Journal of Environmental Management 341 (2023) 117908

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: www.elsevier.com/locate/jenvman

Research article

A multi-modal wildfire prediction and early-warning system based on a novel machine learning framework

Rohan T. Bhowmik [a,b,*], Youn Soo Jung [a], Juan A. Aguilera [a], Mary Prunicki [a], Kari Nadeau [a]

[a] Stanford University School of Medicine, Stanford, CA, 94305, USA
[b] The Harker School, San Jose, CA, 95129, USA

ARTICLE INFO

Handling Editor: Lixiao Zhang

Keywords:
Wildfire
Artificial intelligence
Machine learning
Remote sensor network
Disaster prediction

ABSTRACT

Wildfires are increasingly impacting the environment and human health. Among the top 20 California wildfires, those in 2020–2021 burned more acres than the last century combined. Lack of an adequate early warning system impacts the health and safety of vulnerable populations disproportionately and widens the inequality gap. In this project, a multi-modal wildfire prediction and early warning system has been developed based on a novel spatio-temporal machine learning architecture. A comprehensive wildfire database with over 37 million data points was created, including the historical wildfires, environmental and meteorological sensor data from the Environmental Protection Agency, and geological data. The data was augmented into 2.53 km × 2.53 km square grids to overcome the sensor network coverage limitations. Leading and trailing indicators for the wildfires are proposed, classified, and tested. The leading indicators are correlated to the risks of wildfire conception, whereas the trailing indicators are correlated to the byproducts of the wildfires. Additionally, geological data was incorporated to provide additional information for better assessment on wildfire risks and propagation. Next, a novel U-Convolutional Long Short-Term Memory (ULSTM) neural network was developed to extract key spatial and temporal features of the dataset, specifically to address the spatial nature of the location of the wildfire and time-progression temporal nature of the wildfire evolution. Through iterative improvements and optimization, the final ULSTM network architecture, trained with data from 2012 to 2017, achieved >97% accuracy for predicting wildfires in 2018, as compared to ~76% using traditional Convolutional Neural Network (CNN) techniques. The final model was applied to conduct a retrospective study for the 2018–2022 wildfire seasons, and successfully predicted 85.7% of wildfires >300 K acres in size. This technique could enable fire departments to anticipate and prevent wildfires before they strike and provide early warnings for at-risk individuals for better preparation, thereby saving lives, protecting the environment, and avoiding economic damages.

1. Introduction

Wildfires are increasingly threatening human lives, safety, and the environment at a global scale. Within the United States, the annual average of burned acreage by wildfires since 2000 is double that of the 1990s (Hoover and Hanson, 2022). Among California's top 20 wildfires, those in 2020–2021 burned more areas than the last century combined (Porter et al., 2020). Instances of particularly large wildfires, such as those with over 25,000 acres burned have been documented throughout California's history. A large influx of people settling into California sparked a wave of conflagrations that peaked in the 1920s, leading to over 4.2 million acres of land burned in that decade (Hoover and Hanson, 2022). However, its severity is dwarfed in comparison to the 2020 wildfire season, which burned over 4.2% of California's land area or 4.3 million acres of land (High, 2022). In addition, wildfires cause significant damage to the infrastructure and economy, California's 2018 wildfire season caused economic damages of $148.5 billion (roughly 1.5% of California's annual gross domestic product), with $27.7 billion (19%) in capital losses, $32.2 billion (22%) in health costs and $88.6 billion (59%) in indirect losses (Dennison et al., 2014; Ullrich et al., 2018).

Rapid climate change leads to extreme temperatures, decrease in precipitation, and reduced moisture in vegetation (Miller, 2020; Williams, 2016). The change in moisture levels within the soil, air, and vegetation negatively impacts the water content that would otherwise act as a buffer against fire spread (Balch et al., 2017). Under these

* Corresponding author. Stanford University School of Medicine, Stanford, CA, 94305, USA.
E-mail address: rbhowmik@stanford.edu (R.T. Bhowmik).

https://doi.org/10.1016/j.jenvman.2023.117908
Received 11 December 2022; Received in revised form 16 February 2023; Accepted 9 April 2023
Available online 12 May 2023

conditions, wildfires could easily burn out of control, leading to major conflagrations before firefighters can properly manage, mitigate, and control the burns.

Land and fire management has also exacerbated the hazard (Halofsky et al., 2020). The common fire suppression strategy from wildfire management services such as the U.S. Forest Service is to immediately control and put out wildfires upon reporting. Forest overgrowth and lack of controlled burns lead to a gradual accumulation of unburnt foliage, building up a large amount of combustible fuel for future fires. Decades worth of additional forestry coupled with population and economic growth at the wildland-urban interface has dramatically increased the human exposure to fires. The combined results increase the wildfire risks (Taylor and WoolfordCB Dean and DL Martell).

In order to combat wildfires, these issues must be addressed from their root cause. New technologies for wildfire prediction and prevention also need to be developed and deployed. Specifically, analyzing and correlating the effects of the aforementioned environmental circumstances on wildfire occurrences could help forecast impending wildfires, allowing for faster wildfire response, management, and damage mitigation (Kitzberger et al., 2017). As such, the ability to predict wildfire location and severity with high precision would be a significant breakthrough to prevent small flames from transgressing into larger conflagrations so that wildfire spread can be controlled intelligently (Wang et al., 2021; Marlon et al., 2012).

In this project, we proposed, developed, and evaluated a multi-modal wildfire prediction and early warning system based on a novel machine learning framework, aiming to aid the suppression of major fires before they could spread. Historical environmental, meteorological, and geological data along with past wildfire data were compiled to create a comprehensive California wildfire database with over 37 million data points, the largest of its kind. We architected a novel U-Convolutional Long Short-Term Memory (ULSTM) neural network model which incorporates aspects of the U-Net Convolutional Neural Networks (CNNs) architecture and Long Short-Term Memory (LSTM) Recurrent Neural Networks (RNNs) architecture to address the spatial and temporal nature of the wildfire indicators.

The ULSTM model was trained using the wildfire database and optimized through iterative permutations of training parameters. The model conducts risk assessments of the onset of major wildfire within each given location, the results of which could be instructive for wildfire management resources allocation optimization and for evacuation readiness. While smaller wildfires are indeed necessary for keeping forest growth in check and preventing wildfire crises in the future, it is critical to prevent these smaller wildfires from uncontrollably spreading (Ban et al., 2020; Molina-Pico et al., 2016). This ULSTM model can also be used to predict the likelihood of small fires spreading and thus enable proactive prevention.

2. Prior research

Early warning systems for natural disasters have been widely examined in both scientific research and governmental policies. Current early warning systems offer two methods of mitigating losses from natural disasters (Molina-Pico et al., 2016; Kelleher et al., 2018; Varela et al., 2020; Sayad and MousannifHassan Al Moatassime, 2019; Jain et al., 2020). The first method involves predicting the likelihood of a natural disaster occurring in the near future within a region by comparing its meteorological and geographic conditions to those recorded preceding historical disasters (Molina-Pico et al., 2016; Kelleher et al., 2018; Jain et al., 2020). The second method is based on detecting the onset of an imminent disaster by observing definite precursors and signs for these geological events (Varela et al., 2020; Sayad and MousannifHassan Al Moatassime, 2019).

Different types of natural disasters benefit differently from each early warning method. Disasters whose severity progresses over time, such as wildfires, tend to show noticeable indicators and follow predictable trends. These characteristics could be leveraged to alert, warn, and advise evacuation of the potentially impacted populations.

Previous work on wildfire early-warning systems primarily focuses on wildfire detection, rather than prediction. These solutions require infrastructure support, such as weather stations, sensor networks, and lookout points, the deployment cost of which could become prohibitive at scale. Furthermore, existing prediction techniques are limited in scope and accuracy (Kelleher et al., 2018; Varela et al., 2020; Sayad and MousannifHassan Al Moatassime, 2019; Jain et al., 2020).

The previous research conducted on wildfire analysis, detection, and prediction is summarized in Table 1 (Molina-Pico et al., 2016; Kelleher et al., 2018; Varela et al., 2020; Sayad and MousannifHassan Al Moatassime, 2019; Jain et al., 2020; Costa et al., 2017). The major achievements and key limitations are compared and discussed. Jain et al. provided an overview of wildfire management techniques, with a focus on machine learning to tackle wildfire detection, monitoring, and prediction (Costa et al., 2017). Jain proposes that machine learning techniques hold potential for wildfire applications, yet this is an emerging field needing future development. Ban et al. created a wildfire progression prediction tool using a Convolutional Neural Network (CNN) trained on Synthetic Aperture Radar (SAR) data (Molina-Pico et al., 2016). Ban's technique showed high accuracy in tracking wildfire progression regardless of geological conditions. However, this technique can only detect burnt areas after it has already been ravaged by the fire. The focus of their research is on fire progression prediction, not on forecasting impending wildfires. Molina-Pico et al. used wireless sensor networks for monitoring basic meteorological factors related to wildfires, setting the foundation for sensor networks in this field of study (Kelleher et al., 2018). However, this method requires a dense network of sensors, as its accuracy decreases with increasing distance from the sensor nodes. Kelleher et al. used an Outdoor Aerosol Sampler (OAS) network, which measures environmental levels in the air. The OAS devices are durable and spatially efficient, which is suitable for remote wildfire spread monitoring (Varela et al., 2020). However, OAS devices are battery dependent and inaccurate for extreme particulate matter reading. Varela et al. created a regression model for detecting wildfires, which achieved perfect accuracy in a controlled environment using only simple detection criteria (Sayad and MousannifHassan Al Moatassime, 2019). However, this technique doesn't perform as well in the real world where the sensors could be under sun exposure. Sayad et al. created a wildfire prediction technique using satellite data based on geological and meteorological factors (Jain et al., 2020). While this technique demonstrated high accuracy, it focuses on localized areal analysis. This limits the scalability of this technique for a broad impact.

As the maintenance costs from deployment, regular quality inspections, and sunk costs from false alarms and inaccurate reporting go down, overall benefit from the system goes up. Molina-Pico et al. have proposed methods to evaluate the cost-benefits of implementing early warning systems (Kelleher et al., 2018). Thus, recently proposed early warning systems which utilize modern technology for efficient and accurate natural disaster alerts are more attractive. Notably, the sharp spike in the number of active satellites due to their increase in widespread viability has enabled continuous monitoring of the entire globe. Different sensors placed on satellite networks create aerial analysis of surface temperature, water elevation, and vegetation as well as enable the detection of potential natural disasters (Varela et al., 2020). Some recent examples include the particulate level monitoring for wildfire prediction techniques by the researchers from Colorado State University (Varela et al., 2020). Another form of early warning system is based on the emerging remote sensor network: similar to satellites, wireless, land-situated sensors installed throughout regions of interest relay readings to central stations, which then process recorded data to predict or detect natural hazards. Comparatively, remote sensor networks are less costly to deploy and have higher precision, but they are narrower in scope and provide less coverage than satellite networks (Sayad and MousannifHassan Al Moatassime, 2019; Jain et al., 2020).

Table 1
A summary of previous research conducted on wildfire analysis, detection, and prediction, detailing each technique's achievements and limitations (F1 = Harmonic Mean of Specificity and Sensitivity; ACC = Accuracy).

References	Detection Technique	Technical Merit	Equipment Used	Technique Achievements	Technique Limitations
Ban et al., 2020 (Kelleher et al., 2018)	CNN on Synthetic Aperture Radar (SAR) for Wildfires	87% F1 84% ACC	SAR Satellites	• Accurate wildfire progression monitoring in challenging topographic conditions	• Can only detect burnt areas after being ravaged by fire
Molina-Pico et al., 2016 (Varela et al., 2020)	Wireless Sensor Networks (WSN) for Wildfires	95% Sensor Transmission	ZigBee for computing, Meteorological sensors	• Lays groundwork for WSNs on wildfires	• Network effectiveness drops significantly within few kilometers
Kelleher et al., 2018 (Sayad and MousannifHassan Al Moatassime, 2019)	Outdoor Aerosol Sampler (OAS) Network for Wildfires	N/A	PM Sensors, Processing Systems, Battery/Solar Panels	• Compact, weatherproof, and battery-powered	• Prone to failure in high $PM_{2.5}$ environment • Dependent on battery
Varela et al., 2020 (Jain et al., 2020)	Regression-based Wildfire Detection	100% Controlled ACC	WSN for Temperature, Humidity	• Simple detection criteria	• Conducted in controlled environment • Must not be in direct sunlight
Sayad et al., 2019 (Costa et al., 2017)	Wildfire Prediction using Remote Sensing Techniques	98.3% ACC	Satellites for various vegetation and meteorological data	• Created remote sensing database • Highly accurate classification	• Very narrow scope, hard to scale up
Jain et al., 2020 (Singla et al., 2020)	Overview of Wildfire Management	N/A (review)	Various previous research and literature	• Showed that machine learning holds potential in successful wildfire management	• Limited work in wildfire prediction

Overall, these techniques reveal a clear need for a solution that can utilize key wildfire indicators, predict wildfires with high accuracy, and allow real-time applications with data that is provided by the existing sensor networks.

3. Methods

This project has two major stages. In the first part, a comprehensive large dataset consisting of historical wildfires as well as key wildfire related leading and trailing indicators was developed. In the second part, a novel machine learning model with targeted spatial and temporal analytical features was developed specifically for wildfire prediction application. The Artificial Intelligence (AI) model is trained, optimized, and validated based on the aforementioned datasets to assess wildfire risks for locations across California.

3.1. Development of a comprehensive wildfire dataset

In order to train and evaluate the AI models effectively, a wildfire database needs to contain the location and severity of historical wildfires as well as the environmental, meteorological, and geological data that reveal indicative probabilities of future wildfires. Specifically, the objective of the training process is to teach the neural network patterns that are encoded within the historical data which corresponds to the wildfire's location and severity, enabling advanced prediction of future wildfires based on the real time data from the current sensor network.

We utilized various resources to construct the database, starting with the integration and extension of the WildfireDB dataset, which consists of the wildfires' longitude and latitude coordinates and the Fire Radiative Power (FRP) data based on the infrared readings from satellite imagery (Singla et al., 2020; Environmental Protection Agency). We expanded upon the WildfireDB to include three meteorological leading indicators, four environmental trailing indicators, and two geological factors to enable efficient AI-network training, culminating in the largest California wildfire dataset with over 37 million data points. We acquired the leading and trailing indicators from the Environmental Protection Agency's (EPA) and PurpleAir's websites, retrieved the geological factor data using the Google Earth Engine's code editor, and compiled the wildfire data from the WildfireDB dataset to form this comprehensive database (PurpleAir; Google Earth Engine; National Wildfire Coordinating Group, 1874).

3.1.1. Identifying wildfire leading and trailing indicators

We studied the causes and mechanisms of wildfires through literature and identified the key environmental and meteorological indicators related to wildfire inception and propagation independently (Gutierrez et al., 2021). We categorized these parameters into leading indicators that demonstrate certain environmental and meteorological trends in areas before the onset of the wildfires and trailing indicators that demonstrate certain environmental and meteorological trends in areas during the wildfire outbreaks.

We hypothesized that the leading and trailing indicators can be derived from the causes and effects of wildfires. In addition, we hypothesized that geological factors could also modulate the process of wildfire conception and spread.

Leading indicators, which lead to the conceptions of wildfires, illustrate particular trends in areas where sparks or lightning could produce flames. Certain patterns in environmental factors in an area produce conditions where a spark would be exacerbated into a wildfire rather than be smothered and extinguished. Additionally, optimal weather conditions would fuel small fires into raging conflagrations that would yield a much larger acreage and damage. The leading indicators include:

- Temperature (°C) governs the conditions such as precipitation, water evaporation, etc., which would set an ideal environment for wildfires (Hamadeh et al., 2016).
- Dew point (°C), the temperature to which air must be cooled to be fully saturated with water vapor, can be used to find relative humidity which indicates the dryness and thus inflammability of the area (Cruz and Alexander, 2019).
- Wind speed (knots/hour) provides a supply of air to fuel the fire and a means of carrying the burning soot to spread fire to its surroundings (Liu et al., 2016).

The trailing indicators, which often correlate with byproducts of the wildfires, show abnormal trends before a small flame transgresses into a large wildfire. After a fire begins, toxic gas and particles rise up and diffuse into the atmosphere. As compared to fires which are dependent on ideal landscape and environmental conditions to propagate, smoke spreads quickly through the air regardless of wind speed. Nearby sensors can detect the uptick and be used to determine the presence and severity of a wildfire. Hence, trailing indicators trail after the conception of wildfires. The trailing indicators include:

- $PM_{2.5}$ ($\mu g/m^3$), the concentration of particulate matter in the air with a diameter $\leq$2.5 μm (Bela et al., 2022).

- PM_{10} ($\mu g/m^3$), the concentration of particulate matter in the air with a diameter $\leq$10 μm (Bela et al., 2022).
- CO or Carbon Monoxide levels (parts per million, or ppm) (Griffin et al., 2021).
- NO_2 or Nitrogen Dioxide levels (parts per billion, or ppb) (Ramsey et al., 2020).

The city of Paradise in California was destroyed by the notorious Camp Fire in 2018 (EPA, 2021). The seven graphs depicting the leading and trailing indicator trends for the city of Paradise from June to December 2018 are shown in Fig. 1. They are inclusive of the Camp Fire time period of Nov. 8–25, 2018. The graphs show the readings from a set of sensors located ~50 miles from the Camp Fire during that period. The two weeks preceding and during the fires are highlighted. The leading indicators shown in the left graph in Fig. 1 display successive surging and dropping trends in temperature, dew point, and wind speed before the onset of the Camp Fire (Oct. 25 - Nov. 8). The trailing indicators shown in the right graph in Fig. 1 display the sudden spikes in particulate, carbon monoxide, and nitrogen dioxide levels after Nov. 8, which lasted past the end of the Camp Fire date of Nov. 25. These trends provide the basis of predicting the risk of onset and the spread of potential wildfires.

3.1.2. Processing environmental and meteorological data

The leading and trailing indicators discussed in Section 3.1.1 were acquired from the Environmental Protection Agency (EPA) historical database (Gonzalez and Woods, 2008). Each environmental and meteorological factor contains readings from weather sensors across the United States. These data were partitioned into matrices of daily sensor values at each sensor location.

The daily average sensor readings were plotted onto a two-dimensional longitude and latitude spatial plot, which was then overlaid onto the map of California. Since the sensor network coverage is not uniform across the state, denser at populous locations and sparse in the wilderness, we interpolated the data to create an extrapolated environmental/meteorological factor map, turning a series of discrete readings into an analog map. All the data was transformed through a logarithmic function defined by $y = log_2(x + 1)$. This function compresses the extreme readings during wildfires, where some stations' sensor measurements spiked up by $> 50 \times$ from their normal levels. Because environmental and meteorological data tend to be mostly around normal levels with a few days/instances of large changes, a logarithmic transformation helps bound the output range (e.g., a large linear range from 1 to 1024 maps to a compact range of 0–10 using log-base-2 transformation). This enables a more uniform output distribution that enables smoother neural network training and classification and as such is frequently used in computer vision and image processing applications (Mansoor et al., 2022). After data processing, each spatial sensor reading map was saved as a 500 $\times$ 500 pixel image, with each pixel representing an area that is 2.53 km $\times$ 2.53 km.

3.1.3. Processing geological data

Terrain characteristics, including elevation and local vegetation productivity, are important in dictating the likelihood of the onset of a wildfire and its potential to spread. Terrain elevation dictates much of the environmental and atmospheric processes occurring within the area (CAL Fire). At higher elevations, a lower level of oxygen and humidity coupled with higher temperatures mean that fires will have a lower likelihood of spreading. Additionally, higher altitudes usually signify mountainous areas with sloped terrain, which could mean a lusher

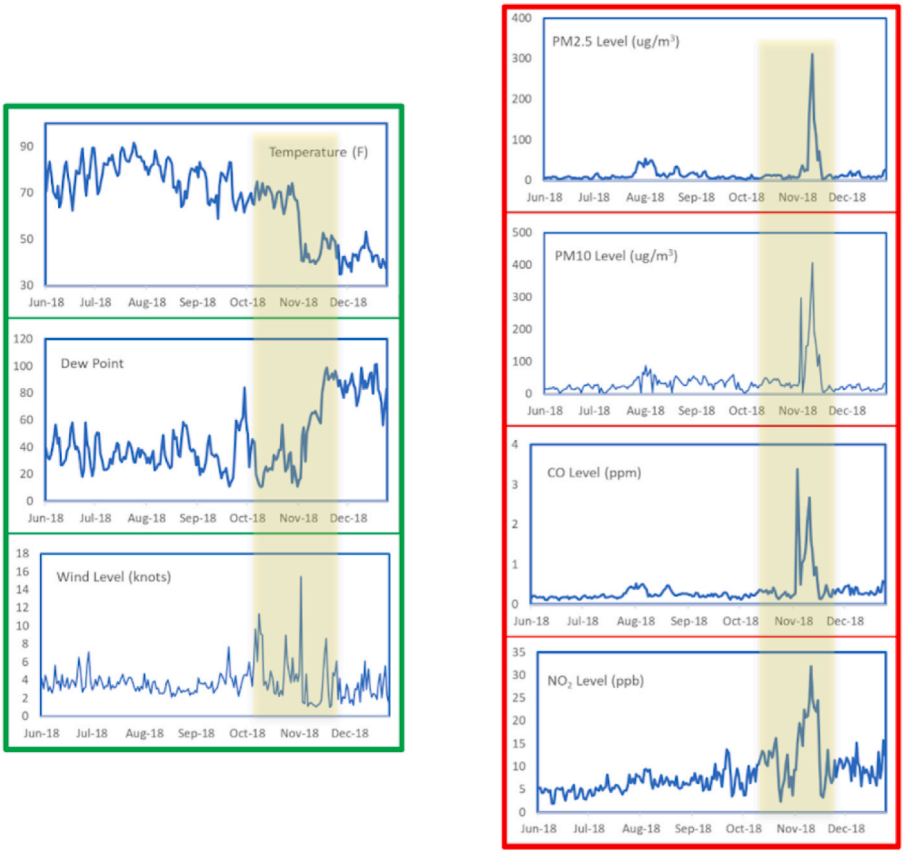


Fig. 1. Graphs of leading indicators including temperature, dew point, wind speed (shown in the left figure), and trailing indicators including $PM_{2.5}$, PM_{10}, CO, and NO_2 concentrations (shown in the right figure), from June to December of 2018 over Paradise, California. The relevant time window associated with the devastating Camp Fire's period is highlighted on the graphs.

windward side that is less prone to serious fires or a drier landscape on the leeward side's rain shadow that is more combustible (Huang et al., 2021). Along with the environmental and meteorological indicators, the location where the model makes a prediction should also depend on the geological specifications of the area under consideration, which should be integrated into the wildfire prediction model.

3.1.4. Processing vegetation data

Vegetation productivity, measured by the Normalized Difference Vegetation Index (NDVI), is an indicator of the amount of infrared energy radiated given the absorbed influx of solar energy into a certain area (Deshmukh et al., 2020). The working chlorophyll cells in the productive plants emit infrared light. The higher the NDVI index, the more productive the vegetation, and vice versa.

Plant health and productivity greatly reflect the conditions of the terrain surrounding the plant: unhealthy plants could indicate an unusually dry or hot environment and increased volatility for large fires. Additionally, a lower NDVI index could indicate withering plants or dead foliage, the perfect fuel for a conflagration to burn through rapidly.

Similar to the leading and trailing indicator data, the elevation (in meters) and NDVI (no units) were processed to generate two-dimensional spatial plots measuring 500 × 500 pixels in size. This ensures that at each location, the leading indicator, trailing indicator, elevation, and NDVI plots all have the same data resolution.

3.1.5. Processing wildfire data

We collected historical wildfire data in California from WildfireDB, an open-source dataset maintained by UC Riverside, Vanderbilt University, and Stanford University (Environmental Protection Agency). The dataset originally contained 17 million data points with daily wildfire details from 2012 to 2018. Specifically, the wildfire activity is quantified through Fire Radiative Power (FRP), which is a measure of radiated infrared energy per unit of time directly correlating to the rate of biomass burning (Singla et al., 2020). To ensure that smaller fires with low FRP measurements are not ignored, we transformed the data through a fourth-root scale as defined by $y = \sqrt[4]{x}$ to express fire activity as a function of temperature according to the Stefan-Boltzmann law (which states that $P \propto T^4$).

Additional data processing was done in order to utilize the dataset for neural network training. We transformed the dataset to form two-dimensional images, with the x and y coordinates of each pixel corresponding to its respective latitude and longitude location. The pixel intensity value represents the measured or extrapolated level of an indicator at that specific location. As these indicators largely follow cyclic trends that revolve around the human circadian rhythm, the average measured indicator level at each location over a 24-h period is used to represent the daily values of that location. For each indicator, the final output is a sequence of image arrays in chronological order. The 2D data maps contain patterns within the spatial domain. Trends over time introduce a temporal progression to the two-dimensional images sequences. Both the spatial and temporal aspects of the database must be analyzed together for wildfire prediction. This insight led to the development of a unique machine learning architecture as narrated below.

3.2. Spatio-temporal U-convolutional-LSTM neural network development

To exploit the spatial and temporal nature of the input data consisting of the leading and trailing indicators of wildfires, we created a spatio-temporal U-Convolutional-Long Short-Term Memory (LSTM) network. The structure of this unique neural network framework is depicted in Fig. 2. The name U-Convolutional-LSTM (ULSTM) references its architecture and purpose. Spatio-temporal refers to the model's ability to effectively process and analyze the data using both spatial and temporal feature extraction capabilities. Hybrid and deeply meshed spatio-temporal neural network architectures have shown to be effective in analyzing complex datasets where key information is encoded in both spatial and temporal dimensions concurrently (Kim et al., 2020; Ronneberger et al., 2015). Information in the spatial domain includes data represented as arrays and matrices, making spatial analysis important in image-related tasks such as picture classification, object recognition, and image generation. Information within the temporal domain includes data represented in sequences, making temporal analysis important in chronological time-series data such as text analysis, music generation, speech recognition, and video manipulation.

The first fundamental aspect of the ULSTM machine learning architecture is the U-Convolutional network, or U-Net architecture. The U-Net framework was originally proposed for biomedical imaging applications by Ronneberger et al., in 2015 (Shelhamer et al., 2016) and was built on the fully-convolutional network reported by Shelhamer et al., in 2014 (Hochreiter and Schmidhuber, 1997). The first half of a U-Net consists of a series of convolutional layers, each of which transforms the input image into a smaller, compressed image containing the extracted features. Specifically, the convolutional layers pass a two-dimensional filter over the input image, returning an output matrix with each cell equal to the filter and the region of the input image the filter passes over. At the end of this series of convolutional layers is a compact matrix of key spatial features. The second half of the network is structurally symmetric to the first half of the network mirroring the convolutional layers with a series of deconvolutional layers. While convolutional layers compress an input image, deconvolutional layers expand an input image. These deconvolutional layers perform the same operations that a convolutional layer does but with uniform padding that increases the dimensions of the input image. Overall, the U-Net is a generative network that specifically uses an input image to generate an output image of the same dimensions containing key extracted spatial features.

The second fundamental aspect of the ULSTM framework is the Long Short-Term Memory network. Originally proposed by Hochreiter et al. and subsequently refined by many researchers, the LSTM network is a

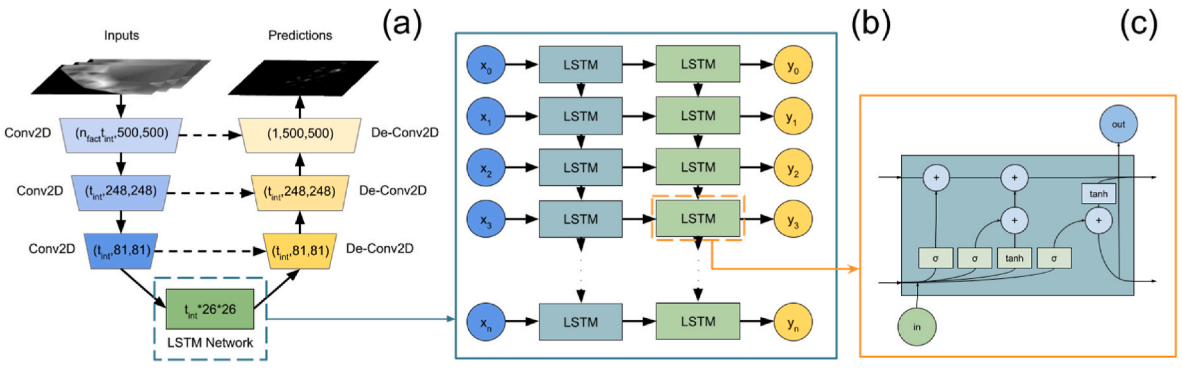


Fig. 2. Architecture of the ULSTM neural network. (a) A U-shaped network framework similar to that of the U-Net, where the dashed arrows indicate symmetry of the U-Net by comparing analogous parts. (b) A two-layer LSTM network component. (c) The architecture of the LSTM unit within the LSTM network.

type of recurrent neural network that specializes in temporal feature analysis (Zhang and Sabuncu, 2018). By selectively learning long-term dependencies from previous inputs, LSTM networks recall stored input sequence information to synthesize outputs. Thus, LSTM networks extract key temporal features from a chronological list of inputs and process them to form outputs. This is especially helpful for data that are time-sequential, such as consecutive frames of a video or the words in a sentence.

The ULSTM network we developed in this project combines the merits from both the U-Net and LSTM network's structures. The ULSTM can take in spatial features from images that are sequenced in a temporal manner as its input. Specifically, we designed the ULSTM architecture to process and analyze two weeks worth of environmental and meteorological factor maps to generate a risk map that details the probability of a fire starting at any given location. The overall architecture of the ULSTM network is analogous to that of the U-Net, with its first half consisting of a series of convolutional layers, which compress the input images to a series of feature matrices. As shown in Fig. 2, these include three Conv2D layers which transform input data from dimensions of ($n_{int} \times t_{int}$, 500, 500) to (t_{int}, 248, 248), then to (t_{int}, 81, 81), and finally to (t_{int}, 26, 26), where n_{int} represents the number of environmental and meteorological factors and t_{int} represents the prior time duration in days that the network analyzes. These feature matrices are then flattened and propagated through a two-layer LSTM network of input and output size $t_{int} \times 26 \times 26$, which analyzes the feature matrices while retaining short-term memory of the matrices it processes. The LSTM network's output is then passed through the second half of the U-Net, where a series of De-Conv2D layers, which are symmetric to the first half of the U-Net, up-scale the feature matrices from (t_{int}, 26, 26) to (t_{int}, 81, 81), then to (t_{int}, 248, 248), and finally to (1, 500, 500) as the predicted wildfire heatmap.

We evaluated the performance of the neural network using the Binary Cross-Entropy function and trained the model to minimize the loss value. Binary Cross-Entropy is a widely used loss function which indicates the proximity of a set of output values to their target labels for true-false classification problems (LeCun et al., 1998). Here, the loss function treats each pixel on a prediction heatmap as an output value and each corresponding pixel on the actual heatmap as its target label. The loss function is defined in Equation (1), where $L(y)$ is the final loss function, y is the set of labels ("1" for yes wildfire, "0" for no wildfire), $p(y)$ is the probability of the given label, and N is the total number of pixels on a prediction heatmap. The closer the predicted output is to the target label, the smaller the loss value is. Thus, iterative tuning of the weights and biases of the neurons of the network using a gradient-descent backpropagation algorithm to minimize this loss value trains and optimizes the model.

$$L(y) = -\frac{1}{N}\sum_{o}^{N} y_i \bullet log(p(y_i)) + (1 - y_i) \bullet log(1 - p(y_i)) \tag{1}$$

The ULSTM neural network, along with all other codes in this project (including the data parsing, 2D extrapolation, probability mapping, and other programs) were written in Python, and the machine learning models were implemented using the PyTorch Libraries.

4. Results

Using the wildfire dataset described in Section 3.1, different designs of the ULSTM neural network were trained and evaluated for wildfire prediction accuracies.

The formula used to define the network's accuracy is given in Equation (2), where t_{int} represents the number of days being analyzed to make a prediction; $F[t']_{i,j}$ and $T[t']_{i,j}$ represent the value of the pixels at position i,j of the predicted risk heatmap and true wildfire heatmap (respectively) on day t'; and t represents the time duration.

$$ACCURACY = 1 - \frac{1}{2}\sum_{t_0=0}^{t}\left(\frac{1}{t_{int}}\sum_{t_1=0}^{t_{int}}\sum_{i,j} f\left(F[t_0 - t_1]_{i,j}\right) - \sum_{i,j} T[t_0]_{i,j}\right)^2 \tag{2}$$

Using a modified version of mean-squared-error analysis, we measured the differences between the network's predicted risk heatmap and the true wildfire heatmap.

We designed four ULSTM networks to analyze the contribution from the various leading and trailing indicators to the network's final prediction accuracies. All ULSTM models share the same architectural features and specifications, but each has a unique set of indicators as inputs. The network structures, input parameters and channels, and the results from these four models are listed in Table 2. For each input data source type, there exists seven channels (one for each heatmap of the prior seven days). For example, Model V3 has twenty-one input channels: seven for wind, seven for temperature, and seven for humidity.

The models were trained to predict impending wildfire risks within the following 24 h based on data from the prior seven days. The environmental, meteorological, geological, and overall wildfire occurrence data were split into training and testing datasets. The training dataset used to train the network included data from 2012 to 2017, and the testing dataset included data from 2018 to 2022.

Model V1 variant of the ULSTM network analyzes all four trailing indicators on air quality ($PM_{2.5}$, PM_{10}, CO, NO_2), achieving an accuracy of 89.3%. Model V2 takes in two meteorological leading indicators and two air-quality trailing indicators (humidity, temperature, $PM_{2.5}$, CO), achieving an accuracy of 92.8%. Model V3 relies on the three meteorological leading indicators (wind speed, temperature, and humidity) to achieve 96.3% in accuracy. Finally, the Model V4 variant takes in all seven total leading and trailing indicators (wind speed, temperature, humidity, $PM_{2.5}$, PM_{10}, CO, NO_2), achieving the overall highest accuracy of 97.1%.

The performance of the proposed ULSTM spatio-temporal neural networks in predicting large wildfires were benchmarked against a traditional Convolutional Neural Network (CNN) architecture represented by the seminal LeNet-5 model (Bhowmik and Most, 2022). The LeNet-5 architecture consists of three sets of convolutional layers followed by two fully connected layers. Using the optimized set of seven leading and trailing indicators used for Model V4, LeNet-5 was trained with the same dataset and used as a benchmark for the four ULSTM network variants. This CNN model achieved an accuracy of 76.4%.

As shown in Table 2, all four versions of the ULSTM network achieved significantly higher accuracy compared to the reference CNN network. Although the CNN is good for extracting key spatial features within the input data for making wildfire predictions, it is not adequate in recognizing the time-related patterns amongst image sequences, leading to a lower accuracy for predicting wildfires based on environmental and meteorological factors. These results illustrate that the ULSTM networks with their unique spatio-temporal architectures are better suited to the wildfire dataset analysis than the traditional CNN.

Also it is demonstrated through this study that the comprehensive incorporation of leading and trailing indicators into the training data improves the model's accuracy. Leading indicators containing patterns or trends can provide insight into the likelihood of onset of wildfires in the near future. It also captures information on land volatility on the potential of small fires evolving into severe conflagrations. Meanwhile, the four air-quality factors contain patterns or trends that trail the initial outbreak of a fire, reflecting noxious emissions from the wildfires.

For wildfire prediction, it is intuitively beneficial to look for trends in leading indicators to foretell potential areas of wildfire days in advance. This conclusion is supported by the comparison results from Models V1 vs. V3 and Models V2 vs. V4. Model V1 was trained on trailing air-quality indicators only and Model V3 was trained with the leading meteorological indicators only. Models V1 and V3 achieved 89.3% and 96.3% prediction accuracy, respectively, demonstrating that leading indicators are better signals of impending wildfires than trailing

Table 2
Four variations of the ULSTM networks are compared with a LeNet-5 CNN as benchmark.

Models	Air-Quality Data	Meteorological Data	Period Used for Prediction	Input Training Parameters	Total # of Inputs	Accuracies
LeNet-5 CNN	4 sources	3 sources	7 days	All Parameters	49 channels	76.4%
Model V1	4 sources	N/A	7 days	$PM_{2.5}$, PM_{10}, CO, NO_2	28 channels	89.3%
Model V2	2 sources	2 sources	7 days	Humidity, Temp, $PM_{2.5}$, CO	28 channels	92.8%
Model V3	N/A	3 sources	7 days	Wind, Temp, Humidity	21 channels	96.3%
Model V4	4 sources	3 sources	7 days	All Parameters	49 channels	97.1%

indicators. Model V2 was trained on two trailing air-quality indicators and two leading meteorological indicators. Model V2's accuracy of 92.8% is between those of Model V1 and Model V3, confirming that leading indicators boost model accuracy more than trailing indicators. Model V4 was trained on the full set of leading and trailing indicators, achieving the highest prediction accuracy of 97.1%. As the four LSTM networks were all trained using the same hyperparameters and have the same overall structure, the difference in performance between these variants lies within the nature of their input data sources. This signifies the importance of leveraging all environmental and meteorological indicators in wildfire prediction.

Overall, the Model V3 variant has a higher accuracy than that of Model V1 and Model V2, showing that leading indicators are better predictors of wildfires than trailing indicators. Additionally, the Model V4 variant has a higher accuracy than that of Model V3. This is attributed to the fact that the trailing indicators do provide some additional insights into wildfire activity. For areas where meteorological factors could not have predicted the onset of wildfires, such as deliberate and expedited arson, trailing indicators provide extra patterns and trends to gather information from them and utilize in computing the predictions. Alternatively, areas with poorer air quality could indicate recent burning from wildfires, allowing the model to determine that wildfires likely will not occur in that area in the near future.

The wildfire predictions using the ULSTM Model V4 were applied to analyzing the 2018 wildfire season, the results of which are shown in Table 3. Predicted wildfire risks are compared with the actual occurrences of five large wildfires in California during that year. Specifically, the predicted wildfire risk-map output of the network is shown with each pixel value scaled between 0 and 1 for the purpose of visualization (0 representing the lowest risk and 1 representing the highest risk). Above each predicted image, the dates on which the wildfires were predicted are annotated. The actual fire maps and dates of occurrences are listed above the images. The predicted wildfire maps highlight areas of high likelihood of wildfires. The names of wildfires and the time intervals from model predictions to actual wildfire occurrences are listed on the top of the table. Each circled area on the map represents a region of high probability of wildfires or the actual wildfires. The circle 1 highlights the Ferguson Fire, circle 2 highlights the Carr Fire, 3 highlights the Mendocino Complex Fire, 4 highlights Hirz Fire, 6 highlights Stone Fire, 7 highlights Camp Fire, and 9 highlights Woosley Fire. Circles 5 and 8 show smaller wildfires predicted by the model which also showed up in the satellite FRP images. For the 2018 wildfire season, seven of California's ten largest wildfires were successfully predicted by Model V4 of the ULSTM network 5–14 days in advance.

Geological information was integrated into the training dataset to further optimize Model V4's wildfire prediction accuracy. The impact of geological factor data (including NDVI vegetation data and terrain elevation data) incorporation to the ULSTM network's prediction accuracy was analyzed over the 2021 wildfire season. The names of the wildfires, the dates of the initial outbreak, and the corresponding locations on the map of California are listed in Table 4. The prediction maps from the network with and without geological input data are listed for comparison. No scaling was applied on these images for accurate comparison. The ULSTM Model V4, trained with geological data, shows higher prediction accuracy as demonstrated by the brighter pixel values indicating higher confidence levels in predicting the Antelope and Beckwourth Complex fires. It also shows a better location prediction accuracy for the Dixie and Beckwourth Complex fires. The inclusion of the geological data provides the ULSTM network with additional information and insights into the likelihood of wildfire outbreaks. Overall, the ULSTM Model V4 with geological data was the best approach to predict impending wildfires with higher confidence and location precision.

Between 2018 and 2021, there were 207 recorded wildfires that were over 1000 acres across California. The wildfires from each fire season are rank-ordered and plotted with their corresponding burned areas in Fig. 3. The severity and frequency of wildfires from each fire season can be visualized clearly.

After finalizing the optimization of the ULSTM Model V4, we applied the trained model to conduct analysis on the 207 California wildfires

Table 3
Comparison between predicted results from the ULSTM (Model V4) and the actual fire map in 2018 fire season. Each circle on the map area is correlated to a reported wildfire:1. Ferguson Fire, 2. Carr Fire, 3. Mendocino Complex Fire, 4. Hirz Fire, 6. Stone Fire, 7. Camp Fire, and 9. Woosley Fires. 5 and 8 smaller wildfires predicted by the model.

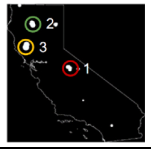
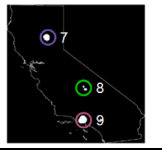
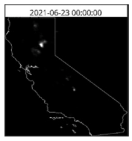
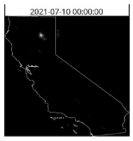
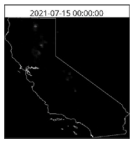
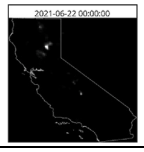
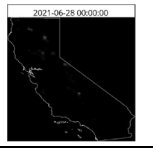
Wildfires	Ferguson	Carr, Mendocino	Hirz	Stone	Camp + Woosley
Advanced Warning	8 days	14 days	5 days	9 days	11 days
Predicted	July 5, 2018	July 13, 2018	August 4, 2018	August 6, 2018	October 27, 2018
					
Actual	July 13, 2018	July 27, 2018	August 9, 2018	August 15, 2018	November 2018
					

Table 4
Impact of geological data on ULSTM model's prediction accuracy. Results compared based on California's 2021 wildfire season.

Wildfire	Date	Map	Prediction Map with Geological Data	Prediction Map without Geological Data
Dixie	07/13			
Antelope	08/01			
Beckwourth Complex	07/03			

from 2018 to 2021 for retroactive statistical study and 20 wildfires in 2022 for forward-looking study. The prediction accuracy statistics for 2018–2021 is plotted in Fig. 4 and summarized in Table 5 for reference. The model's prediction accuracy is better for larger wildfires, detecting the most devastating wildfires (>300 kilo acres) with >85% accuracy. The model's accuracy reduces to 60% for mid-sized fires (15–150 kilo acres) and ~25–50% for small fires (<3 kilo acres).

Thus, the ULSTM model we developed is successful in forecasting the larger wildfires. However, statistically, it is not as successful in predicting the smaller wildfires. There are many reasons contributing to this outcome. Primarily, the prediction rate gap is attributed to the origin of these wildfires, many of the unpredicted wildfires were started due to human factors such as arson, power line failure, or home fires. Some of the wildfires were also caused by lightning, which is not a training input parameter that is comprehended in the dataset. Relative to smaller wildfires, the majority of the large wildfires are more heavily dependent on geological and environmental conditions, therefore being predicted with higher rates.

To further test the neural network model's applicability for real-time wildfire prediction, we applied the model to the 2022 wildfire season (Table 6). The table includes wildfire information, the ULSTM's ability to detect it, and either the date the fire was predicted or the reason for the fire not being predicted. The reasons for a wildfire not being predicted includes "Location Off" (the location of high-probability pixels are located one county away from the actual wildfire epicenter), "Too Faint" (the highest value of the pixels representing a potential fire is less than 0.5), or "None at all" (no coherent group of pixels lies near the wildfire epicenter). If none of the aforementioned reasons are true, the fire was marked as predicted.

As of Oct. 14, 2022, there were 19 fires above 1000 acres in the 2022 wildfire season. The model consistently demonstrated a higher prediction rate for large wildfires, detecting 60% of the 5 largest wildfires and 40% for the 5 smallest wildfires, similar to the results from the 2018 to 2021 wildfire seasons. Overall, the ULSTM Model V4 demonstrated a 42.1% wildfire prediction rate for the 2022 wildfire season as compared to 41.8% over the period between 2018 and 2021. For larger wildfires above 15,000 acres, the model demonstrated a 60% prediction rate for the 2022 wildfires season as compared to 57.1% between 2018 and 2021.

5. Future work

A personalized early warning system, which would give user-specific alerts and precautions given their circumstances and special conditions, would have a great impact on those in areas at risk of wildfires, especially the individuals with disabilities or specific health conditions. A raw percentage value or recent trend of wildfire risk would have varied interpretations from different members of the same community, leading to mixed and likely uncoordinated responses. An early warning system could eliminate such confusion and provide actionable items for civilians and fire protection agencies. Additionally, the personalized aspect of the system would allow more vulnerable populations, such as those with respiratory diseases, sensory impairments, or mobility disabilities,

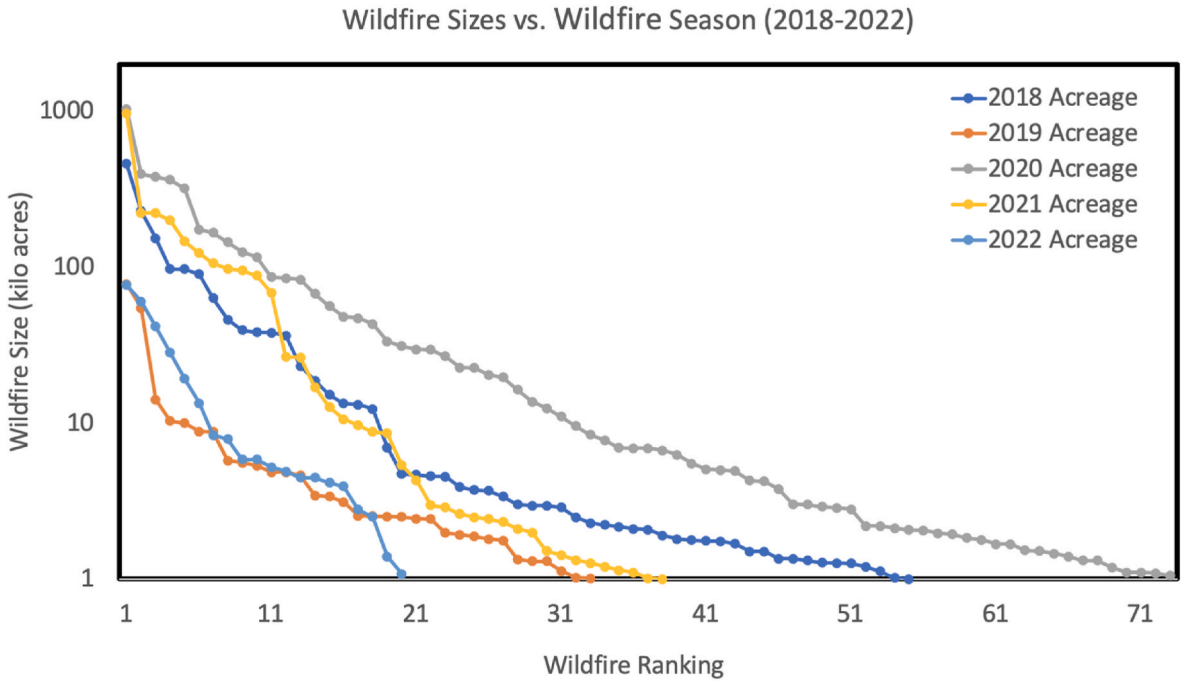


Fig. 3. Wildfire burned acreage distribution between 2018 and 2022. The wildfires are ranked by their burned acreage and color coded by the fire season.

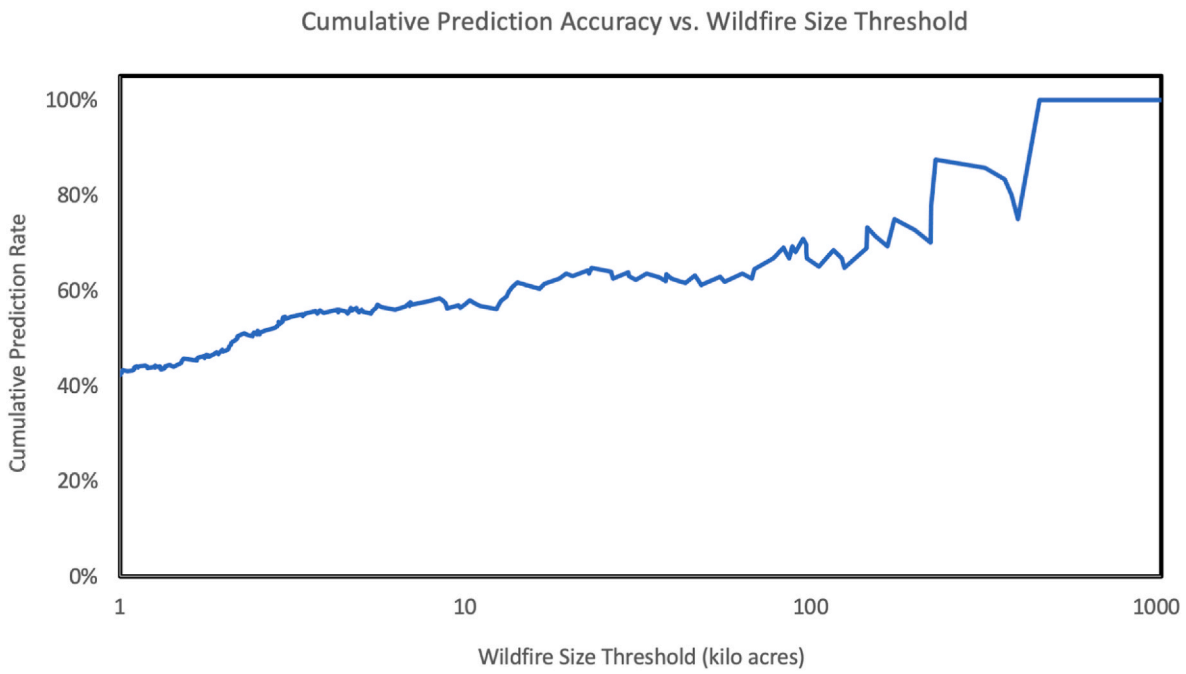


Fig. 4. Summary of all the California wildfires surveyed from 2018 to 2021. The prediction performance of the ULSTM Model V4 versus wildfire size threshold is shown.

Table 5
Wildfire prediction rate statistics based on the burned acreage range for wildfires between 2018 and 2021.

Thresholds	Low Thresholds (acreage)	High Thresholds (acreage)	# Detected	Total #	Prediction Rate
300 and up	300,000	1,500,000	6	7	85.7%
150 to 300	150,000	300,000	4	7	57.1%
75 to 150	75,000	150,000	10	16	62.5%
30 to 75	30,000	75,000	8	15	53.3%
15 to 30	15,000	30,000	8	14	57.1%
7.5 to 15	7500	15,000	9	19	47.4%
3 to 7.5	3000	7500	16	34	47.1%
1 to 3	1000	3000	23	87	26.4%

to better prepare in advance to plan the extra steps they would need to take.

Fig. 5 details the specifications of the proposed personalized early warning system. Utilizing the trained ULSTM neural network model to extract wildfire predictions, the warning system would identify the areas of high wildfire probabilities. The information would then be utilized by fire protection agencies as well as the early warning algorithm to decide if and to whom to send alerts. For example, individuals with special conditions would receive warnings at different risk levels and in proximity to predicted fires. Such adjustments would accommodate for the degree to which certain individuals are impacted by fires and their byproducts and/or are able to evacuate. These populations include those with:

Table 6
Wildfire outbreaks in the 2022 wildfire season were compared with the predicted results to quantify real-time prediction rate. The Airport Fire (highlighted in gray) was excluded due to it occurring before the recording started.

Name	County	Acres	Start Date	End Date/Containment Status	Detected?	Date/Reason
Mosquito Fire	Placer, El Dorado	76788	September 6	95% contained	YES	August 17
McKinney Fire	Siskiyou	60138	July 29	September 7	NO	Location off
Six Rivers Lightning Complex	Humboldt, Trinity	41596	August 5	97% contained	YES	July 25
Fairview Fire	Riverside	28307	September 5	98% contained	YES	August 14
Oak Fire	Mariposa	19244	July 22	August 10	NO	Location off
Mountain Fire	Siskiyou	13440	September 2	September 21	NO	None at all
Red Fire	Mariposa, Madera	8364	August 4	September 28	YES	July 23
Yeti Fire	Siskiyou	7886	July 29	September 1	YES	June 27
Lost Lake Fire	Riverside	5856	May 26	June 8	NO	Location off
Barnes Fire	Modoc	5843	September 7	99% contained	NO	Location off
Route Fire	Los Angeles	5208	August 31	September 7	NO	Too faint
Washburn Fire	Mariposa, Madera	4886	July 7	July 30	NO	None at all
Electra Fire	Amador, Calaveras	4478	July 4	July 28	YES	June 19
Border 32 Fire	San Diego	4456	August 31	September 5	NO	Location off
Airport Fire	Inyo	4136	February 16	February 26	N/A	N/A
Mill Fire	Siskiyou	3935	September 2	September 13	NO	Location off
Rodgers Fire	Tuolumne	2790	August 8	September 26	YES	July 24
Thunder Fire	Kern	2500	June 22	June 27	NO	Too Faint
Summit Fire	Tulare	1394	August 3	90% contained	YES	July 20
Radford Fire	San Bernardino	1079	September 2	September 30	NO	Location off

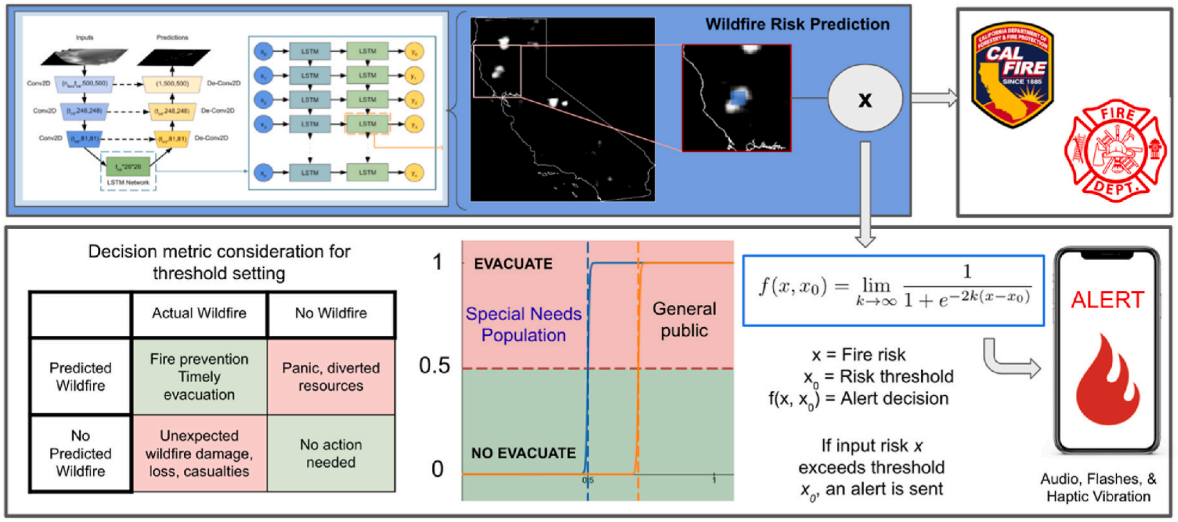


Fig. 5. Diagram of the proposed early warning system based on wildfire prediction. The ULSTM uses environmental and meteorological data to generate risk maps over California, which is directly sent over to CalFire, fire departments, and other wildfire protection agencies. For regions of significantly high risk, impacted communities and fire protection agencies would be notified of potential fires and evacuation instructions. Individuals with specific circumstances, such as respiratory exacerbation risks, would have different warning criteria and content.

- Health ailments that would make one more susceptible to environmental, physiological, or psychological stress,
- Mobility disabilities that would hamper one's ability to evacuate,
- Mental conditions that would alter one's ability to comprehend an act upon alerts and warnings.

For example, individuals with respiratory conditions such as Chronic Obstructive Pulmonary Diseases (COPD) or asthma, who are much more sensitive to smoke and pollution generated from wildfires, would be notified at a lower predicted risk level and larger distance from predicted fires than healthy individuals. Additionally, they would receive specific instructions to protect themselves, such as to stay indoors and wear masks.

Additionally, smoke and particulates from wildfires create a large amount of pollution which negatively affects air quality. Breathing air quality is one of the most crucial factors in human health; poor air quality can cause any person's health to significantly deteriorate and is an increasingly important issue following the advent of rapid industrialization. Especially since their lungs are compromised due to inflammation, individuals with respiratory diseases such as COPD and asthma are extremely susceptible to exacerbation caused by bad air quality, leading to hospitalization. Based on the results of retrospective medical studies, a methodology to estimate the increase in exacerbation risk of respiratory conditions has previously been reported, using the four trailing indicators near the patient's location. Specifically, if a factor falls below the threshold standard, its contribution to the final risk percentage is zero; otherwise, it follows the formula as shown in Equation (3) [46].

$$R = 1\% \cdot \frac{\left([PM_{2.5}] - 12\frac{\mu g}{m^3}\right)}{10.\frac{\mu g}{m^3}} + 0.8\% \cdot \frac{\left([PM_{10}] - 7\frac{\mu g}{m^3}\right)}{10.\frac{\mu g}{m^3}} + 2\% \cdot \frac{\left([NO_2] - 101.23\frac{\mu g}{m^3}\right)}{10.\frac{\mu g}{m^3}} + 2\% \cdot \frac{20°C - T_C}{1°C} \tag{3}$$

6. Conclusions

In summary, a novel spatio-temporal machine learning framework, ULSTM, based on a blended U-Net and LSTM neural network architectures to predict wildfires was developed in this research project. We created a large comprehensive wildfire database containing environmental and meteorological indicators, geological information, and wildfire data spanning from 2012 through 2018. This database, which contains over 37 million data points, was used to train, evaluate, and optimize the ULSTM neural network successfully. The model achieved up to 97% accuracy, outperforming the traditional CNN machine learning architecture. It predicts 7 out of the 10 largest wildfires of the 2018 wildfire season in California up to two weeks in advance. The model is also shown to predict 85% of the wildfires >300 kilo acres between 2018 and 2022.

Despite the ULSTM performing less accurately in predicting smaller wildfires, its ability to predict large fires accurately nevertheless is impactful. Large wildfires propagate due to favorable conditions for spread as well as slow reactions from firefighters due to inaccessibility and/or lack of information, turning small fires into larger ones. For example, the three largest wildfires alone in 2018 accounted for ~70% of monetary loss and ~45% acreage burnt in California. Bringing attention to potential large fires will address a majority of damage caused by wildfires altogether.

Overall, the unique spatio-temporal neural network architecture demonstrated relatively high accuracy in pinpointing the location and severity of the largest wildfires of California ahead of their actual occurrences. Thus, the deployment of such a technique would allow fire protection agencies and firefighters to divert resources to high-risk areas, confront fires early, and mitigate potential damages to the economy, infrastructure, and quality of life. However, this technique is limited in predicting wildfires due to human factors and lightning. This solution should not solely dictate the resulting response towards potential wildfires; rather, it is meant as an assistive tool for ensuring preparedness and optimal decision making during the wildfire season. It can be used to better handle the prescribed burns, station the firefighters, provide early warning to the public, and prioritize targeted evacuations in advance.

Besides the United States, countries such as Australia, Spain, and Italy are suffering from unusually severe wildfire seasons and would benefit from such a predictive solution as well. Thus, a broad-scope wildfire prediction tool could enable proactive intervention, save lives, protect the environment, and prevent or reduce damages. Expanding the scope of this wildfire prediction solution to the entire U.S. and to the world would create a much broader impact. Additionally, the proposed early warning system would be extended in the future to include personalized alerts for at-risk populations with specific health

conditions, such as respiratory diseases, sensory deficiencies, and mobility disabilities.

Credit author statement

Rohan T. Bhowmik: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. Youn Soo Jung: Methodology, Data Curation, Writing - Review & Editing, Supervision, Juan A. Aguilera: Methodology, Data Curation, Writing - Review & Editing, Supervision, Mary Prunicki: Methodology, Data Curation, Writing - Review & Editing, Supervision, Kari Nadeau: Methodology, Writing - Review & Editing, Supervision, Project administration.

Declaration of competing interest

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.

Data availability

Data will be made available on request.

References

Balch, J.K., et al., 2017. Human-started wildfires expand the fire niche across the United States. Proc. Natl. Acad. Sci. U.S.A. 114, 2946–2951.

Ban, Y., Zhang, P., Nascetti, A., Bevington, A.R., Ma, Wulder, 2020. Near real-time wildfire progression monitoring with sentinel-1 SAR time series and deep learning. Sci. Rep. 10, 1322. https://doi.org/10.1038/s41598-019-56967-x.

Bela, M.M., Kille, N., McKeen, S.A., Romero-Alvarez, J., Ahmadov, R., James, E., et al., 2022. Quantifying carbon monoxide emissions on the scale of large wildfires. Geophys. Res. Lett. 49, e2021GL095831 https://doi.org/10.1029/2021GL095831.

Bhowmik, R.T., Most, S.P., 2022. A personalized respiratory disease exacerbation prediction technique based on a novel spatio-temporal machine learning architecture and local environmental sensor networks. Electronics 11, 2562. https://doi.org/10.3390/electronics11162562.

CAL Fire. "CAL fire news release: CAL FIRE investigators determine cause of the Camp fire" [Online]. Available: https://www.fire.ca.gov/media/5121/campfire_cause.pdf.

Costa, B.S.C., et al., 2017. The use of fire radiative power to estimate the biomass consumption coefficient for temperate grasslands in the atlantic forest biome. Revista Brasileira de Meteorologia 32 (2), 55–260. https://doi.org/10.1590/0102-77863220004 [Online]. Available:

Cruz, M.G., Alexander, M.E., 2019. The 10% wind speed rule of thumb for estimating a wildfire's forward rate of spread in forests and shrublands. Ann. For. Sci. 76, 44. https://doi.org/10.1007/s13595-019-0829-8.

Dennison, P.E., Brewer, S.C., Arnold, J.D., Moritz, M.A., 2014. Large wildfire trends in the western United States, 1984–2011. Geophys. Res. Lett. 41, 2928–2933.

Deshmukh, S., Raj, B., Singh, R., 2020. Multi-task learning for interpretable weakly labelled sound event detection. https://doi.org/10.48550/arXiv.2008.07085.

Environmental Protection Agency. Pre-generated data files [Online]. Available: https://aqs.epa.gov/aqsweb/airdata/download_files.html.

EPA, 2021. Air Data: Air Quality Data Collected at Outdoor Monitors across the US. United States Environmental Protection Agency, Washington, D.C., U.S [Online]. Available: https://www.epa.gov/outdoor-air-quality-data. (Accessed 10 August 2022).

Gonzalez, R.C., Woods, R.E., 2008. Digital Image Processing, third ed. Pearson, New Jersey, pp. 109–110.

Google Earth Engine. Earth engine data catalog" [Online]. Available: https://developers.google.com/earth-engine/datasets/.

Griffin, D., et al., 2021. Biomass burning nitrogen dioxide emissions derived from space with TROPOMI: methodology and validation. Atmos. Meas. Tech. 14 (12), 7929–7957. https://doi.org/10.5194/amt-14-7929-2021.

Gutierrez, A.A., et al., 2021. Wildfire response to changing daily temperature extremes in California's Sierra Nevada. Sci. Adv. 7 (47), eabe6417. https://www.science.org/doi/10.1126/sciadv.abe6417.

Halofsky, J.E., Peterson, D.L., Harvey, B.J., 2020. Changing wildfire, changing forests: the effects of climate change on fire regimes and vegetation in the Pacific Northwest. USA. Fire Ecol 16, 4. https://doi.org/10.1186/s42408-019-0062-8.

Hamadeh, N., Karouni, A., Daya, B., Chauvet, P., 2016. Using correlative data analysis to develop weather index that estimates the risk of forest fires in Lebanon & mediterranean: assessment versus prevalent meteorological indices. Case Studies in Fire Safety 7, 12. https://www.sciencedirect.com/science/article/pii/S2214398X16300127.

High, V., 2022. California's hellish 2018 wildfires cost the U.S. Economy $148.5 billion. Dec. 7. https://epic.uchicago.edu/news/californias-hellish-2018-wildfires-cost-the-u-s-economy-148-5-billion/. (Accessed 3 November 2022), 2020.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Hoover, K., Hanson, L.A., 2022. "Wildfire Statistics,". IF10244, Aug. 1. Congressional Research Service, Washington, D.C., U.S. [Online]. Available: https://sgp.fas.org/crs/misc/IF10244.pdf CAL Fire. (Accessed 10 August 2022)

Huang, S., Tang, L., Hupy, J.P., et al., 2021. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. J. For. Res. 32, 1–6. https://doi.org/10.1007/s11676-020-01155-1.

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. Environ. Rev. 28 (4), 478–505. https://doi.org/10.1139/er-2020-0019.

Kelleher, S., Quinn, C., Miller-Lionberg, D., Volckens, J., 2018. A low-cost particulate matter (PM2.5) monitor for wildland fire smoke. Atmos. Meas. Tech. 11, 1087–1097. https://doi.org/10.5194/amt-11-1087-2018.

Kim, C.M., Kim, K.H., Lee, Y.S., Chung, K., Park, R.C., 2020. Real-time streaming image based PP2LFA-CRNN model for facial sentiment analysis. IEEE Access 8, 199586–199602. https://doi.org/10.1109/ACCESS.2020.3034319.

Kitzberger, T., Falk, D.A., Westerling, A.L., Swetnam, T.W., 2017. Direct and indirect climate controls predict heterogeneous early-mid 21st century wildfire burned area across western and boreal North America. PLoS One 12 (12), e0188486. https://doi.org/10.1371/journal.pone.0188486.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86 (11), 2278–2324. https://doi.org/10.1109/5.726791.

Liu, J.C., Mickley, L.J., Sulprizio, M.P., Dominici, F., Yue, X., Ebisu, K., Anderson, G.B., Khan, R.F.A., Ma, Bravo, Bell, M.L., 2016. Particulate air pollution from wildfires in the western US under climate change. Clim. Change 138 (3), 655–666. https://doi.org/10.1007/s10584-016-1762-6.

Mansoor, S., et al., 2022. Elevation in wildfire frequencies with respect to the climate change. J. Environ. Manag. 301, 113769 https://doi.org/10.1016/j.jenvman.2021.113769.

Marlon, J.R., et al., 2012. Long-term perspective on wildfires in the western USA. Proc. Natl. Acad. Sci. USA 109 (9). E535–E543.

Miller, R., 2020. Prescribed burns in California: a historical case study of the integration of scientific research and policy. Fire 3, 44. https://doi.org/10.3390/fire3030044.

Molina-Pico, A., Cuesta-Frau, D., Araujo, A., Alejandre, J., Rozas, A., 2016. Forest monitoring and wildland early fire detection by a hierarchical wireless sensor network. Sensors Env. Monitoring, 8325845. https://doi.org/10.1155/2016/8325845, 2016.

National Wildfire Coordinating Group. Guide to wildland fire origin and cause determination". PMS 412 NFES 1874 (2016). [Online]. Available: https://www.nwcg.gov/sites/default/files/publications/pms412.pdf.

Porter, T.W., Crowfoot, W., Newsom, G., 2020. Wildfire Activity Statistics. CALFIRE, Sacramento, California [Online]. Available: https://www.fire.ca.gov/media/0fdfj2h1/2020_redbook_final.pdf. (Accessed 10 August 2022).

PurpleAir. Sensor data download tool". [Online]. Available: https://www.purpleair.com/sensorlist.

Ramsey, M.L., Murphy, M., Diaz, J., 2020. The Camp Fire Public Report. Butte County District Attorney, Oroville, California [Online]. Available: http://www.buttecounty.net/Portals/30/CFReport/PGE-THE-CAMP-FIRE-PUBLIC-REPORT.pdf. (Accessed 10 August 2022).

Ronneberger, O., Fischer, P., Brox, T., U-Net, 2015. Convolutional networks for biomedical image segmentation. https://doi.org/10.48550/arXiv.1505.04597.

Sayad, Y.O., Mousannif, H., Hassan Al Moatassime, 2019. Predictive modeling of wildfires: a new dataset and machine learning approach. Fire Saf. J. 104, 130–146. https://doi.org/10.1016/j.firesaf.2019.01.006.

Shelhamer, E., Long, J., Darrell, T., 2016. Fully convolutional networks for semantic segmentation. https://doi.org/10.48550/arXiv.1605.06211.

Singla, S., Diao, T., Mukhopadhyay, A., Eldawy, A., Shachter, R., Kochenderfer, M., 2020. WildfireDB: a spatio-temporal dataset CombiningWildfire occurrence with relevant covariates. In: AI for Earth Sciences Workshop at NeurIPS. https://tdiao.github.io/publication/wildfireDB.

SW Taylor, DG Woolford, CB Dean and DL Martell. "Wildfire Prediction to Inform Fire Management: Statistical Science Challenges".

Ullrich, P.A., Xu, Z., Rhoades, A.M., Dettinger, M.D., Mount, J.F., Jones, A.D., Vahmani, P., 2018. California's drought of the future: a midcentury recreation of the exceptional conditions of 2012–2017. Earth's Future 6, 1568–1587. https://doi.org/10.1029/2018EF001007.

Varela, N., Díaz-Martinez, J.L., Ospino, A., Zelaya, N.A.L., 2020. Wireless sensor network for forest fire detection. Proc. Comput. Sci. 175, 435–440. https://doi.org/10.1016/j.procs.2020.07.061.

Wang, D., Guan, D., Zhu, S., et al., 2021. Economic footprint of California wildfires in 2018. Nat. Sustain. 4, 252–260. https://doi.org/10.1038/s41893-020-00646-7.

Williams, JT.Abatzoglou & AP., 2016. Impact of anthropogenic climate change on wildfire across western US forests. Proc. Natl. Acad. Sci. U.S.A. 113, 11770–11775.

Zhang, Z., Sabuncu, M., 2018. Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels," 32nd Conference on Neural Information Processing Systems. NeurIPS, Montréal, Canada [Online]. Available: https://proceedings.neurips.cc/paper/2018/file/f2925f97bc13ad2852a7a551802feea0-Paper.pdf.