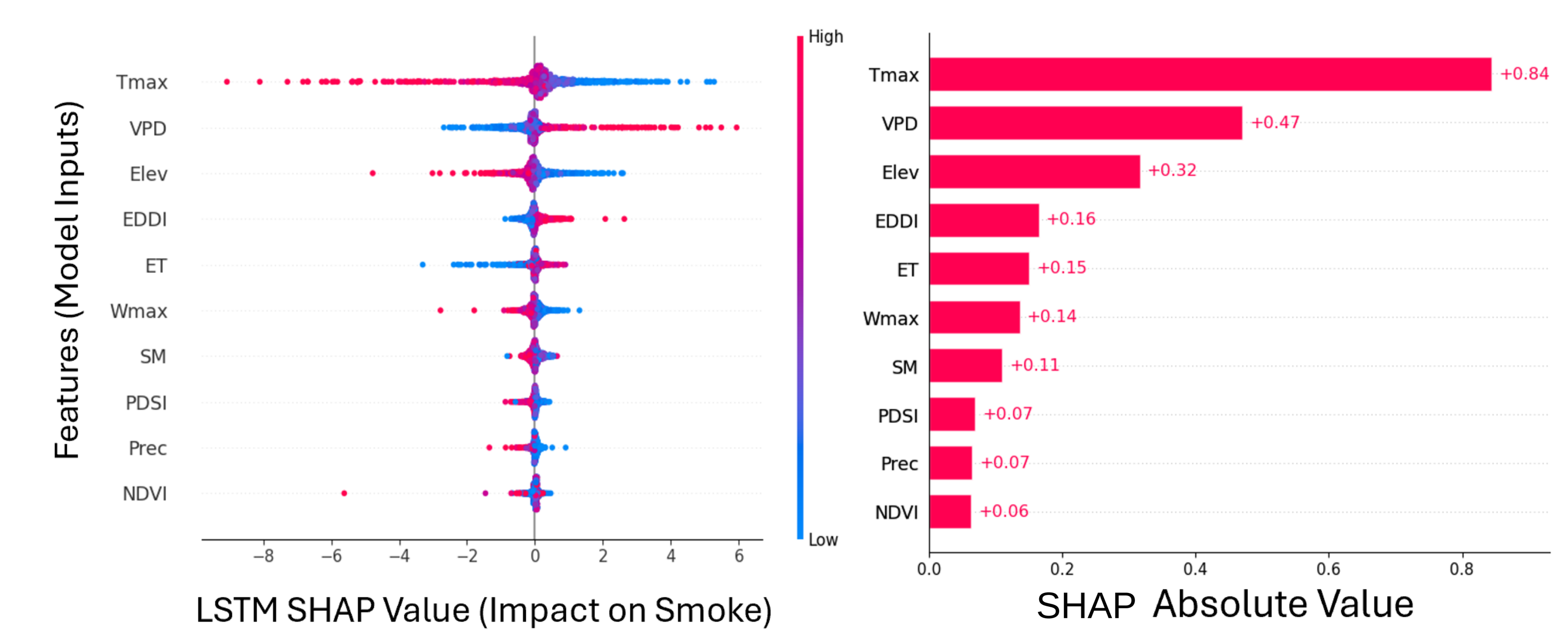
Training Machine Learning (ML) / Deep Learning (DL) Predictive Models

ML/DL Algorithms	
LSTM	**Long Short-Term Memory**
RF	Random Forest Regressor
GBoost	Gradient Boosting regressor
ABoost	Ada Boost
DT	Decision Tree Regressor
Ridge	Ridge Regressor
KNN	K nearest neighbors Regressor



- ML and Deep Learning models are developed based on the training set.
- Prediction is based on the test/validation set ➔ Prepared for unseen data.

Feature Importance: LSTM SHAPley Analysis



- SHAP Value shows effect of **each features** (Model Inputs) on **fire smoke** (Model Output)
- Higher ShAP value ➔ Higher impact on predicted output
- SHAPley Analysis shows **temperature**, **humidity**, and **elevation** are the top important factors in **fire smoke emission**

Summary

- This study leverages AI to analyze large amounts of data, uncovering patterns and relationships that may not be apparent through traditional methods.
- We utilized NASA Big Data, including 24 years of historical environmental variables, to develop ML/DL models.
- LSTM Deep Learning has high performance compared to other ML-based models we developed for predicting fire-induced smoke emissions in fire seasons.

Next steps

- We aim to address spatial variability and bias correction to enhance the ML/DL model performance.
- This includes experimentation and implementation of Convolutional Neural Networks.

Acknowledgments

This material is based upon work supported by the Nevada NASA Space Grant Consortium Undergraduate Fellowship, grant number 80NSSC20M0043.

References

- **Martin M. R**., Yang L., Hosseinpour F. (2025). Leveraging Machine Learning and NASA Satellite Data to Improve Wildfire Smoke Prediction and Environmental Insights, Data Science Conference, http://dx.doi.org/10.13140/RG.2.2.32402.36808
- **Martin M. R**., Yang L., Hosseinpour F. (2025). Harnessing Machine Learning and NASA Satellite Big Data for Enhanced Wildfire Smoke Prediction and Air Quality Forecasting, Data Science Conference, http://dx.doi.org/10.13140/RG.2.2.28208.06406
- **Martin M. R.,** (2024). Advancing Predictive Modeling in Earth Science through Artificial Intelligence, *NASA Space Grant Undergraduate Fellowship*, *Nevada NASA EPSCoR and Space Consortium*, Grant Funded May 2024. http://dx.doi.org/10.13140/RG.2.2.18624.34567.
- **Martin, M. R.**, Mehdizadeh, G., Barjeste Vaezi, R., Erfani, E., Hosseinpour, F. E. (2024). Predictive Modeling in Environmental Science with Machine Learning Algorithms. *American Meteorological Society (AMS)104th Annual Meeting*, Baltimore, MD, January 29, 2024, http://doi.org/10.13140/RG.2.2.18875.18729/2.
- Vaezi, R.B., **Martin, M. R.**, Hosseinpour, F. E. (2024). Impacts of Wildfire Smoke Aerosols on Radiation, Clouds, Precipitation, Climate, and Air Quality, *Atmospheric Environment, Elsevier*, https://dx.doi.org/10.2139/ssrn.5051411.
- Hosseinpour, F., Kumar, N., Tran, T., & Knipping, E. (2024). Using machine learning to improve the estimate of US background ozone. *Atmospheric Environment*, *316*, 120145. https://doi.org/10.1016/j.atmosenv.2023.120145