

Predicting large wildfires in the Contiguous United States using deep neural networks

Sambandh Dhal¹,^{a,*} Shubham Jain¹,^b Krishna Chaitanya Gadepally,^a Prathik Vijaykumar¹,^a Ulisses Braga-Neto¹,^a Bhavesh Hariom Sharma,^a Bharat Sharma Acharya,^c Kevin Nowka,^a and Stavros Kalafatis^a

^aTexas A&M University, Department of Electrical and Computer Engineering, College Station, Texas, United States

^bTexas A&M University, Department of Biological and Agricultural Engineering, College Station, Texas, United States

^cRodale Institute Southeast Organic Center, Chattahoochee Hills, Georgia, United States

ABSTRACT. Over the last several decades, large wildfires have become increasingly common across the United States causing a disproportionate impact on forest health and function, human well-being, and the economy. Here, we examine the severity of large wildfires across the Contiguous United States over the past decade (2011 to 2020) using a wide array of meteorological, land cover, and topographical features in a deep neural network model. A total of 4538 wildfire incidents were used in the analysis covering 87,305 square miles of burned area. We observed the highest number of large wildfires in California, Texas, and Idaho, with lightning causing 43% of these incidents. Importantly, results indicate that the severity of wildfire occurrences is highly correlated with the weather, land cover, and elevation of the study area as indicated from their SHapley Additive exPlanations values. Overall, different variants of data-driven models and their results could provide useful guidance in managing landscapes for large wildfires under changing climate and disturbance regimes.

© The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JRS.18.028501](https://doi.org/10.1117/1.JRS.18.028501)]

Keywords: climate; contiguous United States; deep neural network; land cover; large wildfire

Paper 230267G received Jun. 2, 2023; revised Mar. 20, 2024; accepted Mar. 27, 2024; published Apr. 9, 2024.

1 Introduction

In recent times, data-driven models have played a crucial role in advancing sustainability in the agriculture and forestry sector.¹⁻⁷ In the realm of forestry, various machine learning (ML) algorithms have been employed to explore aspects of forest ecology, such as species distribution models, carbon cycles, hazard assessment, and prediction.^{8,9} Wang et al.¹⁰ and Sharma et al.¹¹ showcased the use of deep learning methods, such as YOLOv4 and YOLOv5m, in forest resource investigation, vegetation coverage statistics, and plant growth monitoring. Similarly, Firebanks-Quevedo et al.¹² employed ML-based methods to formulate forestry policies and identify economic incentives for reforestation. However, a limited number of studies have been conducted to predict the spread of wildfires, a crucial aspect given the multifaceted challenges posed by wildfires, including ecological damage, deteriorating air quality, biodiversity loss, erosion, and soil degradation.

*Address all correspondence to Sambandh Dhal, sambandh@tamu.edu

Wildfires have increased fourfold over the past 40 years primarily due to fuel accumulation and fuel aridity resulting from fire suppression and climatic variability.¹³ In 2022 alone, there were 68,988 wildfires burning a total of 7.8 million acres in the United States. Approximately, 70,000 wildfires have been occurring every year over the past decade burning 7 million acres annually. Indeed, wildfires depend on ecoregions and ignition sources and are reported to cause serious repercussions on climate and ecology.¹⁴ They impair wildlife habitat, alter forest structure and composition, reduce biodiversity,¹⁵ change soil structure and watershed processes,¹⁶ and affect human values, property,¹⁵ health, and well-being.¹⁷ Recently, Burke et al.¹³ estimated that nearly 25% of PM_{2.5} across the United States results from wildfires.¹³ However, a paradigm shift in wildfire policy has been apparent in recent years to counteract long-term risks and restore ecological functionality.^{15,18} Fires and associated problems are increasingly viewed from socio-ecological lenses and different management approaches, such as prescribed fire,¹⁹ fuel treatments (mastication, thinning),²⁰ and polycentric all land management.²¹ Yet, wildfire risk assessment and modeling are challenging due to dynamic climatic variables and complex fire behavior. Improved predictive tools and approaches are, therefore, necessary for wildfire prediction and managing unprecedented fires over time and space scales. Much progress has been made in using artificial neural networks, particularly multilayer perceptron in predicting wildfires, but studies focusing on the use of deep neural networks (DNN) in predicting wildfire spread are generally few. DNN, such as convolutional neural networks and recurrent neural networks (in particular, long short-term memory networks), are deep learning methods that have multiple non-linear hidden networks and have been successfully applied in detecting wildfires from satellite observations²² or predicting wildfire spread using meteorological variables, such as wind, temperature, and humidity.²³ However, many such studies are limited to small spatial and temporal scales. In this short communication paper, we examine the severity of large wildfires across the Contiguous United States over the past decade (2011 to 2020) using a wide array of meteorological, land cover, and topographical features in the DNN model. Here, large wildfires are used to refer to the areas burned being greater than 500 acres in the Eastern and 1000 acres in the Western United States. The data-driven approaches in this paper will be instrumental in understanding different factors influencing the occurrence and severity of wildfires and thereby facilitating wildfire management and policies.

2 Materials and Methods

The study area comprises the Contiguous United States (CONUS), which is divided into 11 Level I Ecoregions and 967 Level IV sub-Ecoregions.²⁴ The western regions of the study area typically experience a higher number of wildfire incidents and encompass larger burned areas compared with the western United States²⁵ due largely to the heterogeneity in the landscape caused by human development and fragmentation of forest land cover.¹⁴ The GIS data for wildfire locations and burned area boundaries were obtained from the Monitoring Trends in Burn Severity (MTBS) program.^{26,27} The program assesses the frequency, extent, and magnitudes of all large wildland fires in the United States. The thresholds for large wildfires are set to greater than 1,000 acres in the western United States and 500 acres in the Eastern United States. A period of 10 years between 2011 and 2020 was selected for analysis, and the “prescribed wildfires” were removed from the dataset. A total of 4,538 wildfire incidents were used in the analysis covering 87,305 square miles of burned area. Additionally, a 1992-2015 spatial wildfire occurrence dataset²⁸ was used to analyze large wildfires.²⁹ In order to identify potential wildfire hotspots, the number of wildfire occurrences and burned areas were also evaluated within each Level IV ecoregion. Figure 1 shows the point locations for the occurrence of large wildfires and potential wildfire hotspots between 2011 and 2020 in the Contiguous United States.

Meteorological variables were obtained from different sources (Table 1) for wildfire prediction. Briefly, monthly climate attributes including total monthly precipitation, mean monthly temperature, and maximum and minimum vapor pressure deficit were obtained from the PRISM dataset. The Palmer Drought Severity Index (PDSI) was obtained from GRIDMET to infer the relative dryness in the region. The index typically ranges from -10 (dry) to +10 (wet).³⁶ The land cover data was obtained from the National Land Cover Dataset (NLCD). The 30-meter NLCD raster for year 2016 was used to obtain land cover percentages around a 4-kilometer buffer at the

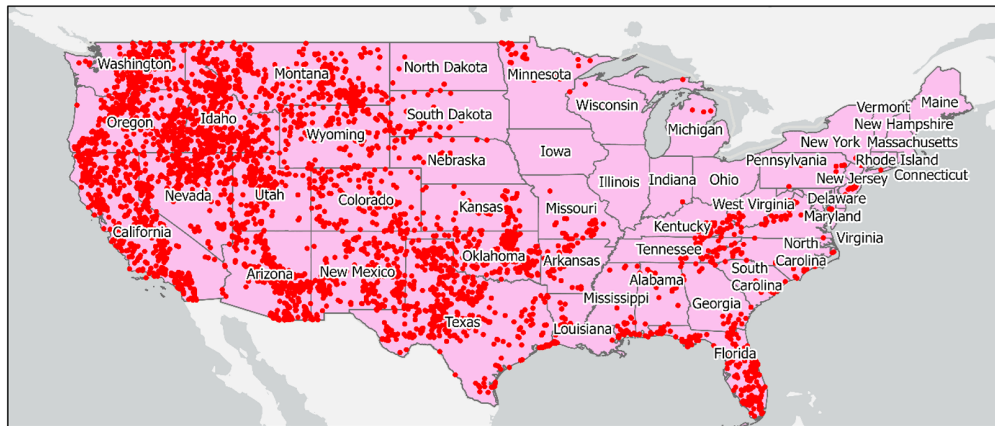


Fig. 1 Large wildfire incidents in the contiguous United States between 2011 and 2020.

Table 1 List of datasets used in the study to model burned areas in large US wildfires.

Category	Dataset	Variables	Source	Resolution
Climate	PRISM	Precipitation, temperature, vapor pressure deficit (min, max)	Ref. 30	4000 m gridded, monthly
	GRIDMET	PDSI, PET	Ref. 31	4000 m gridded, 5-day (PDSI), 1-day (PET)
Land cover	NLCD, 2016	Open water, developed, barren, forests, shrub/scrub, hay/pasture, cultivated crops, wetlands	Ref. 32	30 m gridded
MODIS	MOD13A3 Version 6	NDVI, EVI	Ref. 33	1000 m gridded, monthly
Topography	USGS DEM	Elevation (m)	Ref. 34	100 m gridded
Ecoregion boundaries	US EPA ecoregions	Level I and level IV ecoregions	Ref. 35	Shapefile

point of wildfire occurrence. The 4-kilometer radius was selected based on the mean burned area of all wildfires in the dataset to represent the amount of forest and shrubland available near the fire area that could potentially increase the extent of wildfires. The relationship between land cover and wildfires was examined using NLCD land cover classes within the 4,538 burned area boundaries. Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) within 1 kilometer resolution, were obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite dataset (MOD13A3). Elevation data was obtained from the United States Geological Survey (USGS) Digital Elevation Model (DEM) dataset at 100-meter spatial resolution. All these datasets were spatially and temporally linked to each of the 4,538 wildfires that occurred in the contiguous US between 2011 and 2020 using R 4.3.0 and ArcGIS (Version 10.2).

Different datasets in Table 1 were analyzed using ML models. The best model was selected using the lowest testing Mean Absolute Error (MAE) criterion. Consequently, as observed in Table 2, a DNN model was trained to predict wildfires based on climatological and geological attributes surrounding the point of wildfire origin. The features used in the DNN model to predict large wildfire burned areas are shown in Table 3. Keras and TensorFlow libraries in Python were used to design the deep-learning approach. The dataset was divided into training and testing sets using an 80/20 split. Thus, 80% of the data were used for training and validation and 20% for testing the accuracy of the models. Further, the data was split three times to generate multiple random samples of training and test data to evaluate the accuracy over multiple test set

Table 2 Comparison of the regression models used in predicting spread of large wildfires.

Name of the ML/DL algorithm	Specifications of the model	MAE
Polynomial regression	Degree = 3	0.83
Support vector regression	Kernel = radial basis function, penalty parameter (C) = 0.1 to 100 (using gridsearch to find the best parameter)	0.76
Decision tree regression	Maximum depth = 5	0.72
Random forest regression	Number of trees = 100, maximum depth = 5	0.63
Gradient boosting regression	Number of estimators = 100, learning rate = 0.25, maximum depth = 5	0.7
Four-layered DNN	Number of neurons used in the model = 512, 256, 64, 16, and 1, respectively, Activation function = ReLU, Learning rate = 0.01, No. of epochs = 200	0.055 to 0.06

Table 3 Features used in the DNN model to predict large wildfire burn area with minimum and maximum values in the dataset.

Feature	Description	Min	Max
LATITUDE	Latitude coordinates of wildfire occurrence (decimal degrees)	25.2	49
LONGITUDE	Longitude coordinates of wildfire occurrence (decimal degrees)	-124.1	-72.8
DOY	Wildfire ignition day of year	1	365
ppt	Total monthly precipitation for month of wildfire ignition	0	1063.2
tmean	Average monthly temperature for month of wildfire ignition	-5.3	36.8
vpdmax	Maximum vapor pressure deficit for month of wildfire ignition	2.7	81.8
vpdmin	Minimum vapor pressure deficit for month of wildfire ignition	0	35.3
PDSI	Palmer drought severity index during ignition date	-8.1	7.6
Developed	% NLCD developed around 4-km buffer of wildfire ignition	0	64.2
Forests	% NLCD forests around 4-km buffer of wildfire ignition	0	99.8
Shrub	% NLCD shrub/scrub around 4-km buffer of wildfire ignition	0	100
grass	% NLCD grasslands/herbaceous around 4-km buffer of wildfire ignition	0	100
Pasture	% NLCD hay/pasture around 4-km buffer of wildfire ignition	0	74
Wetlands	% NLCD wetlands around 4-km buffer of wildfire ignition	0	100
NDVI	Normalized difference vegetation index for month of wildfire occurrence	0.1	0.9
EVI	Enhance vegetation index for month of wildfire occurrence	0	0.7
Elevation	Elevation of wildfire occurrence	-2	3507

combinations. Wildfires with missing attributes were removed from the study resulting in a total of 4536 observations for model development. Prior to being used in the DNN model, the wildfire acres were log-transformed to account for any skewness in the observed data and to normalize the target distribution.

Different features in Table 3 were transformed using a standard scalar and were fed as inputs to a DNN model with five layers. The DNN layers had 512, 256, 64, 16, and 1 neuron, respectively. ReLU was used as the activation function for each of the five DNN layers. The DNN model was trained using root mean square optimizer and 0.01 learning rate. Callbacks were used to monitor validation loss. Mean squared error was utilized as the loss function and MAE was used as a performance metric. The model was trained for 200 epochs with a batch size of 32 and a validation split of 0.2. For each of the three values of random seed that was used for generating train and test sets, plots for training loss and validation loss were convex in nature as shown in Fig. 2. A schematic depicting the proposed deep learning framework is given in Table 4. The error rate for the test data was determined using the equation below:

$$\text{Error rate(MAE)} = \frac{\sum_1^N |y_{\text{obs}} - y_{\text{pred}}|}{\sum_1^N |y_{\text{obs}}|}$$

SHapley Additive exPlanations (SHAP) values were used to determine the impact (positive or negative) of each model feature on the burned area. SHAP is a surrogate explanation method for ML models, which computes values that quantify the contribution of each feature to a prediction based on cooperative game theory.³⁷ Thus, SHAP values could be used in interpreting the DNN model and determining the potential drivers of wildfire. For each data point, the model predicted value equals the sum of all feature SHAP values and the average prediction.

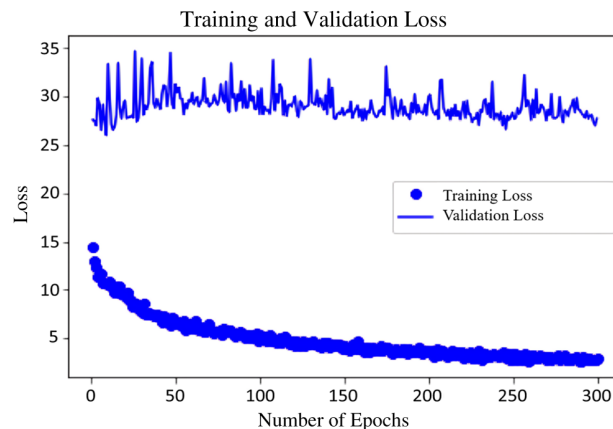


Fig. 2 Training and validation losses for the proposed deep learning framework.

Table 4 Schematic of the proposed deep learning framework.

Layer type	Number of neurons	Number of parameters
Dense layer 1	128	1920
Dense layer 2	64	8256
Dense layer 3	16	1040
Dense layer 4	1	17
Total number of parameters:		11,233
Trainable parameters:		11,233
Non-trainable parameters:		0

A positive SHAP value indicates an increase in the predicted value due to the feature, whereas negative SHAP values indicate a decrease in the predicted value.

3 Results and Discussion

Wildfires are natural or human-induced events occurring in forests, grasslands, and prairies driven by ignition, fuel, droughts, and conducive weather conditions.³⁸ The distribution of total large wildfires by states and potential causes is shown in Fig. 3. The highest number of large wildfires between 2011 and 2020 occurred in California (448 incidents), followed by Texas (434 incidents), and Idaho (426 incidents). About 43% of large wildfires were caused by lightning, followed by “miscellaneous” (18%), unidentified (10%), arson (9%), equipment use (8%), and debris burning (6%). Importantly, our data exclude small wildfires (500 acres) that are more frequent and are caused largely by human activities.³⁹ The percentage of burned area per level IV ecoregion illustrates the severity of wildfires in various ecosystems (Fig. 4). The area consumed by wildfires was higher in Mediterranean California, the Marine West Coast Forest, and North American Desserts, and smaller in Northern and Eastern Temperate Forests (Fig. 4). Most of these burned areas were grassland, forest, and shrub/scrub land covers (Fig. 5). The mean absolute SHAP values for grassland, forest, and shrub cover were 0.6, 0.43, and 0.35, respectively (Fig. 6), indicating their predominant positive role in wildfire spread. It was also observed

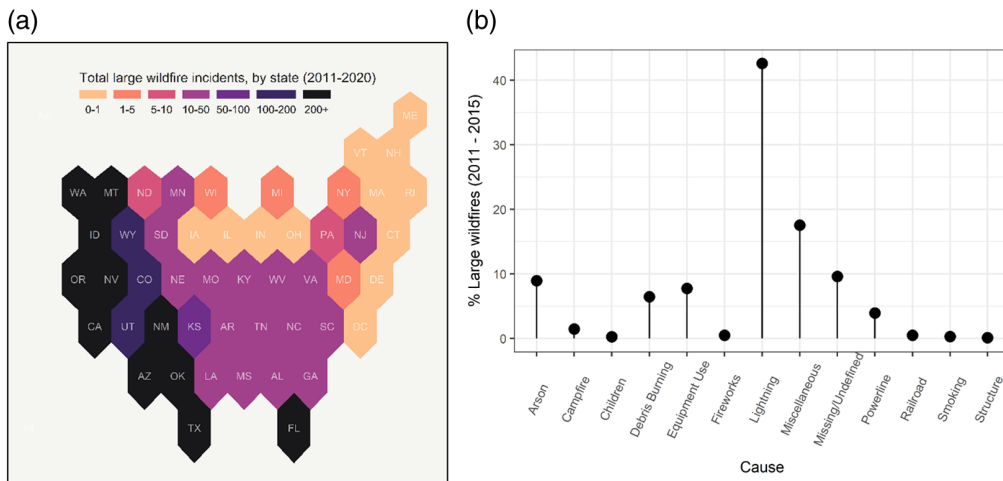


Fig. 3 (a) Average annual large wildfire incidents by states and (b) cause of large wildfires (> 500 acres) in the contiguous United States between 2011 and 2015.

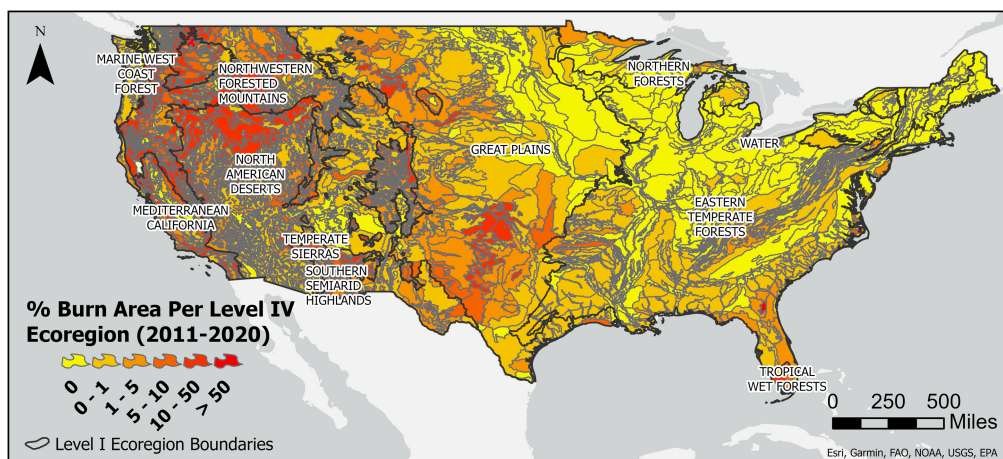


Fig. 4 Percent of level IV ecoregion land burned in large wildfires between 2011 and 2020.

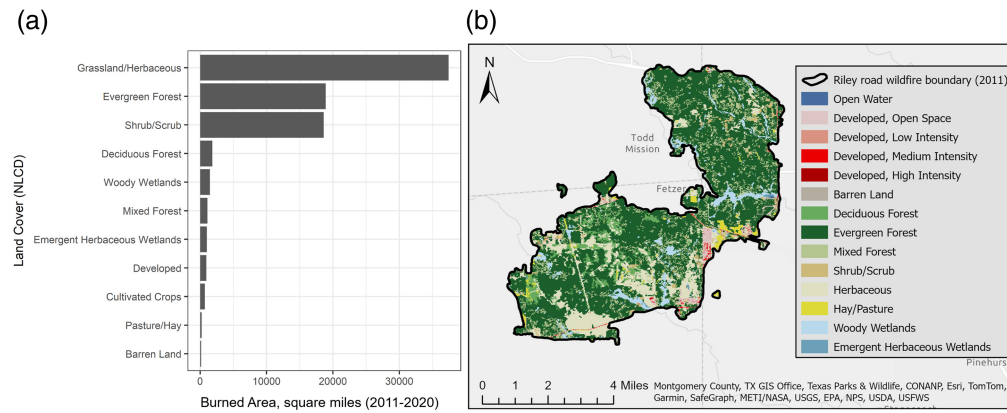


Fig. 5 (a) Burned area by NLCD land cover in large wildfires between 2011 and 2020 and (b) an example of NLCD land cover within the burned area in the September 2011 Riley Road wildfire northwest of Houston burning 19,000 acres of land.

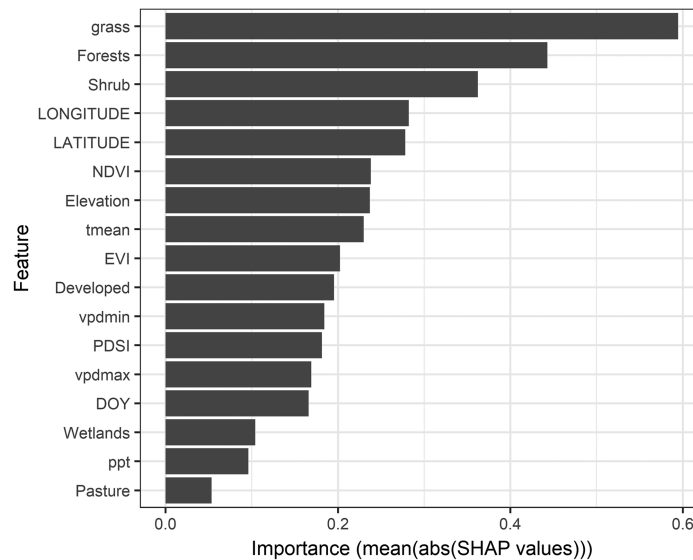


Fig. 6 Feature importance in the DNN model obtained from SHAP values.

that the highest number of wildfires occurred in July and August, which are typically the hottest and driest months. Temperatures in these months were $\sim 21^{\circ}\text{C}$ and 24°C , respectively (Fig. 7). Indeed, warmer temperatures and extended droughts may exacerbate the vulnerability of forests and the occurrence of wildfire events. The climatic dependency of wildfire behavior and spread further highlights the importance of managing fuel and restoring ecology in combating fire hazards and associated impacts.⁴⁰

Here, several ML models, such as polynomial regression,⁴¹ support vector regression,⁴² decision tree regression,⁴³ random forest regression,⁴⁴ gradient boosting regression,⁴⁵ and a DNN model, were utilized to predict wildfires occurrence based on climatological and geological features. Only a few studies have attempted to utilize ML models in wildfire studies. For example, Zhang et al.⁴⁶ compared four multilayer perceptron and CNN architectures in wildfire modeling and reported the highest accuracy in predicting seasonal peaks in fire activity and vulnerable areas with CNN-2D, a DNN model. Langford et al.⁴⁷ used DNN to detect wildfire events in Alaska for the wildfire year 2004 and highlighted the utility of the validation-loss weight selection approach for accurately mapping wildfire on an imbalanced dataset. In another study, deep neural computing optimized by using adaptive moment estimation algorithms showed the highest accuracy in forest fire prediction compared with stochastic gradient descent, root mean square

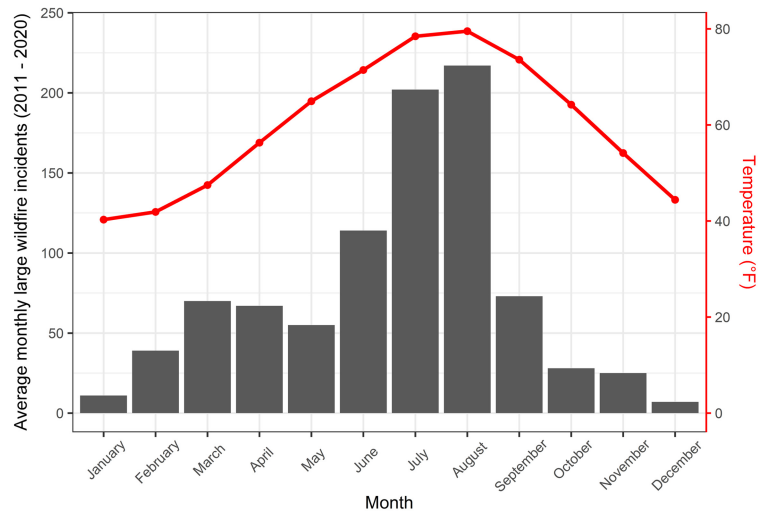


Fig. 7 Plot showing the relationship between average monthly large wildfires (primary y-axis) in the contiguous US and the mean monthly temperature (line).

propagation, and Adadelta optimizers.⁴⁸ In our model, for test sets generated in each of the three values of random seed, the MAE was found to be between 0.055 and 0.06. This lower value of MAE indicates a higher accuracy of wildfire prediction.

The land cover classes around a 4-km buffer at the point of occurrence including the percentage of grasslands/herbaceous, percentage of forests, and percentage of shrublands were found to be the most influential in predicting wildfire burned area within a 4 km radius of the point of wildfire occurrence. Fire activities in such locations are largely associated with fuel loads and flammability. Fuels in grasslands are generally dry, which could easily and rapidly spread fires.⁴⁹ The location of wildfires, as represented by latitude, was also important in predicting burned areas. Indeed, precipitation regimes vary with latitude-longitudes, with lower latitudes exhibiting reduced rainfall and moisture, and drier conditions.

A non-linear relationship existed between features and their impact on the predicted burned area (Fig. 8), consistent with many other global studies.⁵⁰ Also, the predicted burned areas

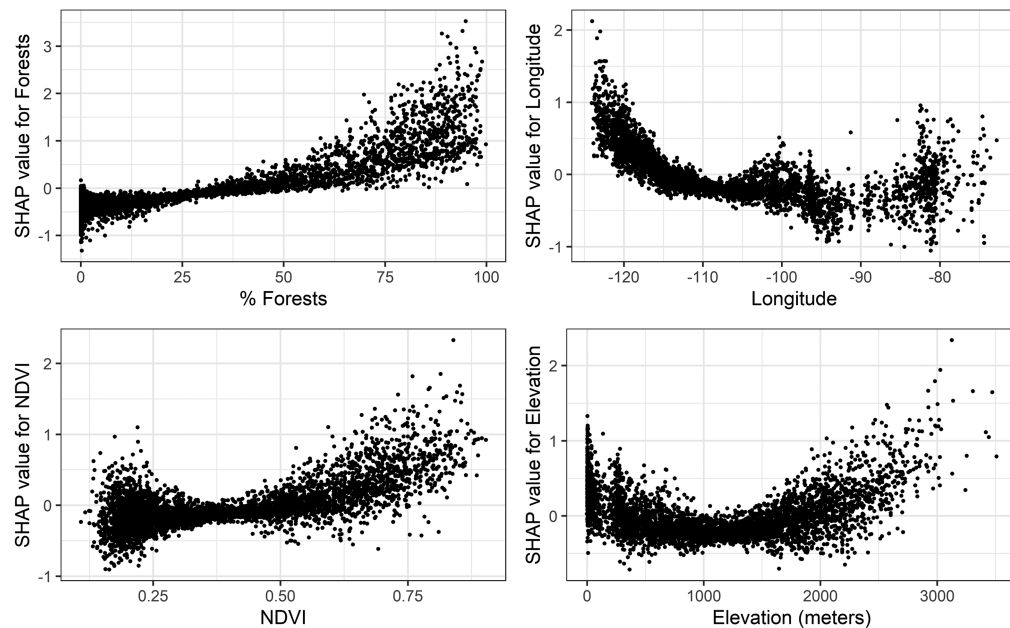


Fig. 8 Partial dependence plots showing the interactions between features and burn area using SHAP values.

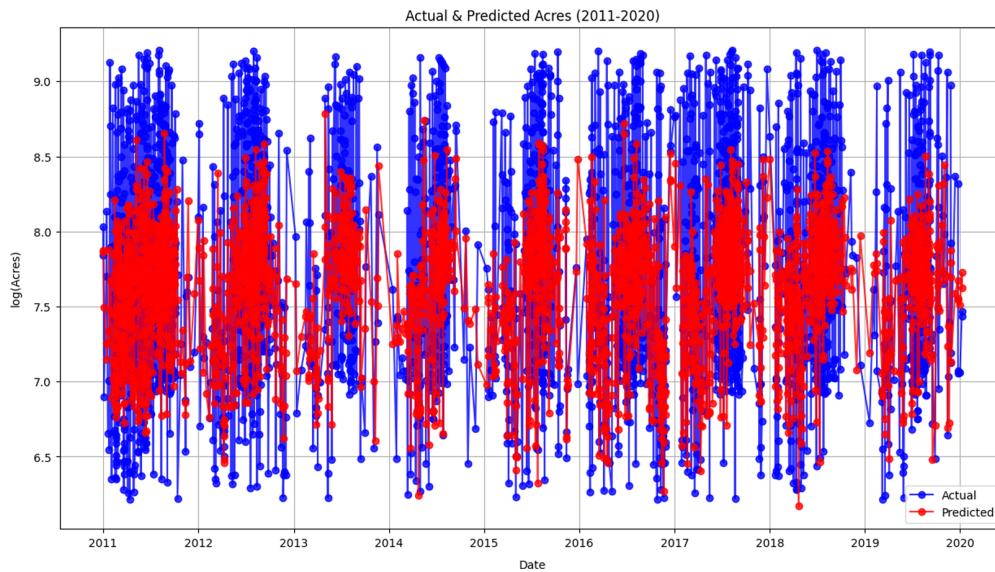


Fig. 9 Plot showing the actual burned area and the predicted burned area over the contiguous United States (2011 to 2020).

exhibited a trend closely resembling that of the actual burned areas (Fig. 9). Large forest cover within a 4-km buffer zone surrounding the point of wildfire occurrence had a large positive impact on the burned area. A forest cover of 30% or more increased the predicted burned area above the mean. More western longitudes presented a significant increase in the burned area. However, higher elevations had positive SHAP values indicating larger burned areas in regions with higher elevations. In general, fire activities are higher in steeper areas.⁴⁹ In the western United States, Westerling et al.⁴⁰ observed the greatest wildfires in the mid-elevation range, occurring mostly as episodic events. These events were further associated with spring snowmelt timing. Topographic features may, however, develop decisively in fire spread when burning conditions are rather less extreme.⁵¹ Finally, we observed that values of NDVI greater than 0.5, indicating vigorous green vegetation, had a positive impact on the burned area, but NDVI values less than 0.5, indicating sparse vegetation, had no net effect on the burned area.

4 Conclusion

This study analyzed and predicted the large wildfires across the contiguous United States from 2011 to 2020. Results showed that the highest number of large wildfires and areas consumed by wildfires occurred in California. Also, wildfires occurred mostly during July and August months. A comparison of different models showed that a four-layered DNN model outperformed other ML models. Further, land cover and the location (latitude and longitude) of wildfire occurrence were most likely to determine the severity and extent of wildfires in the United States as inferred from their SHAP values. Indeed, predictive models utilizing ML and remote sensing tools, climate, and geospatial data are useful in understanding wildfire complexity and predicting and mitigating fire hazards. However, additional features, such as soil characteristics and 100-h fuel moisture, could be integrated into the DNN model to improve model accuracy and prediction.

Code and Data Availability

All data gathered or analyzed in this study are included in the article. Raw data may be available upon appropriate request to the corresponding author.

Disclaimer

The authors are responsible for the views expressed in this paper and do not necessarily represent or reflect the views and policies of the universities. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was presented in the 2023 TAMIDS Data Science Competition. We extend our sincere thanks to different scientists and researchers for their valuable comments and suggestions that helped in designing and improving this paper.

References

1. C. S. Arvind et al., "Edge computing based smart aquaponics monitoring system using deep learning in IoT environment," in *IEEE Symp. Ser. Comput. Intell. (SSCI)*, IEEE, pp. 1485–1491 (2020).
2. A. Dutta et al., "IoT based aquaponics monitoring system," in *1st KEC Conf. Proc.*, Vol. 1, pp. 75–80 (2018).
3. S. B. Dhal et al., "A machine-learning-based IoT system for optimizing nutrient supply in commercial aquaponic operations," *Sensors* **22**(9), 3510 (2022).
4. S. B. Dhal et al., "Nutrient optimization for plant growth in aquaponic irrigation using machine learning for small training datasets," *Artif. Intell. Agric.* **6**, 68–76 (2022).
5. S. B. Dhal et al., "Can machine learning classifiers be used to regulate nutrients using small training datasets for aquaponic irrigation? A comparative analysis," *PLoS One* **17**(8), e0269401 (2022).
6. S. Mahanta, M. R. Habib, and J. M. Moore, "Effect of high-voltage atmospheric cold plasma treatment on germination and heavy metal uptake by soybeans (glycine max)," *Int. J. Mol. Sci.* **23**(3), 1611 (2022).
7. S. B. Dhal et al., "An IoT-based data-driven real-time monitoring system for control of heavy metals to ensure optimal lettuce growth in hydroponic set-ups," *Sensors* **23**(1), 451 (2023).
8. Z. Liu et al., "Application of machine-learning methods in forest ecology: recent progress and future challenges," *Environ. Rev.* **26**(4), 339–350 (2018).
9. L. Xiangdong, "Applications of machine learning algorithms in forest growth and yield prediction," *J. Beijing For. Univ.* **41**(12), 23–36 (2019).
10. Y. Wang et al., "Recent advances in the application of deep learning methods to forestry," *Wood Sci. Technol.* **55**(5), 1171–1202 (2021).
11. S. Sharma et al., "Drones and machine learning for estimating forest carbon storage," *Carbon Res.* **1**(1), 21 (2022).
12. D. Firebanks-Quevedo et al., "Using machine learning to identify incentives in forestry policy: towards a new paradigm in policy analysis," *For. Policy Econ.* **134**, 102624 (2022).
13. M. Burke et al., "The changing risk and burden of wildfire in the United States," *Proc. Natl. Acad. Sci. U. S. A.* **118**(2), e2011048118 (2021).
14. B. D. Malamud, J. D. Millington, and G. L. Perry, "Characterizing wildfire regimes in the United States," *Proc. Natl. Acad. Sci. U. S. A.* **102**(13), 4694–4699 (2005).
15. M. Moritz et al., "Learning to coexist with wildfire," *Nature* **515**(7525), 58–66 (2014).
16. G. G. Ice, D. G. Neary, and P. W. Adams, "Effects of wildfire on soils and watershed processes," *J. For.* **102**(6), 16–20 (2004).
17. R. Xu et al., "Wildfires, global climate change, and human health," *N. Engl. J. Med.* **383**(22), 2173–2181 (2020).
18. T. A. Steelman and C. A. Burke, "Is wildfire policy in the United States sustainable?" *J. For.* **105**(2), 67–72 (2007).
19. C. A. Kolden, "We're not doing enough prescribed fire in the Western United States to mitigate wildfire risk," *Fire* **2**(2), 30 (2019).
20. E. D. Reinhardt et al., "Objectives and considerations for wildland fuel treatment in forested ecosystems of the interior western United States," *For. Ecol. Manage.* **256**(12), 1997–2006 (2008).
21. E. C. Kelly, S. Charnley, and J. T. Pixley, "Polycentric systems for wildfire governance in the Western United States," *Land Use Policy* **89**, 104214 (2019).
22. J. Yao et al., "Predicting the minimum height of forest fire smoke within the atmosphere using machine learning and data from the CALIPSO satellite," *Remote Sens. Environ.* **206**, 98–106 (2018).
23. P. Cortez and A. D. J. R. Morais, "A data mining approach to predict forest fires using meteorological data," <https://hdl.handle.net/1822/8039> (2007).
24. J. M. Omernik, "Ecoregions of the conterminous United States," *Ann. Assoc. Am. Geogr.* **77**(1), 118–125 (1987).
25. R. C. Nagy et al., "Human-related ignitions increase the number of large wildfires across US ecoregions," *Fire* **1**(1), 4 (2018).
26. J. Eidenshink et al., "A project for monitoring trends in burn severity," *Fire Ecol.* **3**(1), 3–21 (2007).
27. <https://data.fs.usda.gov/geodata/edw/datasets.php?xmlKeyword=Burn>
28. K. C. Short, "Spatial wildfire occurrence data for the United States, 1992-2015," [FPA_FOD_20170508] (2017).
29. https://data.fs.usda.gov/geodata/edw/edw_resources/meta/S_USA.FPA_FOD_4thedition.xml.

30. <https://prism.oregonstate.edu/>.
31. <https://www.climatologylab.org/gridmet.html>.
32. <https://www.mrlc.gov/>.
33. <https://lpdaac.usgs.gov/products/mod13a3v006/>.
34. <https://earthworks.stanford.edu/catalog/stanford-zz186ss2071>.
35. <https://www.epa.gov/eco-research/ecoregions>.
36. W. M. Alley, "The Palmer drought severity index: limitations and assumptions," *J. Appl. Meteorol. Climatol.* **23**(7), 1100–1109 (1984).
37. S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Adv. Neural Inf. Process. Syst.* **30** (2017).
38. J. G. Pausas and J. E. Keeley, "Wildfires and global change," *Front. Ecol. Environ.* **19**(7), 387–395 (2021).
39. J. P. Prestemon and J. P. Prestemon, *Wildfire Ignitions: A Review of the Science and Recommendations for Empirical Modeling*, p. 24, US Department of Agriculture, Forest Service, Southern Research Station, Asheville, North Carolina (2013).
40. A. L. Westerling et al., "Warming and earlier spring increase western US forest wildfire activity," *Science* **313**(5789), 940–943 (2006).
41. E. Ostertagová, "Modelling using polynomial regression," *Procedia Eng.* **48**, 500–506 (2012).
42. M. Awad et al., "Support vector regression," in *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, pp. 67–80 (2015).
43. M. Xu et al., "Decision tree regression for soft classification of remote sensing data," *Remote Sens. Environ.* **97**(3), 322–336 (2005).
44. L. Breiman, "Random forests," *Mach. Learn.* **45**, 5–32 (2001).
45. G. Ke et al., "LightGBM: a highly efficient gradient boosting decision tree," in *Adv. Neural Inf. Process. Syst.* **30** (2017).
46. G. Zhang, M. Wang, and K. Liu, "Deep neural networks for global wildfire susceptibility modeling," *Ecol. Indic.* **127**, 107735 (2021).
47. Z. Langford, J. Kumar, and F. Hoffman, "Wildfire mapping in Interior Alaska using deep neural networks on imbalanced datasets," in *IEEE Int. Conf. Data Mining Workshops (ICDMW)*, IEEE, pp. 770–778 (2018).
48. H. Van Le et al., "A new approach of deep neural computing for spatial prediction of wildfire danger at tropical climate areas," *Ecol. Inf.* **63**, 101300 (2021).
49. I. Stavi, "Wildfires in grasslands and shrublands: a review of impacts on vegetation, soil, hydrology, and geomorphology," *Water* **11**(5), 1042 (2019).
50. L. M. Li et al., "Artificial neural network approach for modeling the impact of population density and weather parameters on forest fire risk," *Int. J. Wildland Fire* **18**(6), 640–647 (2009).
51. M. G. Turner and W. H. Romme, "Landscape dynamics in crown fire ecosystems," *Landsc. Ecol.* **9**, 59–77 (1994).

Sambandh Dhal received his PhD in computer engineering from the Department of Electrical and Computer Engineering, Texas A&M University, College Station. His research focus is on developing data-driven approaches and error estimation techniques in dealing with sparse agricultural datasets. He is the recipient of IEEE-HKN membership in 2022 and the Texas Instruments Graduate Mentoring Fellowship in 2024.

Shubham Jain is a research assistant at Texas A&M AgriLife, Blackland Research Center, a PhD student with a demonstrated history of working in the water resources industry, and a licensed engineer in training. He is a strong research professional with his MS degree focused on water management and hydrological sciences from Texas A&M University and also received his Bachelor of Science degree in civil engineering.

Krishna Chaitanya Gadepally is a former MS student in the Department of Electrical and Computer Engineering, Texas A&M University, College Station. He received his BTech degree in electronics and instrumentation engineering from BITS, Pilani, in Hyderabad, in 2020. He works extensively in deep learning techniques and novel machine learning techniques in the agricultural sector.

Prathik Vijaykumar is pursuing his MS degree in computer engineering at Texas A&M University. His research focuses on resource allocation and orchestration for a Docker-based microservice system, using RL to reduce system bottlenecks and enhance efficiency.

Ulisses Braga-Neto is currently a professor of electrical and computer engineering at Texas A&M University, specializing in statistical pattern recognition, machine learning, signal and

image processing, and systems biology. Renowned for his work on error estimation in pattern recognition and machine learning, he co-authored the first dedicated book on the subject. His contributions extend to mathematical morphology in signal and image processing.

Bhavesh Hariom Sharma is a former MS student in the Department of Electrical and Computer Engineering, Texas A&M University, College Station. His major research interests include VLSI design, computer architecture, and hardware logic design. He is currently an architecture modeling engineer at NXP Semiconductors, Austin.

Bharat Sharma Acharya is the research director at Rodale Institute Southeast Organic Center, Chattahoochee Hills, Georgia. He is a soil scientist with multiple years of experience in vadose zone processes, hydrology, and water quality. Currently, he serves as an associate editor for the Soil Science Society of America Journal. He is also a recipient of the 2022 Education and Public Service Award from the Universities Council on Water Resources.

Kevin Nowka received his PhD from Stanford University in electrical engineering in 1996. After that, he was in IBM Research, Austin, for 22 years, before joining Texas A&M University in 2018 as a professor of practice. He works mostly in the applications of data-driven approaches in the agricultural domain.

Stavros Kalafatis received his BS degree from the University of Surrey in 1989 and his MS degree from the University of Arizona, Tucson, in 1991. He is the associate department head and a professor of practice in the ECE Department at Texas A&M University. He has been the Capstone program director since 2016. Before that, he worked for Intel Corp as a senior director of R&D in the CPU division.