Earth's Future

RESEARCH ARTICLE

10.1029/2024EF005187

Special Collection:
Advancing Interpretable AI/ML Methods for Deeper Insights and Mechanistic Understanding in Earth Sciences: Beyond Predictive Capabilities

Key Points:

- The spatial and temporal distribution of wildfires is captured by associated biological, physical, social, and management attributes
- Wildfires are increasingly reported with an unknown cause (5-fold increase from 1992 to 2020), hindering targeted prevention strategies
- ML models can identify the cause of wildfire ignitions and determine the contributions of diverse attributes to classification of cause

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

M. Sadegh,
mojtabasadegh@gmail.com

Citation:

Pourmohamad, Y., Abatzoglou, J. T., Fleishman, E., Short, K. C., Shuman, J., AghaKouchak, A., et al. (2025). Inference of wildfire causes from their physical, biological, social and management attributes. *Earth's Future*, *13*, e2024EF005187. https://doi.org/10.1029/2024EF005187

Received 30 JUL 2024
Accepted 8 DEC 2024

Author Contributions:

Conceptualization: Yavar Pourmohamad, John T. Abatzoglou, Karen C. Short, Mojtaba Sadegh
Data curation: Yavar Pourmohamad
Formal analysis: Yavar Pourmohamad
Investigation: Yavar Pourmohamad, Mojtaba Sadegh

Inference of Wildfire Causes From Their Physical, Biological, Social and Management Attributes

Yavar Pourmohamad[1], John T. Abatzoglou[2], Erica Fleishman[3], Karen C. Short[4], Jacquelyn Shuman[5], Amir AghaKouchak[6], Matthew Williamson[7], Seyd Teymoor Seydi[1], and Mojtaba Sadegh[1,8]

[1]Department of Civil Engineering, Boise State University, Boise, ID, USA, [2]Management of Complex Systems Department, University of California, Merced, CA, USA, [3]College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR, USA, [4]USDA Forest Service, Rocky Mountain Research Station, Missoula, MT, USA, [5]Earth Science Division, NASA Ames Research Center, Moffett Field, CA, USA, [6]Department of Civil and Environmental Engineering, University of California, Irvine, CA, USA, [7]Human-Environment Systems, Boise State University, Boise, ID, USA, [8]United Nations University Institute for Water, Environment and Health, United Nations University, Hamilton, ON, Canada

Abstract Effective wildfire prevention includes actions to deliberately target different wildfire causes. However, the cause of an increasing number of wildfires is unknown, hindering targeted prevention efforts. We developed a machine learning model of wildfire ignition cause across the western United States on the basis of physical, biological, social, and management attributes associated with wildfires. Trained on wildfires from 1992 to 2020 with 12 known causes, the overall accuracy of our model exceeded 70% when applied to out-of-sample test data. Our model more accurately separated wildfires ignited by natural versus human causes (93% accuracy), and discriminated among the 11 classes of human-ignited wildfires with 55% accuracy. Our model attributed the greatest percentage of 150,247 wildfires from 1992 to 2020 for which the ignition source was unknown to equipment and vehicle use (21%), lightning (20%), and arson and incendiarism (18%).

Plain Language Summary Prevention of human-caused wildfires, which account for >60% of ignitions across western United States (WUS), is an effective way of reducing wildfire risk. However, the reported cause of an increasing number of wildfires (>50% in recent years) is unknown, challenging the development of targeted prevention strategies. Here, we leverage the spatial and temporal characteristics of wildfires to infer the cause of 150,247 wildfires from 1992 to 2020 in WUS with an unknown cause. Each ignition cause was associated with distinct attributes. For example, lightning-ignited fires generally start at higher elevations and lower vegetation greenness, whereas wildfires started by power infrastructure mainly ignite during dry, hot, windy weather. Model accuracy in assigning an ignition cause to each wildfire among the 12 available classes was >70% when applied to test data. Across WUS, global human modification index, elevation, discovery day of year, fire year, and temperature on the day of ignition made the greatest contributions to determining the causes of wildfires. The contribution of weather was lower because wildfires start during a subset of weather conditions that are generally common between wildfires. The overall accuracies of models developed for individual states ranged from 60% (California) to 81% (Nevada), and the most influential attributes varied among states.

1. Introduction

The number of large wildfires—hereafter fires—increased substantially in the western United States (WUS) over the past five decades (Abatzoglou et al., 2016; Dennison et al., 2014), with severe short-to long-term human and environmental impacts (Khorshidi et al., 2020; Westerling et al., 2006). Climate change has substantially increased the probability of coincident dry, hot, and windy conditions (Khorshidi et al., 2020). These conditions, along with higher fuel density and fuel load, increased the number of fires that resulted in loss of human life or property (Parks et al., 2015). Furthermore, synchronous extreme fire danger across the WUS (Cattau et al., 2020) and extended fire seasons have strained fire suppression resources, and contributed to escalating fire activity due to lack of attack or late initial attack (Balch et al., 2017). Eighty-four percent of fire ignitions in the United States (>60% in the WUS) are human caused (Balch et al., 2017), and the intersection of human-caused ignitions and environmental change has resulted in several fires with catastrophic effects on people in recent years (e.g., Nauslar et al., 2018).

Methodology: Yavar Pourmohamad, John T. Abatzoglou, Mojtaba Sadegh
Project administration: Yavar Pourmohamad
Resources: Yavar Pourmohamad, John T. Abatzoglou, Mojtaba Sadegh
Software: Yavar Pourmohamad
Validation: Yavar Pourmohamad, Mojtaba Sadegh
Visualization: Yavar Pourmohamad
Writing – original draft: Yavar Pourmohamad, Erica Fleishman, Mojtaba Sadegh
Writing – review & editing: Yavar Pourmohamad, John T. Abatzoglou, Erica Fleishman, Karen C. Short, Jacquelyn Shuman, Amir AghaKouchak, Matthew Williamson, Seyd Teymoor Seydi, Mojtaba Sadegh

Human-caused ignitions exhibit temporal and spatial patterns that can be analyzed using various attributes, facilitating their modeling and the development of targeted fire prevention strategies (Thomas et al., 2013). The primary factors that are often considered in modeling human-caused ignitions are social, economic, and environmental (e.g., vegetation, meteorology, topography, distance to roads) (Arndt et al., 2013; Bergado et al., 2021; Earl et al., 2015; Fusco et al., 2016; Pelletier et al., 2005; Prestemon et al., 2010; Rodrigues & De la Riva, 2014; Syphard et al., 2007, 2008; Vilar del Hoyo et al., 2011). Generally, vegetation (fuels) that have become dry due to high temperatures and arid conditions increase the likelihood of a fire igniting and spreading (Rodrigues & De la Riva, 2014). Higher ignition likelihoods are seen in ecosystems with intermediate fuel densities (e.g., shrublands and open-canopy forests) (Earl et al., 2015) as well as in intermediate population and housing densities (CPR, 2021). Human access to wildland vegetation, for example, through road networks, is also positively correlated with the likelihood of ignitions (Arndt et al., 2013; Thomas et al., 2013).

Fire prevention strategies are devised and implemented on the basis of specific fire ignition sources, also referred to as fire causes (Syphard & Keeley, 2015). However, not all fires are formally investigated, and an increasing number of fires are reported with unknown or missing causes due to factors such as the destruction or inaccessibility of the fire's point of origin, lack of evidence (CPR, 2021; Sari, 2023), or changes in reporting practices. Although such fires are mostly recognized as ignited by human activity (Short, 2014), lack of knowledge of their cause challenges developing risk management strategies (Sari, 2023). The National Wildfire Coordination Group identifies 13 distinct fire causes: arson or incendiarism; debris and open burning; equipment and vehicle use; firearms and explosives use; fireworks; missing or unknown; misuse of fire by a minor; natural; other; power generation, transmission, or distribution; railroad operations and maintenance; recreation and ceremony; and smoking (Short & Finney, 2022).

Advanced machine learning models enable modeling of causes of ignitions on the basis of patterns captured in historical fires (Jain et al., 2020). In this study, we leveraged a recently developed data set of >2.3 million fires across the United States from 1992 to 2020 that includes nearly 270 attributes associated with each fire, and multiple ensemble machine learning (ML) and deep learning (DL) models, to answer two questions. First, what were the causes of fires recorded with missing or unknown ignition sources? Second, what is the relative contribution of different attributes in predicting the cause of fires across the WUS, and how does that contribution vary regionally?

2. Data and Methods

2.1. Fires and Their Attributes

We built and tested our models on the Fire Program Analysis Fire Occurrence Database-Attributes (FPA FOD-Attributes) data set (Pourmohamad et al., 2024), which augments the sixth edition of the FPA FOD (Short, 2022) with nearly 270 attributes that coincide with the date and location of each fire ignition in the United States. FPA FOD contains information on the location, jurisdiction, discovery time and date, cause, and final size of >2.3 million fires from 1992 to 2020. FPA FOD-Attributes offers comprehensive information on physical (e.g., weather, climate, topography, infrastructure), biological (e.g., land cover, normalized difference vegetation index), social (e.g., population density, social vulnerability index), and management (e.g., national and regional preparedness level, jurisdiction) attributes corresponding to each fire.

First, we filtered all fires that occurred in the WUS (Arizona, California, Colorado, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming) on the basis of whether their cause was known (534,113) or "missing or unknown" (218,348). We used ordinary least square linear regression analysis to test whether the number of fires ignited by each known cause increased or decreased from 1992 to 2020. We considered the trend as being significant if the p-value from the Mann-Kendall test was less than 0.05. We also summarized the area burned by fires ignited by each cause.

Attributes in FPA FOD-Attributes are either predictive (124 attributes: useful for machine learning and deep learning models) or descriptive (183 attributes: not applicable to these modeling methods) (Pourmohamad et al., 2024). We employed correlation analysis and expert judgment to select the most relevant predictive attributes with the least overlapping information as input to our models. First, we calculated the Pearson correlation coefficient between all pairs of 124 predictive attributes (Figure S1a in Supporting Information S1). Then, we selected 39 attributes (Figure S1b in Supporting Information S1) with the least collinearity and greatest relevance

10.1029/2024EF005187

to fire cause. We classified these predictive attributes as weather and fire danger indices (12 attributes), topography and land cover (six attributes), social (eight attributes), and management (13 attributes) (Tables S1 and S2 in Supporting Information S1).

2.2. Hierarchical Clustering and Similarity Analysis

We used multiple techniques to identify similarities and differences among the attributes associated with each ignition cause. We used a kernel density function to investigate the distributions of the attributes, and dendrograms to hierarchically cluster fire ignition sources (Alswaitti et al., 2018). A dendrogram is a tree-like diagram that, in our case, estimates distances between the attributes of each pair of ignition groups. We plotted dendrograms for each of the four classes of attributes (weather, topography, social, and management). To assess similarity of classes associated with each ignition source, we used the average attribute value, and a cophenetic distance function (Saraçli et al., 2013), which is the distance between attributes of two ignition sources along each linkage. The leaves of the dendrogram represent the ignition sources, and the branches represent the hierarchical relationships between them. The height of each branch indicates the dissimilarity between the two clusters that it connects.

2.3. Machine Learning and Deep Learning Modeling

Machine learning and deep learning (ML and DL) models are powerful tools to address major challenges in fire science (Jain et al., 2020) including characterizing and predicting fuel load, density, connectivity, and moisture (Alizadeh et al., 2024; Li et al., 2023; Seydi et al., 2024; Wu et al., 2023), detecting and mapping fires from remotely sensed images (Buch et al., 2023; Huot et al., 2020; Kondylatos et al., 2022; Seydi & Sadegh, 2023), understanding fire risk and climate influences (Gualdi et al., 2022; Shao et al., 2022), modeling fire behavior and spread (Allaire et al., 2021; Khennou et al., 2021), evaluating fire effects on ecosystems (Wang et al., 2021; Zou et al., 2019), and optimizing fire management decisions (Bot & Borges, 2022; Nawaz et al., 2021; Rad et al., 2023).

We used multiple ML (ensemble models) and DL (one-dimensional convolutional neural networks; CNN-1D) models (Chollet, 2017) to (a) model fire ignition sources on the basis of their attributes, (b) enhance the understanding of the contribution of various attributes to classifying ignition sources, and (c) attribute an ignition source to fires for which the cause in FPD FOD was recorded as missing or unknown. We first identified WUS fires for which the set of attributes was complete, resulting in an ML-ready sample of 369,262 and 150,427 fire incidents with a known cause and missing or unknown cause, respectively. We then divided the set of ML-ready fire records with a known cause into training (65%), validation (15%), and test (20%) data. We used training data to tune model parameters, validation data to tune model hyperparameters, and out-of-sample test data to assess model accuracy. We also leveraged the trained ML and DL models to ascribe an ignition source to each fire with a missing or unknown cause on the basis of its attributes.

We developed two CNN-1D architecture models with PyTorch and Keras, and developed five ensemble ML models (AdaBoost, Gradient Boosting, Random Forest, LightGBM, and XGBoost). We transformed fire attributes to the appropriate structure for inclusion in a given model. For instance, to feed the DL models with PyTorch, we transformed the Dataframe to a Tensor, and for the XGBoost model, we transformed data to a DMatrix. We tuned model hyperparameters through GridSearch cross validation. We sequentially narrowed the search window for each hyperparameter until the overall accuracy did not change more than 0.01%. We applied different architectures, sequentially changing the number of convolution layers, paddings, number of channels, dropout ratios, number of neurons, optimizers, and activation and loss functions, to identify the optimum PyTorch and Keras CNN-1D architectures. Because the number of fires associated with different ignition sources was widely different (unbalanced target), we trained all ML and DL models with and without class weights. In all models, we were interested in fire ignition cause without any consideration of the final fire size. Fire records include incidents that coincided with environmental conditions conducive to large fires, but timely and effective initial attack (i.e., suppression) prevented them from growing to large sizes. Excluding a subset of ignitions based on fire size unnecessarily remove important information from our models. We then selected the best model on the basis of overall accuracy, Kappa coefficient, Mathew's correlation coefficient, and balanced accuracy. We used the SHapley Additive exPlanations (SHAP) technique to explain the contribution of each attribute to the final output of each model (Bagheri, 2022).

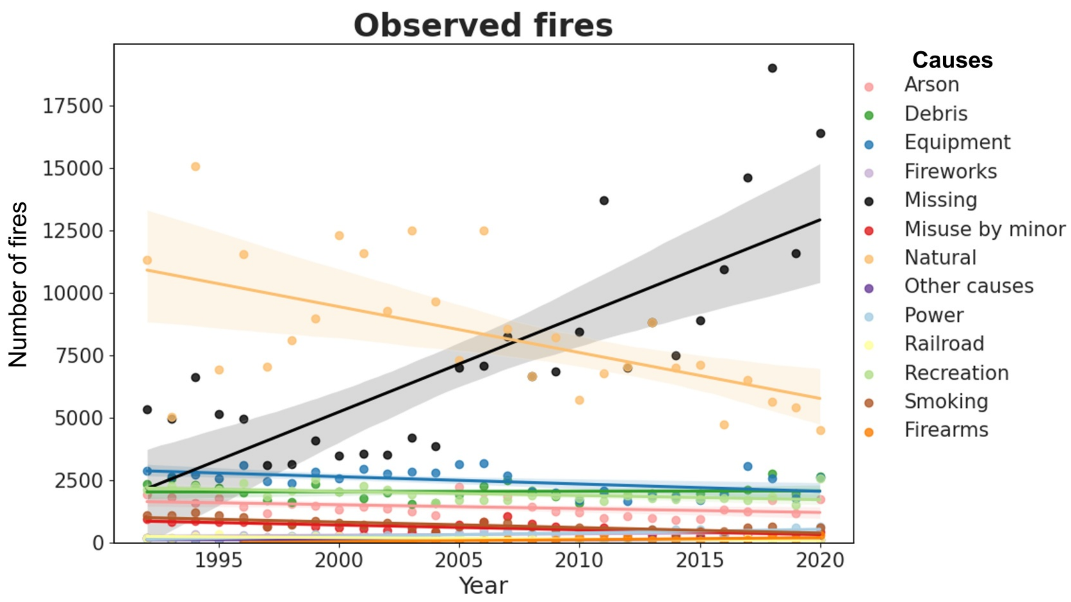

Figure 1. Trends in the annual number of fires by cause in the western United States.

3. Results

3.1. Trends and Patterns in Fire Ignitions

The number of fires ignited by lightning (i.e., natural cause—acknowledging that a marginal percentage of natural fires are not started by lightning) declined by an average of −183* per year across the WUS from 1992 to 2020 (Mann-Kendall p-value = 0.00; hereafter, all trends statistically significant at a 5% level are marked with *). Among the human-ignited fires with a specified cause, the trend in number of fires per year caused by arson and incendiarism (−15* fires/year), equipment and vehicle use (−29* fires/year), misuse of fire by a minor (−19* fires/year), railroad operations and maintenance (−6* fires/year), recreation and ceremony (−16* fires/year), and smoking (−22* fires/year) was significantly negative (Figure 1). The trend in number of fires per year caused by firearms (8* fires/year), fireworks (5* fires/years), and power generation, transmission, or distribution (14* fires/year) was significantly positive. Other studies have noted increases in the number of fires caused by power infrastructure in California (Keeley & Syphard, 2018) with notable societal impacts. For example, the 2018 Camp Fire, the deadliest fire on record in California, was ignited by a powerline failure. The number of fires with a missing or unknown cause increased by a statistically significant rate of 384* fires/year. In fact, fires with a missing or unknown cause accounted for ~50% of all fires during 2018–2020, challenging the development and implementation of targeted fire prevention strategies.

Causes of fire ignitions had distinct spatial patterns across the WUS (Figure 2a). Lightning accounted for a majority of ignitions in all states except California, Arizona, and Colorado, in which fires of missing or unknown cause were most prevalent. Natural fires accounted for the largest cumulative area burned in all states (62%), including 82% in Nevada (Figure 2a). California accounted for the greatest percentage of all ignitions (33%), followed by Arizona (14%). The percentage of burned area also was greatest in California (20%), followed by Idaho (16%) (Figure 2b). Across the WUS, the greatest percentages of fires were ignited by lightning (32%), missing or unknown causes (29%), and equipment or vehicle use (9%) (Figure 2d), and the resulting fires also accounted for the greatest percentages of burned areas (62%, 18%, and 7%, respectively) (Figure 2c).

3.2. Physical, Biological, Social, and Management Attributes of Fires

Fire attributes associated with different sources of ignitions had distinct distributions (Figure 3). Fires ignited by lightning, recreation and ceremony, debris and open burning, and power generation, transmission, or distribution were associated with lower fire danger as captured by the energy release component (erc), whereas ignitions from firearms and arson and incendiarism were associated with higher erc (Figure 3a). The distribution of erc for different fire causes was not always the same as the distribution relative to other fire danger indices with higher

10.1029/2024EF005187

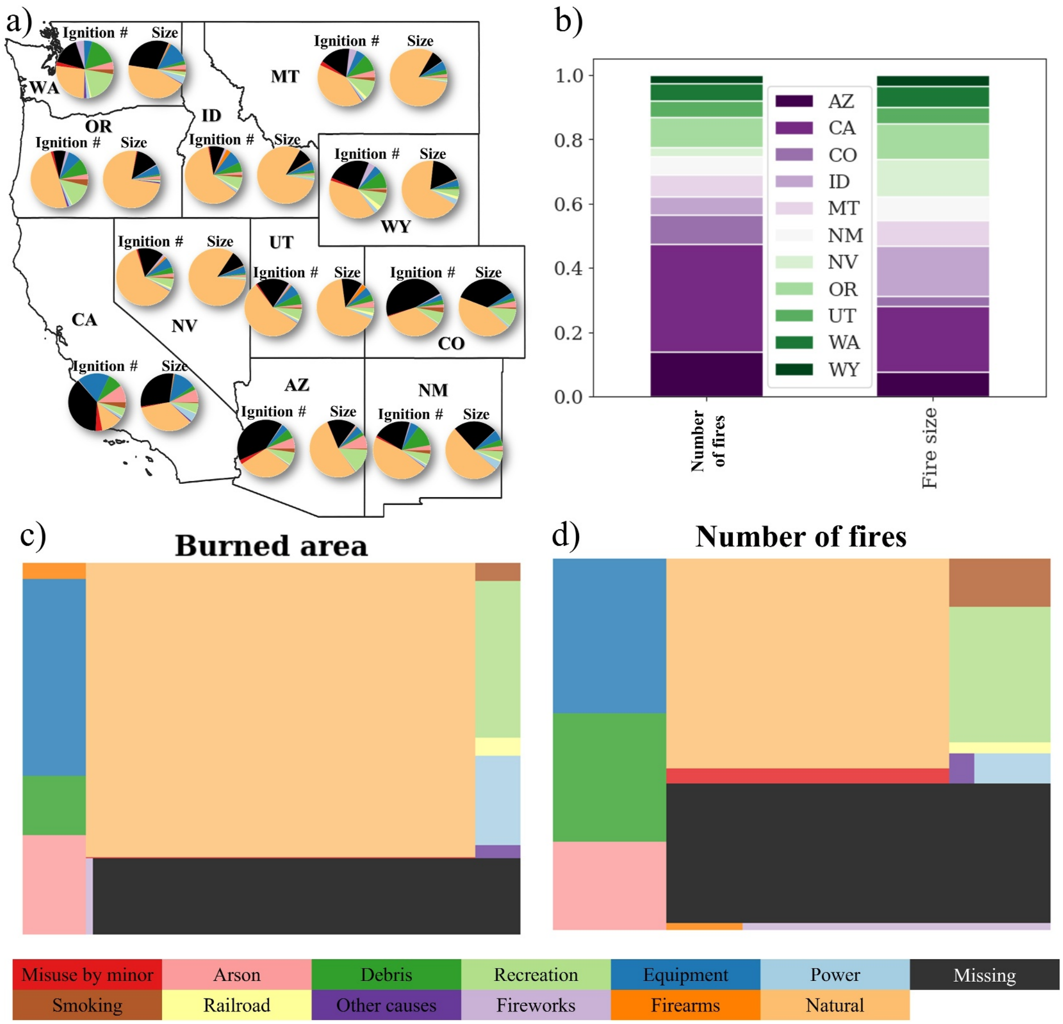

Figure 2. Distributions of the number of fires and cumulative area burned by fires ignited by different causes across the western United States from 1992 to 2020. (a) Percentage of fires caused by each ignition source and corresponding burned area in each state (AZ, Arizona; CA, California; CO, Colorado; ID, Idaho; MT, Montana; NM, New Mexico; NV, Nevada; OR, Oregon, UT, Utah; WA, Washington; WY, Wyoming). (b) Proportion of fires and burned areas in each state. (c) Percentage of burned area attributable to each ignition source. (d) Percentage of fires caused by each ignition source.

sensitivity to temperature and shorter memory, such as vapor pressure deficit (vpd) and 100-hr dead fuel moisture (fm100). Fires caused by recreation and ceremony and by debris and open burning ignited during periods with relatively low vpd and fm100, but fires caused by lightning and power generation, transmission, and distribution tended to occur when vpd and fm100 were high (Figure 3a). Fires caused by arson and incendiarism and firearms consistently ignited during dry, hot weather as reflected by indices with both short and long memory. Geography and seasonality of fires explain these patterns.

Elevations of fire ignitions varied by cause. Fires ignited by lightning occurred at distinctly higher elevations than those ignited by other causes (Figure 3b). Fires caused by fireworks, railroad operations and maintenance, firearms, and recreation and ceremony also ignited at relatively high elevations (Figure 3b). Fires ignited by natural causes and recreation and ceremony occurred on steeper slopes and in areas with a lower normalized difference vegetation index (lower vegetation greenness) than fires ignited by other causes (Figure 3b). Fires caused by lightning, smoking, and recreation and ceremony occurred where the global human modification index was lower (Figure 3c). Fires caused by firearms, fireworks, lightning, and railroad operations and maintenance started in areas with a lower gross domestic product (Figure 3c). Additionally, misuse of fire by a minor, arson and incendiarism, debris and open burning, smoking, and fireworks ignited in areas with a noticeably higher overall

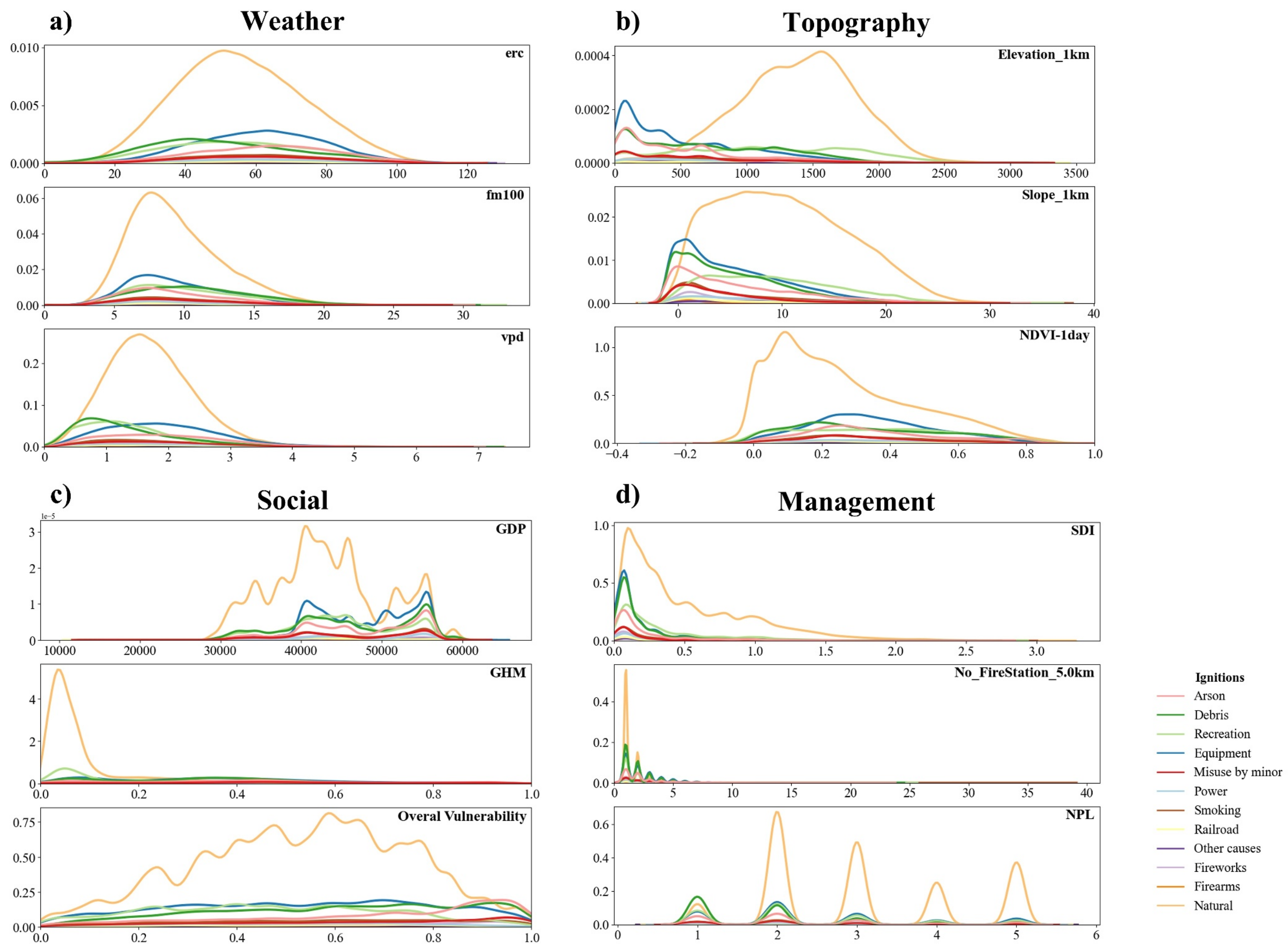

Figure 3. Attributes in the four general classes of (a) weather, (b) topography and biology, (c) social, and (d) management associated with specific fire ignition causes across the western United States from 1992 to 2020. erc: energy release component, fm100: 100-hr dead fuel moisture, vpd: vapor pressure deficit, NDVI: normalized difference vegetation index, GDP: gross domestic product, GHM: global human modification index, SDI: suppression difficulty index, No_FireStation_5.0 km: number of fire stations in a 5 km radius from the ignition point, NPL: national preparedness level.

social vulnerability index (Figure 3c). Management variables also showed distinct patterns; for example, debris and open burning was the most common cause of fires when the national preparedness level was relatively low (Figure 3d).

Figures S2–S5 in Supporting Information S1 illustrate distributions of attributes associated with different causes of fire ignitions in each state from 1992 to 2020.

3.3. Hierarchical Clustering of Fire Ignitions

Hierarchical clustering also revealed both similarities and differences in the attributes associated with fire ignition sources (Figure 4). For instance, the weather attributes associated with fires ignited by natural causes were distinct from those associated with other causes (Figure 4a). Topographic, social, and management attributes of fires ignited by natural causes were fairly similar to those associated with fires ignited by recreation and ceremony, railroad operations, and maintenance, and fireworks (Figure 4c). Fires ignited by arson and equipment and vehicle use were associated with similar weather (Figure 4a) and topographic (Figure 4b) attributes, but different social (Figure 4c) and management (Figure 4d) attributes. The clusters changed markedly at the state level (Figure S6 in Supporting Information S1).

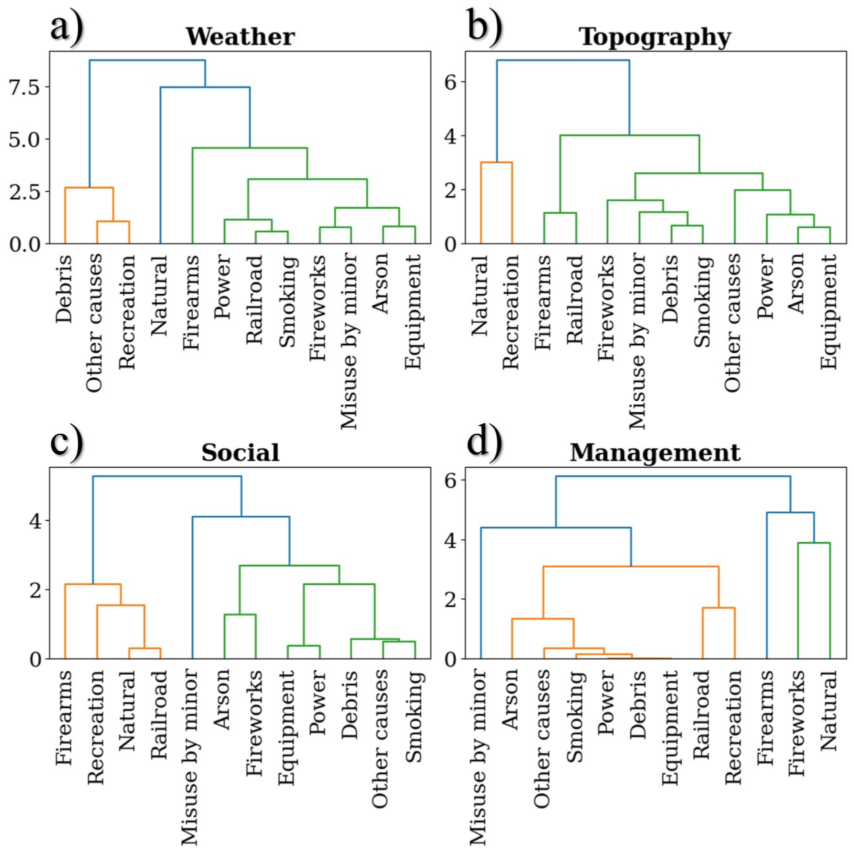

Figure 4. Hierarchical clustering of ignition sources on the basis of (a) weather, (b) topographic, (c) social, and (d) management attributes.

3.4. Machine Learning Models of Fire Ignitions

All ML and DL models effectively classified the cause of fires in the test data into one of the 12 categories with an overall accuracy >62% across the WUS (Figure 5a). All four performance metrics considered herein (overall accuracy (>70%), Kappa's coefficient, Mathew's correlation coefficient, and balanced accuracy) indicated that XGBoost outperformed other models (Figure 5a). Therefore, we adopted the XGBoost model for the remaining analyses. Literature has shown that the outcomes of XGBoost typically are accurate and generalizable (Tran et al., 2024).

The model's performance with respect to different fire causes was unequal, which in large part reflected distinctness of their associated attributes (Figure 5b). The rows in the confusion matrix are true labels—observed fire causes in the test data—and the columns are causes predicted by the XGBoost model. The diagonal elements represent correct classifications and off-diagonal elements represent misclassifications. The model most accurately classified fires with natural causes (94% accuracy), and least accurately classified fires of "other causes" (1% accuracy) (Figure 5b). The model generally classified fires caused by lightning, equipment and vehicle use, debris and open burning, recreation and ceremony, and firearms more accurately than fires with other sources of ignition (Figure 5b; Table S3 in Supporting Information S1).

The imbalance in the percentage of fires with different causes, from natural causes (32%) to "other causes" (<1%), may impact model performance (Thabtah et al., 2020). Although tree-based models, including XGBoost, are generally robust to class imbalances (Le et al., 2022), we tested whether including class weights may improve model accuracy. XGBoost with class weights had an overall accuracy of nearly 70%, and similar class-specific accuracies to those reported in Figure 5b; and including class weights did not improve performance. Therefore, we attribute the low accuracy with which the model classified certain fire causes to similarity among attributes associated with these causes, uncertainty in attributes, and the stochasticity of the fire ignition process. If our goal

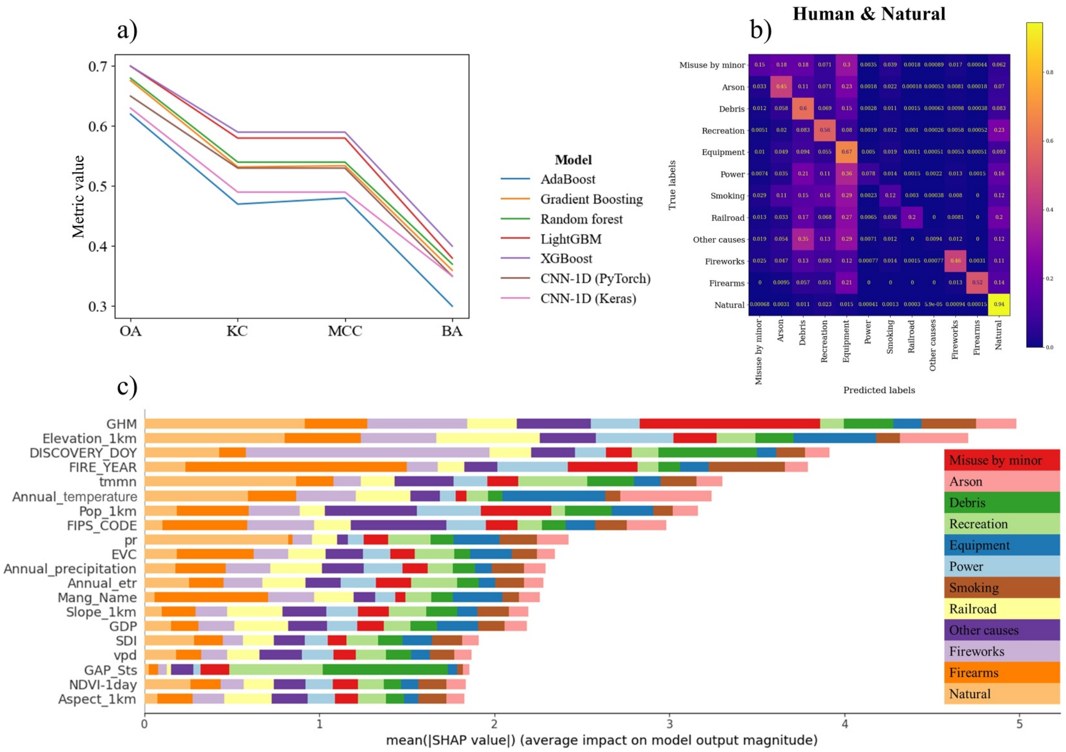

Figure 5. Machine learning classification of fire ignition sources. (a) Performance of seven models in classifying fire ignition source in out-of-sample test data across the western United States. (b) Confusion matrix based on the XGBoost model. (c) Attribute importance in the XGBoost classification model for the western United States. OA: overall accuracy, KC: Kappa coefficient, MCC: Mathew's correlation coefficient, BA: balanced accuracy, GHM: global human modification index, DOY: day of year, tmmn: minimum temperature, Pop_1 km: average population density within 1 km radius of the fire ignition point, pr: precipitation amount at the date of fire ignition, EVC: existing vegetation cover, etr: daily reference evapotranspiration, Mang_Name: management jurisdiction, GDP: gross domestic product, SDI: suppression difficulty index, vpd: vapor pressure deficit, GAP_Sts: protection status, NDVI: normalized difference vegetation index.

only was to classify fires as naturally versus human caused, the model accuracy would increase to 93% (Figure S7b in Supporting Information S1). Conversely, if our focus was to classify only human-caused fires, a more difficult classification task (Rodrigues & De la Riva, 2014), the overall model accuracy would drop to 55% (Figure S7a in Supporting Information S1).

SHAP-based analysis revealed that the global human modification index (GHM) was the greatest contributor to the XGBoost model's overall classification, followed by elevation, discovery day of year, fire year, and minimum daily temperature on the ignition date (Figures 5c and S8 in Supporting Information S1). GHM contributed the most to classifying fires caused by lightning and misuse of fire by a minor. The latter is in part due to the prevalence of natural ignitions in remote areas with low GHM values and the prevalence of ignitions from misuse of fire by minors in areas with high GHM values. Day of year contributed to classification of fires ignited by fireworks and debris and open burning due to the seasonality of these sources. For example, the number of fires ignited by fireworks peaks on 4 July, US Independence Day. We believe that fire year functions as a proxy for trends in the number of fires, which is driven by factors such as fire prevention efforts and population or demographic dynamics (Chen & Jin, 2022). Figure S9 in Supporting Information S1 shows attribute importance for each fire cause, acknowledging that the correlation between fire attributes (Figure S10 in Supporting Information S1) may impact the attribute importance analysis.

The overall accuracy of XGBoost models trained for each state ranged from 60% (California) to 81% (Nevada) (Table 1). Model performances are to a large extent governed by the percentage of naturally caused fires, which are the easiest to classify. California had the lowest percentage of naturally caused fires, and Nevada had the highest percentage (Figure 1 and Table 1). The global human modification index contributed the most to classification of fire cause in California, Colorado, and Wyoming. Elevation contributed the most to classification of

Table 1
The Overall Accuracy of the XGBoost Model Across the Western United States and in Each State

	West	AZ	CA	CO	ID	MT	NM	NV	OR	UT	WA	WY
Overall accuracy (%)	70.15%	73.72%	60.04%	75.89%	80.19%	70.49%	78.66%	80.99%	72.81%	78.92%	62.39%	70.54%
Number of fires with a known cause (×1,000)	369.3	42.6	107.9	25.3	26.9	29.8	23.6	12.2	46.1	20.8	23.7	10.4

Note. AZ, Arizona; CA, California; CO, Colorado; ID, Idaho; MT, Montana; NM, New Mexico; NV, Nevada; OR, Oregon; UT, Utah; WA, Washington; WY, Wyoming.

fire cause in Arizona, Nevada, Oregon, and Utah, annual precipitation to classification in Idaho, and discovery day of year to classification in Montana, New Mexico, and Washington (Figure S11 in Supporting Information S1).

3.5. Classifying Ignitions of Fires With a Missing or Unknown Cause

We used the trained XGBoost model to classify the ignition source of the 150,247 fires in the WUS for which the cause was missing or unknown cause and the set of attributes was complete. The model uses the set of attributes of each fire to assign a probability to each of the 12 causes. We first adopted the fire cause with the highest probability as the deterministic cause. The causes assigned to the highest percentages of fires were equipment and vehicle use (21%), natural (20%), arson and incendiarism (18%), and debris and open burning (15%) (Figure 6, middle column). Next, we adopted a more stringent selection criterion, in which we also considered the cause with the second highest probability if the model did not assign >90% probability to the first cause (Figure 6, right column). Although the threshold is arbitrary, we are confident in the classification if a >90% probability is assigned to a cause. The model assigned causes with >90% probability to roughly 22% of the 150,247 fires with a missing or unknown cause.

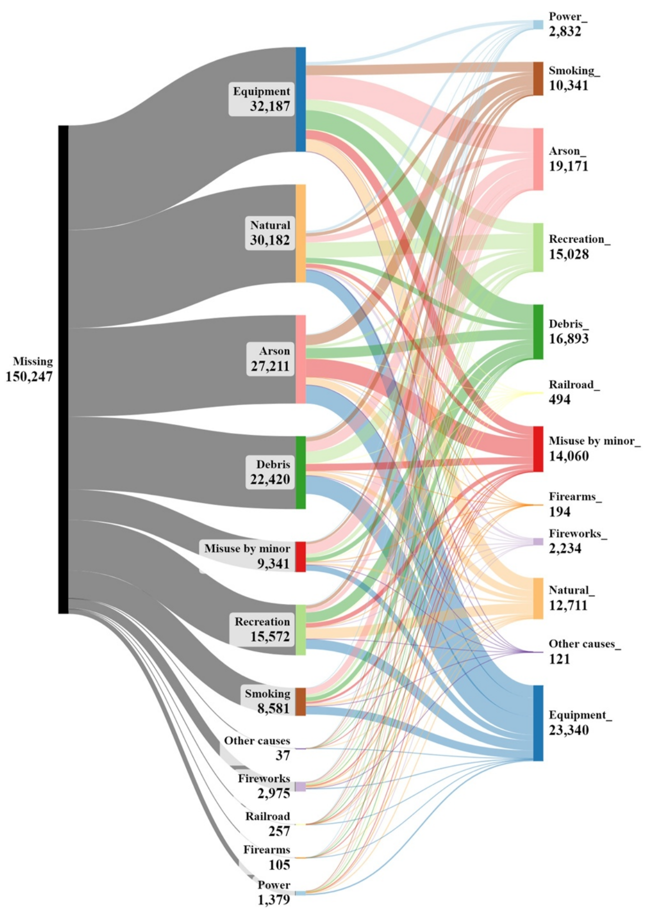

Figure 6. Ignition sources attributed to fires in the western United States from 1992 to 2020 with a missing or unknown cause. Middle column: causes attributed with highest probability. Right column: causes attributed with second-highest probability to fires for which the probability of the first cause was <90%.

Trends in the annual number of fires that were assigned to six classes were statistically significant: arson and incendiarism (86/year), equipment and vehicle use (68/year), debris and open burning (45/year), natural (28/year), recreation and ceremony (15/year), and power generation, distribution, and transmission (6/year) (Figure 7a). Next, we labeled the fires that were not classified with a >90% probability as "indecisive." The number of indecisive fires increased at a statistically significant rate of 218/year. But even in this case, the increase in the annual number of fires attributed to arson and incendiarism (25/year), debris and open burning (10/year), and recreation and ceremony (4/year) was statistically significant (Figure 7b).

Nearly 95% of all fires that were classified as arson and incendiarism-caused occurred in California (55%) or Arizona (40%) (Figure 7c), and the number of these fires increased over time (Figure 7d). Generally, a majority of fires with missing or unknown causes occurred in California and Arizona, and therefore they accounted for a disproportionately large percentage of fires in each class (Figure 7c). A high percentage of classifications attributed to debris and open burning occurred in Colorado and New Mexico, to firearms in Nevada and Utah, to fireworks in Colorado and Montana, to natural causes and smoking in Colorado, to other causes in Oregon and Nevada, to railroad operations and maintenance in Montana and Wyoming, and to recreation and ceremony in Colorado and Washington.

3.6. Classifying Ignitions of Fires With a Missing or Unknown Cause on the Basis of Reporting Agency

We trained separate models for fires reported by federal and non-federal agencies, as their reporting protocols and resources for investigating fire

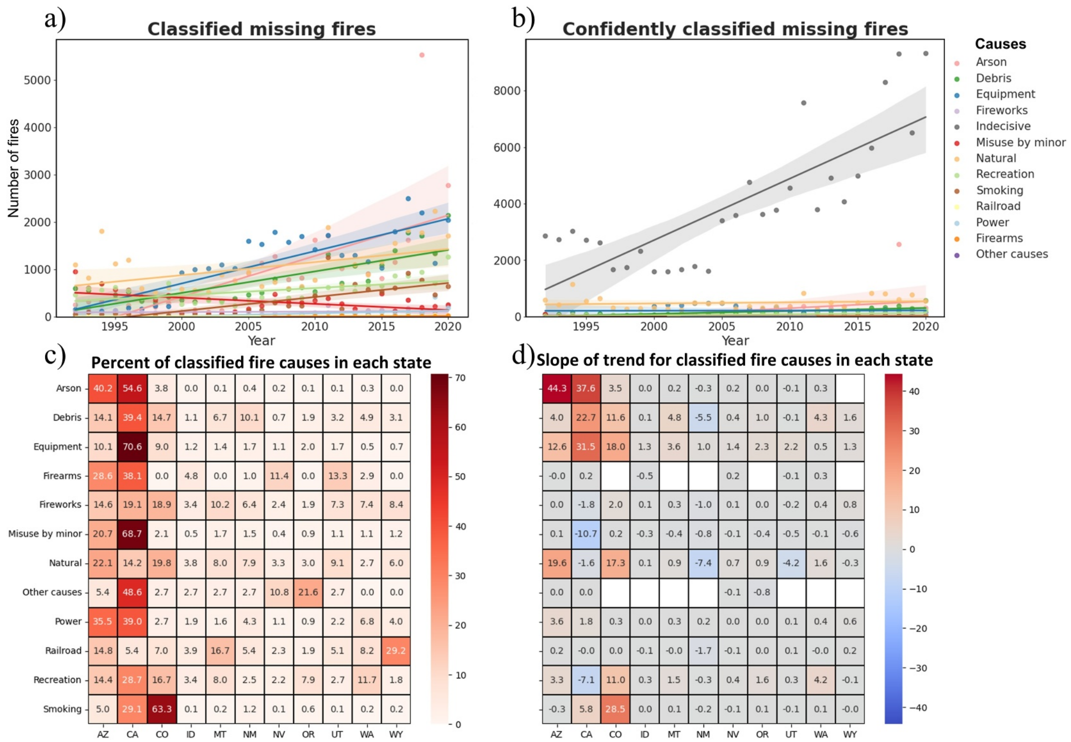

Figure 7. Trends in classified fires for which the cause originally was recorded as missing or unknown. (a) Trends in the annual number of deterministically classified fire ignitions by cause in the western US from 1992 to 2020. (b) Similar to (a) but fires for which the classification probability was <90% were indicated as indecisive. (c) Percentage of deterministically classified fire causes in each state. (d) Trends in deterministically classified fire causes in each state.

causes differ markedly across space (e.g., different counties across the West), time (e.g., recording procedures in 1992 vs. those in 2020), and reporting agency (e.g., state fire marshal vs. U.S. Forest Service). Federal and non-federal agencies reported roughly 37,300 and 113,200 fires with an unknown or missing cause, which accounted for 13% and 48% of their reports, respectively. Separately trained XGBoost models achieved 77% and 57% overall accuracy for test fires reported by federal and non-federal agencies, respectively (Table S4 in Supporting Information S1). Confusion matrices indicated widely different performances between the two models (Figures S12a and S12b in Supporting Information S1), and distinct attribute importances (Figures S12c and S12d in Supporting Information S1). The county/census tract in which the fire occurred (FIPS_Code), a proxy for reporting agency, was the most informative attribute in the non-federal model, whereas elevation was the most informative input to the federal model (Figures S12c and S12d in Supporting Information S1). In both models, a large percentage of fires with an unknown or missing cause were attributed to equipment and vehicle use, lightning, arson and incendiarism, and debris and open burning cause, with markedly higher percentage of fires attributed to recreation and ceremony in the federal model and to smoking in the non-federal model (Figure S13 in Supporting Information S1). In both models, the percentage of fires that were classified as caused by equipment and debris burning and lightning increased from 1992 to 2020, but the percentage of fires attributed to arson and incendiarism increased only in the non-federal model (Figure S14 in Supporting Information S1).

4. Discussion and Conclusion

Fire prevention is efficient and effective with strategies largely dependent on the source of ignition. For example, public safety power shutoffs are increasingly implemented by electric utilities during dry, hot, and windy conditions to prevent fires. As another example, installing fire pits in campgrounds reduces the chance of a recreational fire escaping and initiating a wildfire. Although engineering-, education-, and enforcement-based fire prevention strategies have been adopted across the United States (Hesseln, 2018), the increasing number of fires with a missing or unknown cause hinders targeted fire prevention. In fact, the rate of increase in the number of fires with a missing or unknown cause (384 fires/year—statistically significant) was greater than the rate of increase among other fire causes across the WUS from 1992 to 2020, and these fires accounted for ∼50% of all fires between 2018 and 2020. To address this, we leveraged the spatial and temporal patterns associated with known fire causes to classify the causes of fires for which the cause was missing or unknown. These patterns were captured through a comprehensive set of biological, physical, social, and management attributes. For example, natural fires generally occurred at higher elevations with lower global human modification index, lower gross domestic product (i.e., far from higher-income urban areas), and steeper slopes.

We developed a cohort of machine learning and deep learning models trained on historical fires with one of 12 known causes, and used this model to attribute an ignition source to the fires with a missing or unknown cause. The ensemble-based XGBoost model, with an accuracy of >70% on out-of-sample test data over the WUS, outperformed other models. The most influential attributes for classifying fire causes included the global human modification index, elevation, discovery day of year, fire year, and minimum daily temperature on the ignition date. The contribution of these attributes, however, varied among fire causes. For example, the global human modification index was more relevant for fires caused by lightning and misuse of fire by a minor, elevation for fires caused by lightning, and discovery day of year for fires caused by fireworks and debris and open burning. Certain attributes, although contributing little to overall accuracy, played a significant role in classifying specific fire causes. For example, precipitation on the day of ignition, while only the ninth-ranked contributor overall, was vital for classifying lightning-started fires. The latter is consistent with the literature that identified dry lightning (lightning that does not occur concurrent with precipitation) as the primary cause of natural fires (Herndon & Whiteside, 2018). Additionally, land management jurisdiction contributed to classification of fires ignited by firearms, which aligns with the large percentage of such fires that start on or near military installations (Lake & Christianson, 2020). Protected area status (GAP status) contributed to classification of fires caused by recreation and ceremony and debris and open burning through a negative feedback—that is, GAP status resulted in lower probabilities assigned to recreation or debris burning causes (U.S. Geological Survey Gap Analysis Project, 2018).

The contributions of weather and human population density were not as substantial as suggested by other studies (Abatzoglou & Kolden, 2011; Littell et al., 2009). This is in large part due to the target of the analysis. Much work has focused on determining presence or absence of fires, which is strongly affected by weather and human presence. However, we concentrated on classifying the causes of fires, and therefore environmental conditions were already conducive to an ignition. Also, in contrast to other studies (Narayanaraj & Wimberly, 2012), distance to road was not an important attribute in our model. The bias in reporting the location of fires that are marked in close proximity to roads is known (Mostafa et al., 2024), and a majority of fires in our sample were near roads.

Attributes that influence some causes of ignitions were not included in our analysis due to limited data availability. For example, information on the age and maintenance status of power distribution and transmission systems is not readily available and was not included in the training data, and model accuracy for fires ignited by power generation, distribution, and transmission was only 8%.

The greatest percentage of fires with missing or unknown causes were attributed to equipment and vehicle use (21%), natural causes (20%), arson and incendiarism (18%), and debris and open burning (15%). An increasing number of fires (86/year from 1992 to 2020) were attributed to arson and incendiarism by our model. We note that intentionally ignited fires in the built environment accounted for an average $590M (2021 US dollar) building loss from 2012 to 2021, with an increasing trend of $17 M/year ("Residential building fire dollar-loss causes (2013-2022)," 2023). While we have not conducted any forensic analysis, our model indicated that a large and increasing number of fires in the wildland areas with a missing or unknown cause were associated with biological, physical, social, and management conditions that were most similar to the fires known to be caused by arson and incendiarism. We also note the reporting uncertainties in the FPA FOD data set, especially for small fire reported by non-federal agencies.

We strived to develop explainable artificial intelligence models for classification of fire causes across the WUS. We designed the forcings (inputs) of the models on the basis of our understanding of the physical conditions associated with fire ignitions to ensure that our models only made predictions that reflected scientifically robust relations (Lattimer et al., 2020). Not only was the relative contribution of individual attributes to classification of fire causes consistent with our understanding of physical conditions, but our models resolved the complex interactions among various drivers. While our model is mainly designed for diagnostic purposes, our findings can inform future fire prevention efforts. Improved understanding of fire causes and their drivers informs fire prevention and management efforts that can save lives, property and other resources, and millions to billions of dollars in direct and indirect costs of fires. Finally, our models can also be readily adapted for prediction purposes.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

All the data used and models developed in this study are publicly available at: Pourmohamad, Y. (2024). Inference of Wildfire Causes from Their Physical, Biological, Social and Management Attributes [Data set and Software] (Version 1.0). Zenodo. https://doi.org/10.5281/zenodo.11510677.

Acknowledgments
This research was supported by the Joint Fire Science Program (Grants L21AC10247 and L24AC00277). The authors appreciate the contributions of Dr. Jeffrey Prestemon and Dr. Erin Belval from US Forest Service.

References

Abatzoglou, J. T., & Kolden, C. A. (2011). Relative importance of weather and climate on wildfire growth in interior Alaska. *International Journal of Wildland Fire*, *20*(4), 479–486. https://doi.org/10.1071/wf10046

Abatzoglou, J. T., Kolden, C. A., Balch, J. K., & Bradley, B. A. (2016). Controls on interannual variability in lightning-caused fire activity in the western US. *Environmental Research Letters*, *11*(4), 45005. https://doi.org/10.1088/1748-9326/11/4/045005

Alizadeh, M. R., Adamowski, J., & Entekhabi, D. (2024). Land and atmosphere precursors to fuel loading, wildfire ignition and post-fire recovery. *Geophysical Research Letters*, *51*(2), e2023GL105324. https://doi.org/10.1029/2023gl105324

Allaire, F., Mallet, V., & Filippi, J.-B. (2021). Emulation of wildland fire spread simulation using deep learning. *Neural Networks*, *141*, 184–198. https://doi.org/10.1016/j.neunet.2021.04.006

Alswaitti, M., Albughdadi, M., & Isa, N. A. M. (2018). Density-based particle swarm optimization algorithm for data clustering. *Expert Systems with Applications*, *91*, 170–186. https://doi.org/10.1016/j.eswa.2017.08.050

Arndt, N., Vacik, H., Koch, V., Arpaci, A., & Gossow, H. (2013). Modeling human-caused forest fire ignition for assessing forest fire danger in Austria. *iForest: Biogeosciences and Forestry*, *6*(6), 315–325. https://doi.org/10.3832/ifor0936-006

Bagheri, R. (2022). Introduction to shap values and their application in machine learning. *The Medium*.

Balch, J. K., Bradley, B. A., Abatzoglou, J. T., Nagy, R. C., Fusco, E. J., & Mahood, A. L. (2017). Human-started wildfires expand the fire niche across the United States. *Proceedings of the National Academy of Sciences*, *114*(11), 2946–2951. https://doi.org/10.1073/pnas.1617394114

Bergado, J. R., Persello, C., Reinke, K., & Stein, A. (2021). Predicting wildfire burns from big geodata using deep learning. *Safety Science*, *140*(January 2020), 105276. https://doi.org/10.1016/j.ssci.2021.105276

Bot, K., & Borges, J. G. (2022). A systematic review of applications of machine learning techniques for wildfire management decision support. *Inventions*, *7*(1), 15. https://doi.org/10.3390/inventions7010015

Buch, J., Williams, A. P., Juang, C. S., Hansen, W. D., & Gentine, P. (2023). SMLFire1. 0: A stochastic machine learning (SML) model for wildfire activity in the western United States. *Geoscientific Model Development*, *16*(12), 3407–3433. https://doi.org/10.5194/gmd-16-3407-2023

Cattau, M. E., Wessman, C., Mahood, A., & Balch, J. K. (2020). Anthropogenic and lightning-started fires are becoming larger and more frequent over a longer season length in the USA. *Global Ecology and Biogeography*, *29*(4), 668–681. https://doi.org/10.1111/geb.13058

Chen, B., & Jin, Y. (2022). Spatial patterns and drivers for wildfire ignitions in California. *Environmental Research Letters*, *17*(5), 055004. https://doi.org/10.1088/1748-9326/ac60da

Chollet, F. (2017). Deep learning with Python. https://doi.org/10.23919/ICIF.2018.8455530

CPR. (2021). Colorado's biggest wildfires are mostly human-caused—And unsolved | Colorado public radio. Retrieved from https://www.cpr.org/2021/11/11/colorados-biggest-wildfires-are-mostly-human-caused-and-unsolved/

Dennison, P. E., Brewer, S. C., Arnold, J. D., & Moritz, M. A. (2014). Large wildfire trends in the western United States, 1984–2011. *Geophysical Research Letters*, *41*(8), 2928–2933. https://doi.org/10.1002/2014gl059576

Earl, N., Simmonds, I., & Tapper, N. (2015). Weekly cycles of global fires—Associations with religion, wealth and culture, and insights into anthropogenic influences on global climate. *Geophysical Research Letters*, *42*(21), 9579–9589. https://doi.org/10.1002/2015gl066383

Fusco, E. J., Abatzoglou, J. T., Balch, J. K., Finn, J. T., & Bradley, B. A. (2016). Quantifying the human influence on fire ignition across the western USA. *Ecological Applications*, *26*(8), 2388–2399. https://doi.org/10.1002/eap.1395

Gualdi, B., Binet-Stéphan, E., Bahabi, A., Marchal, R., & Moncoulon, D. (2022). Modelling fire risk exposure for France using machine learning. *Applied Sciences*, *12*(3), 1635. https://doi.org/10.3390/app12031635

Herndon, J. M., & Whiteside, M. (2018). California wildfires: Role of undisclosed atmospheric manipulation and geoengineering. *Journal of Geography, Environment and Earth Science International*, *17*(3), 1–18. https://doi.org/10.9734/jgeesi/2018/44148

Hesseln, H. (2018). Wildland fire prevention: A review. *Current Forestry Reports*, *4*, 178–190. https://doi.org/10.1007/s40725-018-0083-6

Huot, F., Hu, R. L., Ihme, M., Wang, Q., Burge, J., Lu, T., et al. (2020). *Deep learning models for predicting wildfires from historical remote-sensing data* (pp. 1–11). NeurIPS. Retrieved from http://arxiv.org/abs/2010.07445

Jain, P., Coogan, S. C. P., Subramanian, S. G., Crowley, M., Taylor, S., & Flannigan, M. D. (2020). A review of machine learning applications in wildfire science and management. *Environmental Reviews*, *28*(4), 478–505. https://doi.org/10.1139/er-2020-0019

Keeley, J. E., & Syphard, A. D. (2018). Historical patterns of wildfire ignition sources in California ecosystems. *International Journal of Wildland Fire*, *27*(12), 781–799. https://doi.org/10.1071/wf18026

Khennou, F., Ghaoui, J., & Akhloufi, M. A. (2021). Forest fire spread prediction using deep learning. In *Geospatial informatics XI* (Vol. 11733, pp. 106–117). SPIE. https://doi.org/10.1117/12.2585997

Khorshidi, M. S., Dennison, P. E., Nikoo, M. R., AghaKouchak, A., Luce, C. H., & Sadegh, M. (2020). Increasing concurrence of wildfire drivers tripled megafire critical danger days in Southern California between 1982 and 2018. *Environmental Research Letters*, *15*(10), 104002. https://doi.org/10.1088/1748-9326/abae9e

Kondylatos, S., Prapas, I., Ronco, M., Papoutsis, I., Camps-Valls, G., Piles, M., et al. (2022). Wildfire danger prediction and understanding with deep learning. *Geophysical Research Letters*, *49*(17), 1–11. https://doi.org/10.1029/2022gl099368

Lake, F. K., & Christianson, A. C. (2020). Indigenous fire stewardship. In *Encyclopedia of wildfires and wildland-urban interface (WUI) fires* (pp. 714–722). Springer.

Lattimer, B. Y., Hodges, J. L., & Lattimer, A. M. (2020). Using machine learning in physics-based simulation of fire. *Fire Safety Journal*, *114*, 102991. https://doi.org/10.1016/j.firesaf.2020.102991

Le, T.-T.-H., Oktian, Y. E., & Kim, H. (2022). XGBoost for imbalanced multiclass classification-based industrial internet of things intrusion detection systems. *Sustainability*, *14*(14), 8707. https://doi.org/10.3390/su14148707

Li, F., Zhu, Q., Riley, W. J., Zhao, L., Xu, L., Yuan, K., et al. (2023). AttentionFire_v1. 0: Interpretable machine learning fire model for burned-area predictions over tropics. *Geoscientific Model Development*, *16*(3), 869–884. https://doi.org/10.5194/gmd-16-869-2023

Littell, J. S., Mckenzie, D., Peterson, D. L., & Westerling, A. L. (2009). Climate and wildfire area burned in western U.S. ecoprovinces, 1916-2003. *Ecological Applications*, *19*(4), 1003–1021. https://doi.org/10.1890/07-1183.1

Mostafa, M., Elia, M., Giannico, V., Lafortezza, R., & Sanesi, G. (2024). Assessing forest road network suitability in relation to the spatial occurrence of wildfires in mediterranean forest ecosystems. *Fire*, *7*(6), 175. https://doi.org/10.3390/fire7060175

Narayanaraj, G., & Wimberly, M. C. (2012). Influences of forest roads on the spatial patterns of human-and lightning-caused wildfire ignitions. *Applied Geography*, *32*(2), 878–888. https://doi.org/10.1016/j.apgeog.2011.09.004

Nauslar, N. J., Abatzoglou, J. T., & Marsh, P. T. (2018). The 2017 north bay and southern California fires: A case study. *Fire*, *1*(1), 18. https://doi.org/10.3390/fire1010018

Nawaz, M., Iqbal, M. F., & Mahmood, I. (2021). Validation of CHIRPS satellite-based precipitation dataset over Pakistan. *Atmospheric Research*, *248*, 105289. https://doi.org/10.1016/j.atmosres.2020.105289

Parks, S. A., Miller, C., Parisien, M.-A., Holsinger, L. M., Dobrowski, S. Z., & Abatzoglou, J. (2015). Wildland fire deficit and surplus in the western United States, 1984–2012. *Ecosphere*, *6*(12), 1–13. https://doi.org/10.1890/es15-00294.1

Pelletier, J., Mayer, L., Pearthree, P., House, P., Demsey, K., Klawon, J., & Vincent, K. (2005). An integrated approach to flood hazard assessment on alluvial fans using numerical modeling, field mapping, and remote sensing. *Geological Society of America Bulletin*, *117*(9–10), 1167–1180. https://doi.org/10.1130/B255440.1

Pourmohamad, Y. (2024). Inference of wildfire causes from their physical, biological, social and management attributes [Dataset and Software]. *Zenodo*. https://doi.org/10.5281/zenodo.11510677

Pourmohamad, Y., Abatzoglou, J. T., Belval, E. J., Fleishman, E., Short, K. C., Reeves, M. C., et al. (2024). Physical, social, and biological attributes for improved understanding and prediction of wildfires: FPA FOD-attributes dataset. *Earth System Science Data*, *16*(6), 3045–3060. https://doi.org/10.5194/essd-16-3045-2024

Prestemon, J. P., Butry, D. T., Abt, K. L., & Sutphen, R. (2010). Net benefits of wildfire prevention education efforts. *Forest Science*, *56*(2), 181–192. https://doi.org/10.1093/forestscience/56.2.181

Rad, A. M., Abatzoglou, J., Fleishman, E., Mockrin, M. H., Radeloff, V. C., Pourmohamad, Y., et al. (2023). Social vulnerability of the people exposed to wildfires. *Science Advances*, *9*(38), 10. https://doi.org/10.1126/sciadv.adh4615

Residential building fire dollar-loss causes (2013-2022). (2023). Retrieved from https://www.usfa.fema.gov/statistics/residential-fires/dollar-loss.html

Rodrigues, M., & De la Riva, J. (2014). An insight into machine-learning algorithms to model human-caused wildfire occurrence. *Environmental Modelling & Software*, *57*, 192–201. https://doi.org/10.1016/j.envsoft.2014.03.003

Saraçli, S., Doğan, N., & Doğan, İ. (2013). Comparison of hierarchical cluster analysis methods by cophenetic correlation. *Journal of Inequalities and Applications*, *2013*, 1–8. https://doi.org/10.1186/1029-242x-2013-203

Sari, F. (2023). Identifying anthropogenic and natural causes of wildfires by maximum entropy method-based ignition susceptibility distribution models. *Journal of Forestry Research*, *34*(2), 355–371. https://doi.org/10.1007/s11676-022-01502-4

Seydi, S. T., Abatzoglou, J. T., AghaKouchak, A., Pourmohamad, Y., Mishra, A., & Sadegh, M. (2024). Predictive understanding of links between vegetation and soil burn severities using physics-informed machine learning. *Earth's Future, Under Revi*, *12*(8). https://doi.org/10.1029/2024ef004873

Seydi, S. T., & Sadegh, M. (2023). Improved burned area mapping using monotemporal Landsat-9 imagery and convolutional shift-transformer. *Measurement*, *216*, 112961. https://doi.org/10.1016/j.measurement.2023.112961

Shao, Y., Feng, Z., Sun, L., Yang, X., Li, Y., Xu, B., & Chen, Y. (2022). Mapping China's forest fire risks with machine learning. *Forests*, *13*(6), 856. https://doi.org/10.3390/f13060856

Short, K. C. (2014). A spatial database of wildfires in the United States, 1992-2011. *Earth System Science Data*, *6*(1), 1–27. https://doi.org/10.5194/essd-6-1-2014

Short, K. C. (2022). Spatial wildfire occurrence data for the United States, 1992-2020 [FPA_FOD_20221014].

Short, K. C., & Finney, M. A. (2022). Agency records of wildfires caused by firearms use in the United States. *Fire Safety Journal*, *131*, 103622. https://doi.org/10.1016/j.firesaf.2022.103622

Syphard, A. D., & Keeley, J. E. (2015). Location, timing and extent of wildfire vary by cause of ignition. *International Journal of Wildland Fire*, *24*(1), 37–47. https://doi.org/10.1071/wf14024

Syphard, A. D., Radeloff, V. C., Keeley, J. E., Hawbaker, T. J., Clayton, M. K., Stewart, S. I., & Hammer, R. B. (2007). Human influence on California fire regimes. *Ecological Applications*, *17*(5), 1388–1402. https://doi.org/10.1890/06-1128.1

Syphard, A. D., Radeloff, V. C., Keuler, N. S., Taylor, R. S., Hawbaker, T. J., Stewart, S. I., & Clayton, M. K. (2008). Predicting spatial patterns of fire on a southern California landscape. *International Journal of Wildland Fire*, *17*(5), 602–613. https://doi.org/10.1071/wf07087

Thabtah, F., Hammoud, S., Kamalov, F., & Gonsalves, A. (2020). Data imbalance in classification: Experimental evaluation. *Information Sciences*, *513*, 429–441. https://doi.org/10.1016/j.ins.2019.11.004

Thomas, D., Butry, D., & Prestemon, J. (2013). The effects of wildfire prevention activities. In T. F. Shupe & M. S. Bowen (Eds.), *Proceedings of the natural resources symposium* (pp. 101–115). LSU AgCenter.

Tran, T. T. K., Janizadeh, S., Bateni, S. M., Jun, C., Kim, D., Trauernicht, C., et al. (2024). Improving the prediction of wildfire susceptibility on Hawai'i Island, Hawai'i, using explainable hybrid machine learning models. *Journal of Environmental Management*, *351*, 119724. https://doi.org/10.1016/j.jenvman.2023.119724

U.S. Geological Survey (USGS) Gap Analysis Project (GAP). (2018). *US geological survey–gap analysis Project species range maps CONUS_2001*. US Geological Survey Data Release. https://doi.org/10.5066/F7Q81B3R

Vilar del Hoyo, L., Martín Isabel, M. P., & Martínez Vega, F. J. (2011). Logistic regression models for human-caused wildfire risk estimation: Analysing the effect of the spatial accuracy in fire occurrence data. *European Journal of Forest Research*, *130*(6), 983–996. https://doi.org/10.1007/s10342-011-0488-2

Wang, S. S.-C., Qian, Y., Leung, L. R., & Zhang, Y. (2021). Interpretable machine learning prediction of fire emissions and comparison with FireMIP process-based models. *Atmospheric Chemistry and Physics Discussions*.

Westerling, A. L., Hidalgo, H. G., Cayan, D. R., & Swetnam, T. W. (2006). Warming and earlier spring increase western US forest wildfire activity. *Science*, *313*(5789), 940–943. https://doi.org/10.1126/science.1128834

Wu, X., Zhang, G., Yang, Z., Tan, S., Yang, Y., & Pang, Z. (2023). Machine learning for predicting forest fire occurrence in Changsha: An innovative investigation into the introduction of a forest fuel factor. *Remote Sensing*, *15*(17), 4208. https://doi.org/10.3390/rs15174208

Zou, Y., O'Neill, S. M., Larkin, N. K., Alvarado, E. C., Solomon, R., Mass, C., et al. (2019). Machine learning-based integration of high-resolution wildfire smoke simulations and observations for regional health impact assessment. *International Journal of Environmental Research and Public Health*, *16*(12), 2137. https://doi.org/10.3390/ijerph16122137