

Next Day Wildfire Spread: A Machine Learning Dataset to Predict Wildfire Spreading From Remote-Sensing Data

Fantine Huot¹, R. Lily Hu, Nita Goyal, Tharun Sankar², Matthias Ihme, and Yi-Fan Chen

Abstract—Predicting wildfire spread is critical for land management and disaster preparedness. To this end, we present “Next Day Wildfire Spread,” a curated, large-scale, multivariate dataset of historical wildfires aggregating nearly a decade of remote-sensing data across the United States. In contrast to existing fire datasets based on Earth observation satellites, our dataset combines 2-D fire data with multiple explanatory variables (e.g., topography, vegetation, weather, drought index, and population density) aligned over 2-D regions, providing a feature-rich dataset for machine learning. To demonstrate the usefulness of this dataset, we implement a neural network that takes advantage of the spatial information of these data to predict wildfire spread. We compare the performance of the neural network with other machine learning models: logistic regression and random forest. This dataset can be used as a benchmark for developing wildfire propagation models based on remote-sensing data for a lead time of one day.

Index Terms—Earth Engine, machine learning, remote sensing, wildfire.

I. INTRODUCTION

WILDFIRES can threaten livelihood and properties, and impact environment and health [1], [2]. Over the last few decades, wildfire management has changed profoundly, facing longer fire seasons and more severe fires with more acres burned on average each year [3]. In 2019, more than 775 000 residences across the United States were flagged as at “extreme” risk of destructive wildfire, amounting to an estimated reconstruction cost of 220 billion dollars [4]. Wildfires significantly impact the climate and are estimated to contribute to 10% of the CO₂ emissions per year worldwide [1]. Furthermore, the health impact due to wildfire aerosols is estimated at 300 000 premature deaths per year [2].

There is a need for novel wildfire warning and prediction technologies that enable better fire management, mitigation, and evacuation decisions. In particular, evaluating the fire likelihood [5]—the probability of wildfire burning in a specific location—would provide valuable land management and disaster preparedness capabilities. Predictions of where a fire will spread in the upcoming day are essential to wildfire emergency

response, leading to optimal allocation of resources and quick response to fire activity [6], [7]. In addition, identifying regions of high wildfire likelihood offers the possibility for targeted fire-prevention measures, such as shutting down power lines to minimize electrical fire hazards [8], [9], or compartmentalizing forests by creating vegetation fuel breaks [10], [11].

The increase in the availability of remote-sensing data, computational resources, and advances in machine learning provides unprecedented opportunities for data-driven approaches to estimate wildfire likelihood. Herein, we investigate the potential of deep learning models to predict wildfire spreading from observational data [12]–[15]. For this purpose, we present a new dataset: “Next Day Wildfire Spread.”

This “Next Day Wildfire Spread” dataset is a curated, large-scale, multivariate dataset of historical wildfires over nearly a decade across the United States, aggregated using Google Earth Engine (GEE) [16]. This dataset takes advantage of the increasing availability and capability of remote-sensing technologies, resulting in extensive spatial and temporal coverage from a wide range of sensors (e.g., the Moderate Resolution Imaging Spectroradiometer (MODIS) [17], the Visible Infrared Imaging Radiometer Suite (VIIRS) [18], and the Shuttle Radar Topography Mission (SRTM) [19]). Our dataset combines historical wildfires with 11 observational variables overlaid over 2-D regions at 1 km resolution: elevation, wind direction, wind speed, minimum temperature, maximum temperature, humidity, precipitation, drought index, vegetation, population density, and energy release component (ERC). The resulting dataset has 18 545 fire events, for which we provide two snapshots of the fire spreading pattern, at time t and $t + 1$ day.

Among existing fire datasets based on Earth observation satellites [20]–[22], most do not include this many variables at 1 km resolution. Fire datasets, such as FRY [20], Fire Atlas [21], and the 1.88 million US Wildfires catalog [23], focus on the total burn area and do not provide the temporal resolution or the 2-D information required for characterizing fire spreading patterns. In contrast, GlobFire [22] is designed for characterizing fires and includes daily timestamps of the fire perimeter. However, none of these datasets includes other variables, such as vegetation, weather, drought, or topography data required for fire prediction. It is true that previous studies have combined fire data with other variables [24]–[29], but these datasets have not been publicly released or only partially so. A dataset of fire events in Canada [30] has been released

Manuscript received 4 December 2021; revised 2 March 2022; accepted 2 July 2022. Date of publication 26 July 2022; date of current version 4 August 2022. This work was supported by the Stanford Exploration Project Affiliate Members. (Corresponding author: Fantine Huot.)

The authors are with the Research Department, Google Inc., Mountain View, CA 94043 USA (e-mail: fantinehuot@google.com).

Digital Object Identifier 10.1109/TGRS.2022.3192974

but is limited in geographical coverage and the number of recorded fire events (386 events). Table I compiles a summary of these publicly available wildfire datasets.

Benchmark datasets have played an essential role in driving progress and innovation in machine learning research [31], [32]. They enable performance comparisons across models, which, in turn, can lead to better predictive models. We make the “Next Day Wildfire Spread” dataset available to the broader scientific community for benchmarks, model comparisons, and further insights for wildfire predictions.¹ We share our GEE data aggregation code for users to adapt to their use cases.²

In recent years, deep learning approaches have been increasingly adopted for prediction tasks from remote-sensing data, as reviewed in [15]. For instance, Alonso-Betanzos *et al.* [33] used a deep learning model to estimate the daily fire likelihood from temperature, humidity, and rainfall data from five weather stations in the Galicia region of Spain. On Lebanon data, Sakr *et al.* [34] implemented a neural network for predicting the occurrence of wildfires based only on monthly relative humidity and cumulative precipitation data. Compared to these two datasets, the “Next Day Wildfire Spread” dataset has more data samples and a much greater geographical coverage. In an Australian study, Dutta *et al.* [35] use neural networks on aggregated monthly meteorological data (evaporation, precipitation, incoming solar irradiance, maximum temperature, soil moisture, wind speed, pressure, and humidity) to estimate the risk of fire occurrence. However, with monthly timestamps, their dataset does not have the temporal resolution for estimating fire spreading.

For fire spread prediction, Hodges and Lattimer [36] use a convolutional neural network to predict spreading patterns from environmental variables, such as topography and weather, but only use synthetic data from computational models. In [37], a similar method is used to predict fire spread but from remote-sensing data collected from geographic information systems (GISs), combining topography and weather (pressure, temperature, dew point, wind direction, wind speed, precipitation, and relative humidity). The objectives of this study are similar to the ones in our work but differ in the methodology and data aggregation. The dataset was limited to regions in Colorado and constrained to the data available in the Geospatial Multi-Agency Coordination (GeoMAC) system [38]. Moreover, the datasets from studies [33]–[37] were not publicly released.

In contrast, our “Next Day Wildfire Spread” is an extensive publicly available dataset for ML fire spread prediction. Aggregating large datasets with many observational variables is challenging because it is usually limited by the data availability across multiple data sources. We tackle this challenge by using the GEE framework, which allows capturing a large ensemble of fire events and more observational variables than any of the aforementioned datasets, all at 1 km resolution instead of pointwise locations. In addition, it is the only dataset

of this type to have a variable to represent anthropogenic activity. Moreover, its spatial dimension makes it suitable for computer vision and segmentation tasks.

By releasing the dataset with its interactive GEE aggregation code, we offer an end-to-end framework for taking advantage of GEE capabilities for developing ML datasets. We include the full preprocessing and data interface to make it as accessible as possible to ML practitioners.

To illustrate the usefulness of our “Next Day Wildfire Spread” dataset, we train a deep learning model for fire spread prediction. Given a wildfire, we predict where the fire will spread the following day. We implement a convolutional autoencoder, a specialized type of neural network, to take advantage of the 2-D information of this dataset and compare its performance with two other ML models, namely, logistic regression and random forest.

The remainder of this manuscript has the following structure. Section II describes our data aggregation workflow for historical remote-sensing data. In Section III, we discuss our processing pipeline for the fire spread predictions. Section IV presents the deep learning model, and results are presented in Section V. This article finishes with conclusions in Section VI.

II. DATASET AGGREGATION

With the continuing improvements in quality, resolution, and coverage of remote-sensing technologies, we now have access to increasingly more data for characterizing wildfires. Using data sources available in GEE [16], we present a data aggregation workflow for creating the “Next Day Wildfire Spread” dataset, combining historical fire events with remote-sensing data.

A. Motivation

We collect the data at different locations and times at which wildfires occurred. We extract the data as 64 km × 64 km regions at 1 km resolution to capture all typical active fire sizes [39]–[41]. We process the data from GEE to represent the fire information as a fire mask over each region, showing the locations of “fire” versus “no fire,” with an additional class for uncertain labels (i.e., cloud coverage or other unprocessed data). We include both the fire mask at time t and at time $t+1$ day to provide two snapshots of the fire spreading pattern.

Using GEE, we aggregate data from different data sources and align the data in location and time using the same projection (WGS84). With this methodology, we combine these fire masks with variables that are of direct relevance to wildfire predictions [28], [35], [42], [43]: elevation, wind direction and wind speed, minimum and maximum temperatures, humidity, precipitation, drought index, normalized difference vegetation index (NDVI), and ERC. In particular, NDVI, drought index, and the weather variables provide information relevant to fuel properties. ERC is a calculated output of the National Fire Danger Rating System (NFDRS) [44]. It is considered a composite fuel moisture index as it reflects the contribution of all fuels to potential fire intensity. In addition, we include population density to correlate the wildfires to anthropogenic

¹<https://www.kaggle.com/fantineh/next-day-wildfire-spread>

²https://github.com/google-research/google-research/tree/master/simulation_research/next_day_wildfire_spread

TABLE I
PUBLICLY AVAILABLE FIRE DATASETS

Name	Fire information	Coverage	Period	Spatial resolution	Fire spreading	Other variables
FRY [20]	Total burn area	Worldwide	2005 - 2011	500 m	N/A *	-
Fire Atlas [21]	Total burn area	Worldwide	2003 - 2016	500 m	N/A *	-
GlobFire [22]	Active fire	Worldwide	2001 - 2017	500 m	Daily fire maps	-
1.88 million US Wildfires [23]	Active fire	USA	1992 - 2015	Point coordinates	N/A	-
Fire events in Canada [30]	Active fire	Regions in Canada	August 2014	Point coordinates	N/A	Vegetation, Surface Temperature
Next Day Wildfire Spread	Active fire	USA	2012 - 2020	1 km	t and $t + 1$ day fire maps	Elevation, Wind Direction and Speed, Min. and Max. Temperatures, Humidity, Precipitation, Drought Index, Vegetation, Population density, ERC

*The total burn area by itself does not provide information about the fire spreading pattern.

ignitions and for risk-severity assessment. Humans cause 84% of fires [45], so we use population density as a proxy for anthropogenic activity. Examples from this dataset are illustrated in Fig. 1.

In collecting this dataset, we prioritized data from publicly available data sources that are regularly updated with recent data. This enables the methodologies developed on this dataset to translate to future data predictions, including near real-time prediction. The use of openly available data with generous terms of use allows it to be incorporated into dynamic wildfire prevention and management tools for future applications with minimum licensing constraints.

This feature-rich dataset has multiple applications. It can be used for statistical analysis or to study correlations between various variables. To demonstrate this capability, we use this dataset for predicting wildfire propagation. The dataset can be resampled to different region sizes, spatial resolutions, or any subset of the provided data. In addition to the data itself, we also release the GEE export code that generates the dataset, giving users the flexibility of adapting it to their use cases.² The export code can be changed to incorporate additional data sources, in a different geographic region or worldwide, over different time scales and intervals (hourly, daily, or weekly). For instance, Huot *et al.* [46] present examples of the same dataset extracted as weekly time sequences to create maps of the total burn area.

B. Data Sources From Earth Engine

We compile this dataset from multiple remote-sensing data sources from GEE [16]. We selected data sources with extensive geographical and historical coverage, and reduced missing data.

- 1) **Historical wildfire** data are from the MOD14A1 V6 dataset [17], a collection of daily fire mask composites at 1 km resolution since 2000, provided by NASA LP DAAC at the USGS EROS Center.
- 2) **Topography** data are from the SRTM [19], sampled at 30 m resolution.
- 3) **USA Weather** data are from the University of Idaho Gridded Surface Meteorological Dataset (GRIDMET) [47], a collection of daily surface fields of temperature, precipitation, winds, and humidity at 4 km resolution since 1979, provided by the University of California at Merced.
- 4) **USA Drought** data are from GRIDMET Drought, a collection of drought indices derived from the GRIDMET dataset [48], sampled at 4 km resolution every five days since 1979, provided by the University of California at Merced.
- 5) **Vegetation** data are from the Suomi National Polar-orbiting Partnership (S-NPP) NASA VIIRS Vegetation Indices (VNP13A1) dataset [18], a collection of vegetation indices sampled at 0.5 km resolution every eight days since 2012, provided by NASA LP DAAC at the USGS EROS Center.
- 6) **Population density** data are from the Gridded Population of World Version 4 (GPWv4) dataset by the Center

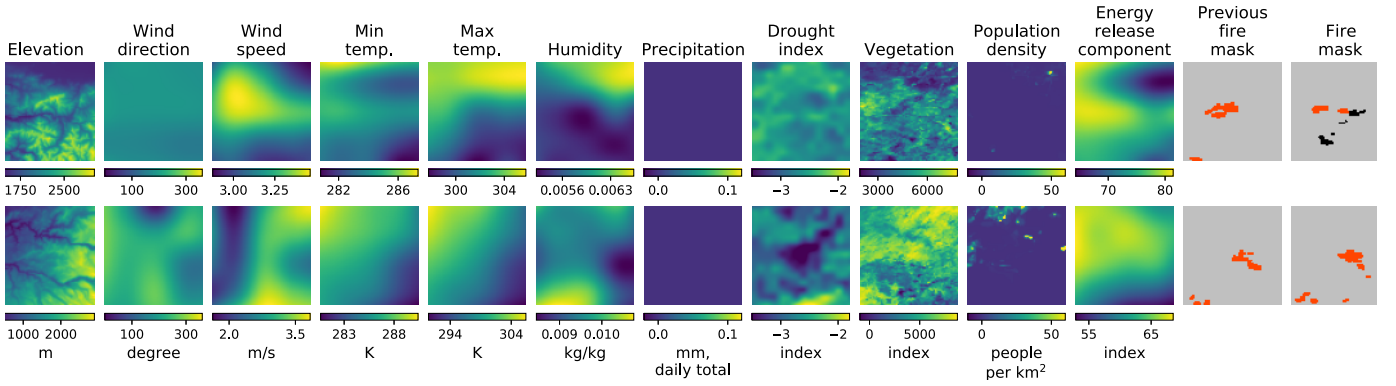


Fig. 1. Examples from our “Next Day Wildfire Spread” dataset. Each example is $64 \text{ km} \times 64 \text{ km}$ at 1 km resolution. Each row corresponds to a location and time t at which a fire occurred, and the columns represent the different variables. In each row, all the weather, drought, vegetation, and population density variables are extracted at that same location at time t . “temp” denotes temperature; “wind direction” is the azimuth in degrees clockwise from north. The “previous fire mask” corresponds to the fire locations at time t , while “fire mask” corresponds to the following day at $t + 1$ day. In the fire masks, red corresponds to fire, and while gray corresponds to no fire. Black indicates uncertain labels (i.e., cloud coverage or other unprocessed data).

for International Earth Science Information Network (CIESIN) [49]. These data contain the distribution of the global human population, sampled every five years at 1 km resolution.

All the variables in our dataset are available from the data sources cited; none were calculated by us.

While we selected these data sources for creating our dataset, users can change the data sources based on their needs using the provided GEE data export code. For instance, other fire mask datasets include the Fire Information for Resource Management System (FIRMS) [50], or GOES-16 and GOES-17 [51], [52]. Additional variables can be extracted and combined from the data from the MODIS Aqua and Terra satellites [53]. As such, the data export code can be adapted to encompass a variety of wildfire problems beyond fire spread prediction, such as identifying fire precursors, probability of ignition, and long-term fire risk patterns.

C. Data Aggregation

From these data sources, we sample examples at all locations and times with active fires. We consider fires separated by more than 10 km as belonging to a different fire. We extract both the fire mask at time t , which we denote “previous fire mask,” and at time $t + 1$ day, which we denote “fire mask.” For characterizing fire spreading, we only keep samples for which there was at least one area on fire within the region at time t , in other words, for which the “previous fire mask” contains any fire at all.

We extract the data over the contiguous United States from 2012 to 2020. We select this time period due to data availability. We split the data between training, evaluating, and testing by randomly separating all the weeks between 2012 and 2020 according to an 8:1:1 ratio, respectively, while keeping a one-day buffer between weeks from which we do not sample data.

Since the data sources have different spatial resolutions, we align all the data to a 1 km resolution, which corresponds to the spatial resolution of the fire masks. We downsample the topography and vegetation data and use bicubic interpolation

for the weather and drought data. The data sources also have different temporal resolutions and are refreshed over different time intervals. For variables that vary slowly over time, we take the last available time stamp at time t . For the weather data, refreshed four times a day, we take the average of each weather variable over the day corresponding to time t .

Using these data sources, we see in Fig. 1 that variables such as elevation, vegetation, and population density show a lot of variation across a $64 \text{ km} \times 64 \text{ km}$ region. Physical quantities, such as temperatures and precipitation, are smoother across the region. The two last columns display the fire locations at times t and $t + 1$ day, where “no fire” is indicated in gray, “fire” in red, and missing data in black.

The resulting dataset contains 18545 samples. In 58% of these samples (10798 samples), the fire increases in size from t to $t + 1$ day. In 39% (7191 samples), the fire decreases in size. In the remaining samples, the fire stays the same size.

III. DATA PREPROCESSING

For the machine learning task, we treat the variables and the “previous fire mask” at time t as data features and the “fire mask” at time $t + 1$ day as labels. The values of each data feature, except for the fire masks, are first clipped between a minimum and a maximum clip value, with different clipping values for each feature. We then normalize each feature separately by subtracting the mean and dividing by the standard deviation.

We clip the data because analysis of the features revealed the presence of extreme values, some of which were not even physically reasonable. Moreover, physical variables that span an extensive dynamic range can lead to vanishing or exploding gradients in the deep learning training process. The clipping values were either based on physical knowledge (e.g., percentages between 0% and 100%) or set to the 0.1th and 99.9th percentiles for each feature. Statistics for processing the input data were calculated over the training dataset. Means and standard deviations were calculated after clipping. This allows us to process the data for inference without knowing whether there is currently a fire.

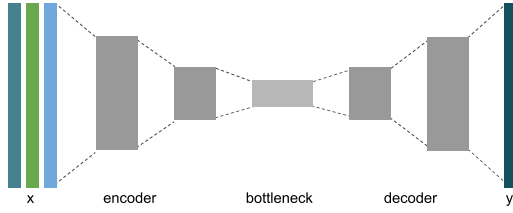


Fig. 2. Convolutional autoencoder for estimating fire spreading

In our dataset, the fire events are generally centered within each $64 \text{ km} \times 64 \text{ km}$ region. Therefore, we perform data augmentation in the machine learning input pipeline to offer data examples of fires occurring at different locations. We do so by randomly cropping $32 \text{ km} \times 32 \text{ km}$ regions from the original $64 \text{ km} \times 64 \text{ km}$ regions.

IV. MACHINE LEARNING APPLICATION

We illustrate the usefulness of the “Next Day Wildfire Spread” dataset by training a deep learning model for fire spread prediction.

A. Models for Segmentation

We frame the ML task as an image segmentation problem where we classify each area as either containing fire or no fire given the location of the fire on the previous day and the data features described in Section II.

We implement a deep learning model that takes advantage of the spatial information in the input data to perform this task. We use a convolutional autoencoder, a specialized type of neural network for precise image segmentation (see Fig. 2). This model was selected because the remote-sensing data can be treated as a multichannel input image and the fire mask as a segmentation map.

In addition to the deep learning model, we also train two machine learning models for comparison: a logistic regression model and a random forest. These non-deep learning models do not take advantage of the 2-D information the way the convolutional autoencoder does. Therefore, with these models, the image segmentation is performed areawise (pixelwise). In this setting, the target label for each training sample is a $1 \text{ km} \times 1 \text{ km}$ area (pixel). The label is the presence of fire in that area at time $t + 1$ day. The input variables are all the features for that area at time t and the eight neighboring areas (pixels) around it.

Additional details on model implementation are provided in Appendix A.

B. Training Details and Hyperparameter Tuning

Since wildfires represent only a small area within a region, we use a weighted cross-entropy loss [54] and explore a range of different weights on the “fire” labels to take into account the class imbalance. Uncertain labels are ignored in the loss and performance calculations.

For the deep learning model, we include data augmentation by random crop, random flip, and random rotation. We train

TABLE II
WILDFIRE SPREADING PREDICTION METRICS

	AUC (PR)	Precision	Recall
Neural Network	28.4	33.6	43.1
Random Forest	22.5	26.3	46.9
Logistic Regression	19.8	32.5	35.3
Persistence	11.5	35.7	27.3

it over 1000 epochs, with 1000 iteration steps per epoch, with an Adam optimizer, on four V100 GPUs. We perform the hyperparameter selection by grid search. The resulting network architecture and hyperparameter tuning details are provided in Appendix A-A.

The training and hyperparameter details of the logistic regression model and the random forest are described in Appendix A-B.

The best model is selected as the one with the best area under the precision–recall curve (AUC PR). We use AUC PR instead of accuracy because AUC PR is a more effective diagnostic metric for imbalanced binary classification [55], and our dataset contains more “no fire” than “fire” samples.

V. RESULTS

We compare the predictions on where the fire will be at time $t + 1$ day given data for a region with an ongoing fire at time t .

A. Prediction Results

Metrics on fire spread prediction per area ($1 \text{ km} \times 1 \text{ km}$) are shown in Table II. All trained models achieve a higher AUC than the persistence baseline, which has an AUC of 11.5%. The model with the highest AUC is the neural network at 28.4%, followed by the random forest and then logistic regression. The precision and recall for the neural network on the positive class are 33.6% and 43.1%, respectively. The AUCs of the logistic regression and random forest models—the non-deep learning models—are within 3% of one another and at least 6% lower than the neural network. The logistic regression baseline achieves nearly the same precision as the neural network at 32.5% but has a lower recall at 35.3%. The random forest baseline has a higher recall than the neural network at 46.9% but has a lower precision at 26.3%.

While the metrics on the positive class seem low, visualizations of some samples with predictions and targets in Fig. 3 show that fires are predicted. The predicted fires are roughly in the target location and are often rounder with smoother borders than the target. Predictions may connect fires that are close together into a single fire. There is a strong dependence on the previous fire mask, with the model predicting that a previous small fire grows, while, in the target, the fire was no longer present. The segmentation errors result from missing small fires and misclassified pixels at the boundary between fire and nonfire and between nearby fire areas.

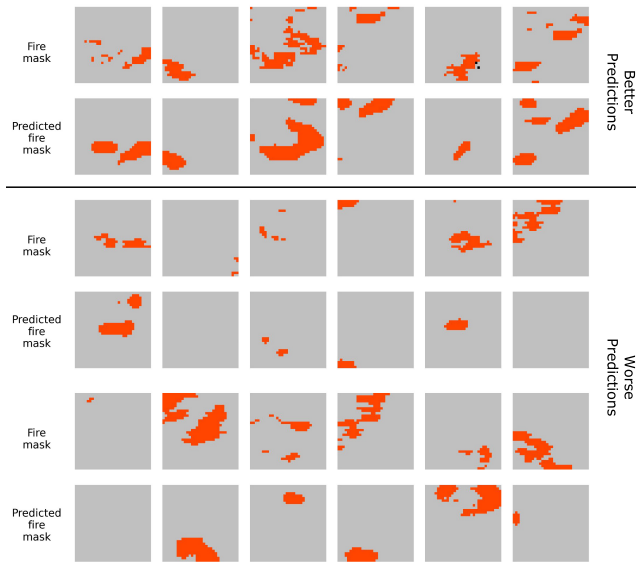


Fig. 3. Examples of target labels for next day fire spread and predictions from the deep learning model. Red corresponds to “fire,” gray corresponds to “no fire,” and black indicates missing data.

To illustrate how to use the ML workflow in practice, we evaluate predictions from the deep learning model on a set of six known historical fires. Using our GEE code, we aggregate the input variables at the start of each event and predict the next day’s fire spreading. Examples of these known fires are shown in Fig. 4. The model achieves an AUC of 38.7% on these historical fires, higher than that of the test set. The precision of 33.4% and recall of 37.2% are similar to the test set metrics. The visualized results are similar to other examples in the dataset, with good recognition of larger fires (such as the Grizzly Creek fire). While the exact outline of the fires is not accurate, the model tends to capture the extent of the fire spreading well (for instance, predicting a large fire for the LNU Lightning Complex). The limitations are similar to those observed on the test set: the model tends to predict smoother boundaries than the target (such as the Bighorn fire), can miss smaller fires entirely (such as the Evans Canyon fire), and can merge multiple separated segments of the fire (such as the LNU Lightning Complex).

Noticing that it is easier to predict the presence or the absence of fire over a larger area, we experiment with predicting fire spread at a coarser, lower resolution than the input data. This is to explore and quantify the tradeoff of prediction performance with resolution, that is, predictions for a $2 \text{ km} \times 2 \text{ km}$ (2×2 pixel) area instead of a $1 \text{ km} \times 1 \text{ km}$ (1×1 pixel) area. This turns the problem framing from pixel classification and segmentation into classification over a larger and larger area, as shown in Fig. 5. Each row illustrates the target and prediction for fire spread at time $t + 1$ day at progressively lower resolution. An area (pixel) within the region is labeled as fire if there is any fire within that area. Similar to previous experiments, the model predicts fire over a larger area than the target. The predictions connect several small areas with fire into a larger fire. The boundaries of

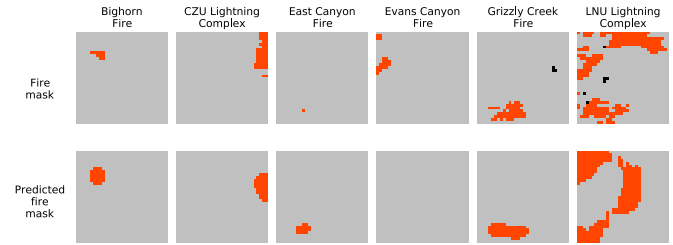


Fig. 4. Examples of historical fires with the target labels for next day fire spread and predictions from the deep learning model. The included fires are the Bighorn fire (Santa Catalina Mountains, AZ, USA; June 7, 2020), CZU lightning complex (Santa Cruz Mountains, CA, USA; August 19, 2020), East Canyon fire (Montezuma County, CO, USA; June 18, 2020), Evans Canyon fire (Yakima, WA, USA; September 2, 2020), Grizzly Creek fire (Glenwood Canyon, CO, USA; August 12, 2020), and LNU lightning complex (Sonoma, CA, USA; August 19, 2020). Red corresponds to “fire,” gray corresponds to “no fire,” and black indicates missing data.



Fig. 5. Examples of target labels for next day fire spread and predictions from a deep learning model, where the resolution of the output is lower than the resolution of the input, with the factor indicated in parenthesis. Each column corresponds to a single sample at different target resolutions. Red corresponds to fire, gray corresponds to no fire, and black indicates missing data.

the predictions are also smoother than the target. The model sometimes misses lone small fires.

The best precision/recall tradeoff for different resolutions is when the output corresponds to an area that is eight times larger than the input area. This results in an AUC of 66.3%. While the prediction metrics improve as the prediction region size increases, the predictions become less useful from an operational perspective due to less localization of the fire. When predicting at lower spatial resolutions, smaller fires are

TABLE III
LOWER RESOLUTION PREDICTIONS

Label	Lower Resolution	AUC (PR)	Precision	Recall
Decimal	2x	38.8	48.4	40.8
	4x	52.3	69.1	39.8
	8x	64.5	88.7	24.7
Binary	2x	39.8	40.3	51.9
	4x	53.5	48.8	61.0
	8x	66.3	62.1	63.1
Persistence	2x	19.5	45.6	36.8
	4x	30.6	56.0	46.8
	8x	41.3	64.0	52.5

missed due to the coarser resolution. We experiment with two methods of combining labels for a larger region: a decimal representing the percent of pixels that are on fire, and a binary value that represents any pixel within that region being on fire. For each of these types of labels, experiments are conducted at different resolutions, as shown in Table III. The neural network performs with higher AUC than the persistence baseline, with most of the gains due to achieving higher precision than the persistence baseline.

B. Feature Analysis

To analyze the effect of different input features on prediction results, we conduct ablation studies using the model with the highest AUC—the deep learning model. The feature analyses on the random forest and the logistic regression model are provided in Appendices A-C and A-D. We do not explicitly program the deep learning model to weigh certain variables more than others. Hence, ablation experiments provide insights into which variables provide the most significant contribution to the model’s predictive capabilities.

First, for each of the input features, we remove that input feature and retrain the remaining ones. The metrics of the resulting models are summarized in Table IV. Removing the current fire location results in the most significant decrease in AUC compared to using all the input features. This result makes intuitive sense and illustrates the importance of the current fire location as a predictor variable. Among the other input features, removing any one of them results in a similar overall performance to using all the input features, which seems reasonable because many of these features are correlated. Removing some of these input features even resulted in slightly better performance. This increase could be due to randomness, as we have observed that our experiments have a 0.5% variation in AUC due to random cropping. In addition, reducing the number of features makes the model simpler, making it less prone to overfitting and reducing the noisiness and redundancies of all the features.

Due to the previous fire mask overwhelmingly contributing to the prediction performance, we also retrain the model

TABLE IV
FEATURE ABLATION: REMOVE ONE FEATURE

Removed feature	AUC (PR)	Precision	Recall
Previous fire mask	6.8	4.6	0.2
Humidity	26.1	35.4	35.7
Max temperature	26.8	35.4	36.6
Wind speed	27.2	33.1	42.4
Vegetation	27.7	35.0	40.0
Precipitation	28.0	32.5	42.0
Population	28.1	32.8	43.0
Wind direction	28.3	32.8	43.1
Min temperature	28.6	35.2	40.9
Elevation	28.8	33.6	43.0
Drought	28.8	36.8	39.0
ERC	28.8	33.0	43.2

TABLE V
FEATURE ABLATION: KEEP PREVIOUS FIRE MASK
AND ONE OTHER FEATURE

Kept feature	AUC (PR)	Precision	Recall
Vegetation	28.2	33.0	41.4
Elevation	27.0	31.3	44.4
Max temperature	26.7	32.7	43.5
Population	26.6	32.6	43.2
Precipitation	26.6	34.1	40.6
Min temperature	26.2	29.9	47.2
Drought	26.0	30.8	44.1
Wind speed	25.9	30.3	45.6
ERC	25.8	30.7	45.0
Humidity	25.7	32.4	41.7
Wind direction	25.5	30.9	43.5

keeping only two features, the previous fire mask, and each of the other features (see Table V). We see that vegetation and elevation result in the best performance. The relative ordering of the features changed between leaving a feature out compared to keeping only that feature along with the previous fire. Because the input features are correlated, we also retrain with only a single feature. Unfortunately, using a single feature resulted in a recall of zero, for all features except for the previous fire mask. Using only the previous fire mask as input results in a recall of 49.6%, which is higher than the recall of using all features. However, the precision and AUC when only using the previous fire mask are lower, at 29.2% and 22.9% respectively.

VI. CONCLUSION

We present this “Next Day Wildfire Spread” dataset as an open data resource for further research to advance our

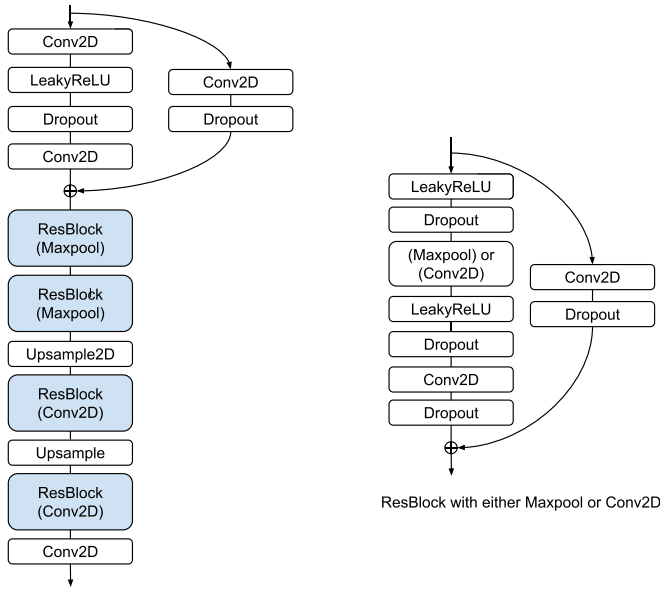


Fig. 6. (Left) Deep learning model architecture for image segmentation. The architecture uses a residual block (ResBlock) module with the variation of the ResBlock indicated in parenthesis. (Right) ResBlock architecture.

collective ability to anticipate and respond to wildfires. Open datasets allow benchmarking and ML model comparisons, making them essential to producing high-quality ML models for real-world fire scenarios. We create this dataset by aggregating nearly a decade of remote-sensing data, combining features including topography, weather, drought index, vegetation, and population density with historical fire records. Using this dataset, we demonstrate the potential of deep learning approaches to predict wildfires from remote-sensing data and illustrate the performance gaps on the example task of day-ahead fire spread prediction. Once the ML model is trained, we apply it to examples of historical fires, illustrating how it could be used for anticipating the extent of fires the following day. This information could be valuable for allocating resources for fire suppression efforts.

However, this experiment has some data limitations that must be addressed before incorporating such data-driven approaches into wildfire warning and prediction technologies. The 1-km spatial and daily temporal resolution of the MODIS data limits the ML model's prediction resolution. As such, it does not provide the fine-scale information required for tactical decision-making. As for the GRIDMET weather data, their spatial and temporal resolutions are insufficient to capture the local wind patterns that drive fire spreading.

Since we use the MODIS data for labeling the active fires, our dataset comes with the same caveats as the MODIS data. For instance, MODIS does not sample the late afternoon when conditions are most favorable for fire intensification. In addition, it does not separate wildfire events from other prescribed fires. To the extent that prescribed fires correlate with land use, we include population density as a proxy for anthropogenic activity. Moreover, since we focus on the fire spreading pattern, our data aggregation does not include the small fires that

TABLE VI
RANDOM FOREST: GINI IMPORTANCE

Feature	Gini Importance
Previous fire mask (5)	0.092
Previous fire mask (6)	0.074
Previous fire mask (4)	0.073
Previous fire mask (2)	0.071
Previous fire mask (1)	0.068
Previous fire mask (8)	0.067
Previous fire mask (9)	0.061
Previous fire mask (3)	0.052
Previous fire mask (7)	0.051
ERC (7)	0.017
ERC (8)	0.014
ERC (5)	0.014
ERC (6)	0.012
ERC (9)	0.01
ERC (2)	0.008
ERC (1)	0.008
ERC (4)	0.008
ERC (3)	0.007
Elevation (7)	0.007
Elevation (9)	0.005

occur only on one day, excluding some of the noisy labels from MODIS. Still, we do not explicitly separate wildfires from prescribed fires. The MODIS fire detection algorithm is also conservative, especially at night or when there is cloud coverage. We include an uncertain label that the ML model ignores to account for these pixels, but artificial patterns in fire-spreading labels can still occur. Last but not least, the MODIS labels are affected by fire suppression efforts. While including the population density map can account for this effect to some degree, this is bound to influence the ML model's predictive capabilities.

These limitations highlight some of the challenges of wildfire prediction from remote-sensing data. Therefore, we release this dataset with its GEE data aggregation code, making it more than just a static dataset but an end-to-end framework for aggregating ML datasets from GEE. We designed this code in a modular fashion to easily change the time, the data sources, the geographical area, or the temporal and spatial sampling. As more curated data sources become available on the GEE platform, our framework can be used to improve the dataset. Future datasets could be expanded to a global scale or cover a more extensive time period. Data from recent years, with more fires and larger burned areas, could be weighted more heavily than data from a decade ago. We could also complement the study with synthetically generated data from high-fidelity

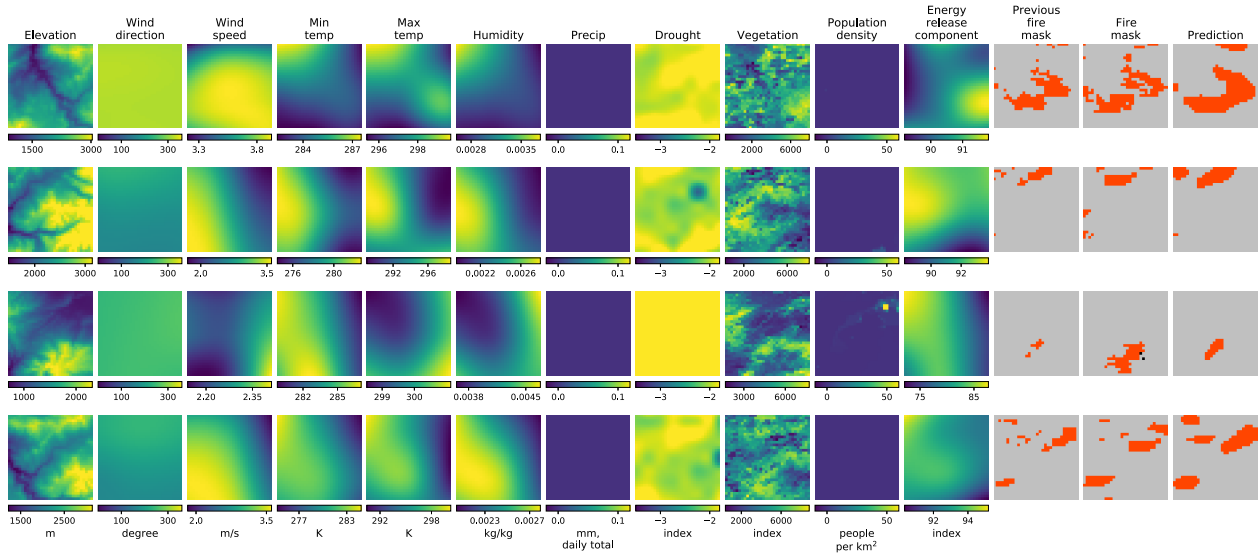


Fig. 7. Examples of better segmentation predictions. Each row is a sample.

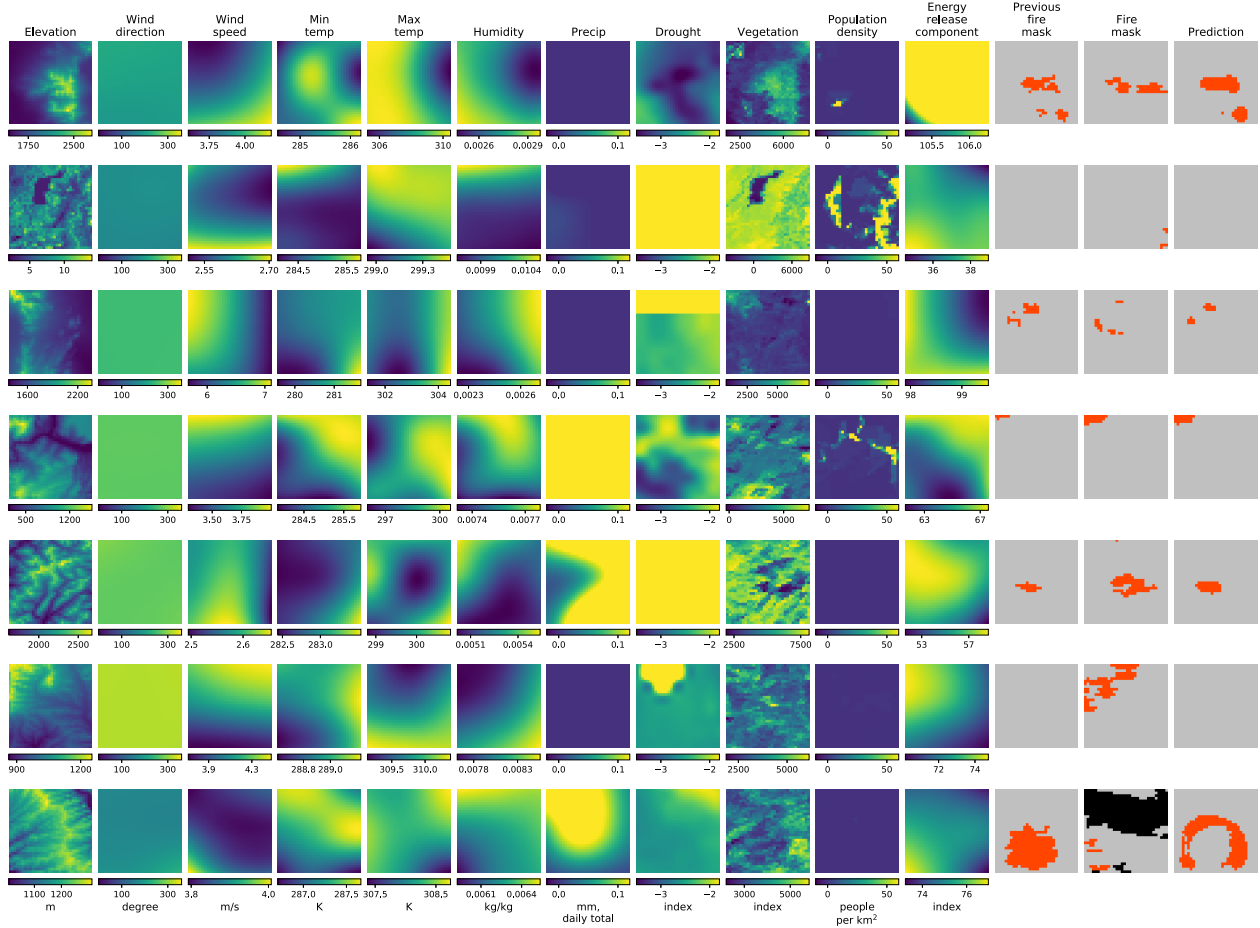


Fig. 8. Examples of worse segmentation predictions. Each row is a sample.

fire simulations. Beyond wildfires, the described workflow and methodology could be expanded to other problems, such as estimating the likelihood of regions to droughts, hurricanes, and other phenomena from historical remote-sensing data.

APPENDIX A MACHINE LEARNING MODEL DETAILS

A. Deep Learning Model

The architecture of the deep learning model used for image segmentation is shown in Fig. 6. All convolutions are 3×3

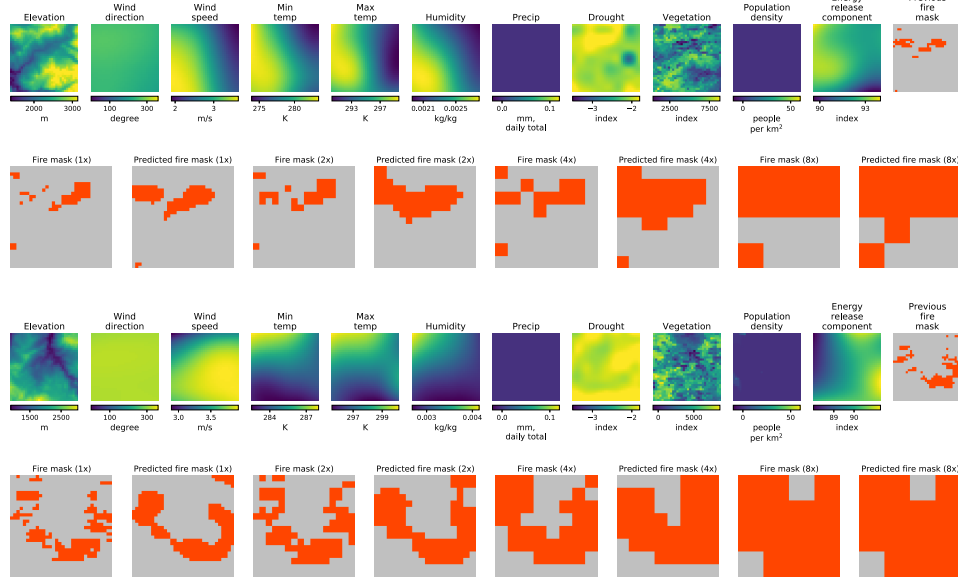


Fig. 9. Examples of better coarse segmentation predictions. Each pair of rows is a sample.

TABLE VII
LOGISTIC REGRESSION: COEFFICIENTS

Feature	Coefficient	Standard Deviation	Units
Min temp. (3)	-7.2	8.98	K
Max temp. (3)	6.01	9.82	K
ERC (1)	-5.69	20.85	index
Max temp. (5)	-5.49	9.82	K
ERC (5)	4.97	20.85	index
Min temp. (5)	4.77	8.98	K
Humidity (1)	-3.64	0.0	kg/kg
Max temp. (8)	-3.5	9.82	K
Min temp. (8)	3.24	8.98	K
Min temp. (4)	-3.16	8.98	K
Max temp. (1)	2.87	9.82	K
ERC (7)	2.71	20.85	index
Precipitation (1)	-2.44	4.48	mm
Precipitation (5)	2.43	4.48	mm
Max temp. (2)	-2.31	9.82	K
Precipitation (9)	-2.09	4.48	mm
Min temp. (2)	1.96	8.98	K
Min temp. (1)	1.87	8.98	K
ERC (4)	-1.81	20.85	index
ERC (3)	1.73	20.85	index

TABLE VIII
LOGISTIC REGRESSION: REMOVE ONE FEATURE

Removed feature	AUC (PR)	Precision	Recall
Previous fire mask	3.1	0.0	0.0
ERC	19.6	32.4	35.5
Elevation	19.6	32.3	35.5
Drought	19.6	32.4	35.4
Min temperature	19.6	32.2	35.5
Wind direction	19.7	32.3	35.7
Max temperature	19.7	32.3	35.5
Vegetation	19.8	32.3	35.6
Precipitation	19.8	32.3	35.6
Population	19.8	32.3	35.6
Humidity	19.8	32.4	35.6
Wind speed	20.0	32.1	35.7

and the last layer has one filter for the segmentation output. The segmentation model is trained with a learning rate of 0.0001 and a weight of 3 in the “fire” class. It is implemented using TensorFlow [56].

We perform the hyperparameter selection for the deep learning model by grid search. The best model is selected as the one with the best AUC PR on the validation dataset. We implement the number of layers and the number of filters in each layer as hyperparameters. We explore the number of filters in the first convolutional block between 16, 32, and 64. We allow the number of residual blocks in the encoder portion of the network to vary between 1 and 4, with the number of filters doubling in each subsequent block. We define the number of filters in the decoder portion of the segmentation model symmetrically.

with a stride of 1×1 , pooling is 2×2 , and the dropout rate is 0.1. All convolutional blocks have 16 filters except for the middle two residual blocks (ResBlocks), which uses 32 filters,

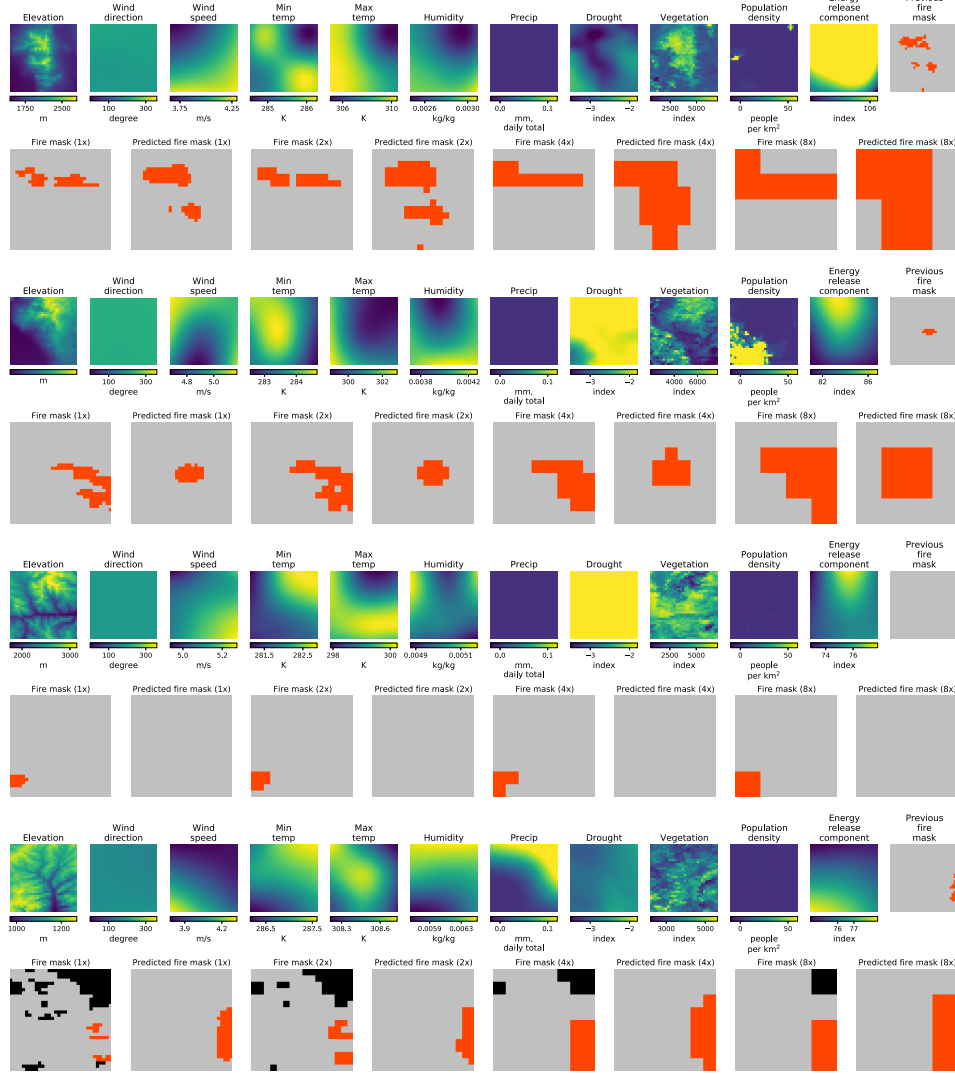


Fig. 10. Examples of worse coarse segmentation predictions. Each pair of rows is a sample.

We explore batch sizes of 32, 64, 128, and 256, learning rates of 0.01, 0.001, 0.0001, and 0.00001, dropout rates in increments of 0.1, and adding L1 and L2 regularization in powers of 10.

B. Nondeep Learning Models

The two nondeep learning models are implemented using Scikit-learn [57]. For these models, the training dataset is balanced to contain the fire and nonfire examples in an approximately 1:1 ratio. Metrics are calculated on the entire test dataset with no balancing. Unknown labels are ignored during training and evaluation.

With these models, the image segmentation is performed areawise (pixelwise). Consequently, the input variables correspond to a single area (pixel) in the input region. To provide spatial context, we include the eight neighboring areas (pixels) in a square around the input area (pixel). To avoid missing data, we exclude the 1-pixel border around the image during training and evaluation.

For the logistic regression model, we use the default hyperparameters from the Scikit-learn package. For the random forest, we sweep over values of the maximum tree depth between 1 and 20. As we increase the maximum depth, precision increases, while recall decreases. Therefore, we select the value with the highest AUC (PR) on the validation set. Using this methodology, we select a maximum depth of 15. All other hyperparameters are set to the default values from the Scikit-learn package.

Initial results from both models show that they overpredict the presence of fire with very low precision and high recall. Therefore, we weigh the “no fire” class using a weight defined as follows:

$$\frac{\text{number of examples}}{\text{number of “no fire” examples}} \times W$$

where W is a factor we adjust to increase the weight. We search over increasing values of W from 1 to 8. We select the value of the W that yields the highest AUC (PR) on the validation dataset. For logistic regression, W is 7, and for the random forest, W is 5.

C. Random Forest Feature Analysis

The 20 most important features in the random forest model are reported in Table VI. Feature importance is considered as the mean decrease in Gini impurity brought by splitting on that feature across the trees. The Gini impurity of a set of points after a split is the probability of misclassification if each point is randomly classified according to the distribution of labels in the set of points. Features are numbered in row-major order for the nine pixels (3×3 square) included in each example. Pixel number 5 is the input pixel. For example, “previous fire mask (3)” is the previous fire mask of the top-right neighbor of the input pixel. Based on this analysis, the previous fire mask of the input pixel is the most important feature, followed by the previous fire mask of the neighboring pixels. ERC follows as the next most important feature, and the remaining features all have very low importance. Precipitation for all pixels is the least important feature.

D. Logistic Regression Feature Analysis

The 20 largest magnitude coefficients of the logistic regression model by absolute value are reported in Table VII. Features are numbered in row-major order for the nine pixels (3×3 square) included in each example. Pixel number 5 is the input pixel. For example, “Min temp. (3)” is the minimum temperature of the top-right neighbor of the input pixel.

We conduct a feature ablation study by comparing the results of removing each feature during training. The resulting metrics are reported in Table VIII. This analysis demonstrates that the previous fire mask is the most important feature of the logistic regression model, similar to the other models.

APPENDIX B ADDITIONAL VISUALIZATIONS

Figs. 7 and 8 provide additional examples of fire spreading prediction results. Figs. 9 and 10 provide additional examples of prediction results at coarser resolution.

ACKNOWLEDGMENT

The authors would like to thank Zack Ontiveros, Carla Cromberg, Jason Hickey, Qing Wang, Cenk Gazen, Shreya Agrawal, Winston Huang, David Tian, Renaud Couzin, Ian Post, Michael Brenner, Fei Sha, John Burge, Zvika Ben-Haim, Avichai Tendler, Sella Nevo, and Omer Nevo.

REFERENCES

- [1] G. R. van der Werf *et al.*, “Global fire emissions estimates during 1997–2016,” *Earth Syst. Sci. Data*, vol. 9, no. 2, pp. 697–720, 2017. [Online]. Available: <https://essd.copernicus.org/articles/9/697/2017/>
- [2] V. Kollanus *et al.*, “Mortality due to vegetation fire—originated PM 2.5 exposure in Europe—Assessment for the years 2005 and 2008,” *Environ. Health Perspect.*, vol. 125, no. 1, pp. 30–37, Jan. 2017.
- [3] A. L. Westerling, H. G. Hidalgo, D. R. Cayan, and T. W. Swetnam, “Warming and earlier spring increase western us forest wildfire activity,” *Science*, vol. 313, no. 5789, pp. 940–943, 2006.
- [4] (2019). CoreLogic. 2019 Wildfire Risk Report. [Online]. Available: https://www.corelogic.com/downloadable-docs/wildfire-report_0919-01-screen.pdf
- [5] D. R. Brillinger, H. K. Preisler, and J. W. Benoit, “Risk assessment: A forest fire example,” *Lect. Notes-Monograph*, vol. 1, pp. 177–196, Jan. 2003.
- [6] M. Thompson *et al.*, “Application of wildfire risk assessment results to wildfire response planning in the southern Sierra Nevada, California, USA,” *Forests*, vol. 7, no. 12, p. 64, Mar. 2016.
- [7] M. P. Thompson *et al.*, “Risk management and analytics in wildfire response,” *Current Forestry Rep.*, vol. 5, no. 4, pp. 226–239, Dec. 2019.
- [8] S. Jazebi, F. de Leon, and A. Nelson, “Review of wildfire management techniques—Part II: Urgent call for investment in research and development of preventative solutions,” *IEEE Trans. Power Del.*, vol. 35, no. 1, pp. 440–450, Feb. 2019.
- [9] A. Sharma, I. Goyal, and S. Figueira, “Pin&post—app-based reporting of electrical fire hazards to prevent wildfires,” in *Proc. IEEE Global Hum. Technol. Conf. (GHTC)*, Oct. 2020, pp. 1–7.
- [10] E. Rigolot *et al.*, “Fuel-break assessment with an expert appraisal approach,” in *Forest Fire Research and Wildland Fire Safety*. Luso, Coimbra, Portugal: Millpress Science Publishers, Nov. 2002, pp. 18–23.
- [11] E. Marino, C. Hernando, R. Planelles, J. Madrigal, M. Guijarro, and A. Sebastián, “Forest fuel management for wildfire prevention in Spain: A quantitative SWOT analysis,” *Int. J. Wildland Fire*, vol. 23, no. 3, pp. 373–384, 2014.
- [12] A. L. Sullivan, “Wildland surface fire spread modelling, 1990–2007. 1: Physical and quasi-physical models,” *Int. J. Wildland Fire*, vol. 18, no. 4, pp. 349–368, Jun. 2009.
- [13] A. L. Sullivan, “Wildland surface fire spread modelling, 1990–2007. 2: Empirical and quasi-empirical models,” *Int. J. Wildland Fire*, vol. 18, no. 4, pp. 369–386, Jun. 2009.
- [14] A. L. Sullivan, “Wildland surface fire spread modelling, 1990–2007. 3: Simulation and mathematical analogue models,” *Int. J. Wildland Fire*, vol. 18, no. 4, pp. 387–403, Jun. 2009.
- [15] P. Jain, S. C. P. Coogan, S. G. Subramanian, M. Crowley, S. Taylor, and M. D. Flannigan, “A review of machine learning applications in wildfire science and management,” *Environ. Rev.*, vol. 28, no. 4, pp. 478–505, Dec. 2020.
- [16] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, “Google Earth engine: Planetary-scale geospatial analysis for everyone,” *Remote Sens. Environ.*, vol. 202, pp. 18–27, Dec. 2017.
- [17] L. Giglio and C. Justice, “MOD14A1 MODIS/terra thermal anomalies/fire daily L3 global 1 km SIN grid V006,” NASA EOSDIS Land Processes DAAC, Washington, DC, USA, Tech. Rep., 2015, doi: [10.5067/MODIS/MOD14A1.006](https://doi.org/10.5067/MODIS/MOD14A1.006).
- [18] K. Didan and A. Barreto, “VIIRS/NPP vegetation indices 16-day L3 global 500 m SIN grid V001,” NASA EOSDIS Land Processes DAAC, Washington, DC, USA, Tech. Rep., 2018, doi: [10.5067/VIIRS/VNP13A1.001](https://doi.org/10.5067/VIIRS/VNP13A1.001).
- [19] T. G. Farr *et al.*, “The shuttle radar topography mission,” *Rev. Geophys.*, vol. 45, no. 2, pp. 1–33, 2007. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005RG000183>
- [20] P. Laurent, F. Mouillot, C. Yue, P. Ciais, M. V. Moreno, and J. M. P. Nogueira, “FRY, a global database of fire patch functional traits derived from space-borne burned area products,” *Sci. Data*, vol. 5, no. 1, pp. 1–12, Dec. 2018.
- [21] N. Andela *et al.*, “The global fire atlas of individual fire size, duration, speed and direction,” *Earth Syst. Sci. Data*, vol. 11, no. 2, pp. 529–552, Apr. 2019.
- [22] T. Artés *et al.*, “A global wildfire dataset for the analysis of fire regimes and fire behaviour,” *Sci. Data*, vol. 6, no. 1, pp. 1–11, Dec. 2019.
- [23] K. C. Short, “Spatial wildfire occurrence data for the United States, 1992–2015 [FPA_FOD_20170508],” Forest Service Res. Data Arch., Fort Collins, CO, USA, Tech. Rep., 2017, doi: [10.2737/RDS-2013-0009.4](https://doi.org/10.2737/RDS-2013-0009.4).
- [24] C. Díaz-Avalos, D. L. Peterson, E. Alvarado, S. A. Ferguson, and J. E. Besag, “Space-time modelling of lightning-caused ignitions in the Blue Mountains, Oregon,” *Can. J. Forest Res.*, vol. 31, no. 9, pp. 1579–1593, Sep. 2001.
- [25] M. A. Krawchuk, S. G. Cumming, M. D. Flannigan, and R. W. Wein, “Biotic and abiotic regulation of lightning fire initiation in the mixed-wood boreal forest,” *Ecology*, vol. 87, no. 2, pp. 458–468, Feb. 2006.
- [26] F. Verdú, J. Salas, and C. Vega-García, “A multivariate analysis of biophysical factors and forest fires in Spain, 1991–2005,” *Int. J. Wildland Fire*, vol. 21, no. 5, pp. 498–509, 2012.
- [27] A. Ganteaume, A. Camia, M. Jappiot, J. San-Miguel-Ayán, M. Long-Fournel, and C. Lampin, “A review of the main driving factors of forest fire ignition over Europe,” *Environ. Manage.*, vol. 51, no. 3, pp. 651–662, Mar. 2013.
- [28] D. Vecín-Arias, F. Castedo-Dorado, C. Ordóñez, and J. R. Rodríguez-Pérez, “Biophysical and lightning characteristics drive lightning-induced fire occurrence in the central plateau of the Iberian Peninsula,” *Agricult. Forest Meteorol.*, vol. 225, pp. 36–47, Sep. 2016.

- [29] D. T. Bui, Q.-T. Bui, Q.-P. Nguyen, B. Pradhan, H. Nampak, and P. T. Trinh, "A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area," *Agricult. Forest Meteorol.*, vol. 233, pp. 32–44, Feb. 2017.
- [30] Y. O. Sayad, H. Mousannif, and H. Al Moatassime, "Predictive modeling of wildfires: A new dataset and machine learning approach," *Fire Saf. J.*, vol. 104, pp. 130–146, Mar. 2019.
- [31] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
- [32] J. Bennett *et al.*, "The Netflix prize," in *Proc. KDD Cup Workshop*, New York, NY, USA, 2007, p. 35.
- [33] A. Alonso-Betanzos *et al.*, "An intelligent system for forest fire risk prediction and fire fighting management in Galicia," *Expert Syst. Appl.*, vol. 25, no. 4, pp. 545–554, Nov. 2003.
- [34] G. E. Sakr, I. H. Elhajj, and G. Mitri, "Efficient forest fire occurrence prediction for developing countries using two weather parameters," *Eng. Appl. Artif. Intell.*, vol. 24, no. 5, pp. 888–894, Aug. 2011.
- [35] R. Dutta, J. Aryal, A. Das, and J. B. Kirkpatrick, "Deep cognitive imaging systems enable estimation of continental-scale fire incidence from climate data," *Sci. Rep.*, vol. 3, no. 1, pp. 1–4, 2013.
- [36] J. L. Hodges and B. Y. Lattimer, "Wildland fire spread modeling using convolutional neural networks," *Fire Technol.*, vol. 55, no. 6, pp. 2115–2142, 2019.
- [37] D. Radke, A. Hessler, and D. Ellsworth, "FireCast: Leveraging deep learning to predict wildfire spread," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 4575–4581.
- [38] S. P. Walters, N. J. Schneider, and J. D. Guthrie, "Geospatial multi-agency coordination (GeoMAC) wildland fire perimeters, 2008," *U.S. Geol. Surv. Data*, vol. 612, no. 6, 2011, doi: [10.3133/ds612](https://doi.org/10.3133/ds612).
- [39] F. P. Schoenberg, R. Peng, and J. Woods, "On the distribution of wildfire sizes," *Environmetrics*, vol. 14, no. 6, pp. 583–592, 2003.
- [40] B. D. Malamud, J. D. A. Millington, and G. L. W. Perry, "Characterizing wildfire regimes in the United States," *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 13, pp. 4694–4699, 2005.
- [41] P. E. Dennison, S. C. Brewer, J. D. Arnold, and M. A. Moritz, "Large wildfire trends in the western United States, 1984–2011," *Geophys. Res. Lett.*, vol. 41, no. 8, pp. 2928–2933, Apr. 2014.
- [42] D. Stojanova, A. Kobler, P. Ogrinc, B. Zenko, and S. Džeroski, "Estimating the risk of fire outbreaks in the natural environment," *Data Mining Knowl. Discovery*, vol. 24, no. 2, pp. 411–442, Mar. 2012.
- [43] A. E. Van Beusekom *et al.*, "Fire weather and likelihood: Characterizing climate space for fire occurrence and extent in Puerto Rico," *Climatic Change*, vol. 146, nos. 1–2, pp. 117–131, Jan. 2018.
- [44] L. S. Bradshaw, J. E. Deeming, R. E. Burgan, and J. D. Cohen, "The 1978 national fire-danger rating system: Technical documentation," Compilers, U.S. Dept. Agricult., Forest Service, Intermountain Forest Range Exp. Station, Ogden, UT, USA, Gen. Tech. Rep. INT-169, 1984.
- [45] J. K. Balch, B. A. Bradley, J. T. Abatzoglou, R. C. Nagy, E. J. Fusco, and A. L. Mahood, "Human-started wildfires expand the fire niche across the United States," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 11, pp. 2946–2951, Mar. 2017. [Online]. Available: <https://www.pnas.org/content/114/11/2946>
- [46] F. Huot *et al.*, "Deep learning models for predicting wildfires from historical remote-sensing data," 2020, *arXiv:2010.07445*.
- [47] J. T. Abatzoglou, "Development of gridded surface meteorological data for ecological applications and modelling," *Int. J. Climatol.*, vol. 33, no. 1, pp. 121–131, Jan. 2013.
- [48] J. T. Abatzoglou, D. E. Rupp, and P. W. Mote, "Seasonal climate variability and change in the Pacific Northwest of the United States," *J. Climate*, vol. 27, no. 5, pp. 2125–2142, Mar. 2014.
- [49] *Gridded Population of the World, Version 4 (GPWv4): Population density, Revision 11*, C. Int. Earth Sci. Inf. Netw., New York, NY, USA, 2018.
- [50] D. K. Davies, S. Ilavajhala, M. M. Wong, and C. O. Justice, "Fire information for resource management system: Archiving and distributing modis active fire data," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 72–79, Jan. 2008.
- [51] W. Schroeder *et al.*, "Early characterization of the active fire detection products derived from the next generation NPOESS/VIRS and GOES-R/ABI instruments," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2010, pp. 2683–2686.
- [52] T. J. Schmit, P. Griffith, M. M. Gunshor, J. M. Daniels, S. J. Goodman, and W. J. Lebar, "A closer look at the ABI on the GOES-R series," *Bull. Amer. Meteorol. Soc.*, vol. 98, no. 4, pp. 681–698, 2017.
- [53] A. Savtchenko *et al.*, "Terra and Aqua MODIS products available from NASA GES DAAC," *Adv. Space Res.*, vol. 34, no. 4, pp. 710–714, Jan. 2004.
- [54] Y. S. Aurelio, G. M. de Almeida, C. L. de Castro, and A. P. Braga, "Learning from imbalanced data sets with weighted cross-entropy function," *Neural Process. Lett.*, vol. 50, no. 2, pp. 1937–1949, Oct. 2019.
- [55] H. R. Sofaer, J. A. Hoeting, and C. S. Jarnevich, "The area under the precision-recall curve as a performance metric for rare binary events," *Methods Ecol. Evol.*, vol. 10, no. 4, pp. 565–577, Apr. 2019.
- [56] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*.
- [57] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.



Fantine Huot received the M.S. degree in science and executive engineering from the École nationale supérieure des mines de Paris, Paris, France, in 2013, and the Ph.D. degree in geophysics from Stanford University, Stanford, CA, USA, in 2022.

She is passionate about AI for science. She has worked in various fields, including seismology, remote sensing, natural language generation, and high-performance computing. She is currently a Researcher and an Engineer with Google Inc., Mountain View, CA.

R. Lily Hu received the B.A.Sci. degree in engineering science from the University of Toronto, Toronto, ON, Canada, in 2012, and the M.S. and Ph.D. degrees in engineering from the University of California at Berkeley, Berkeley, CA, USA, in 2014 and 2016, respectively.

She is currently a Researcher and an Engineer with Google Research, Mountain View, CA. She conducted research in machine learning for the physical sciences and sustainability, drawing upon computer vision, decision analysis, and optimization.

Nita Goyal received the B.Tech. degree in computer science (CS) from IIT Kanpur, Kanpur, India, in 1987, and the Ph.D. degree in CS from Stanford University, Stanford, CA, USA, in 1993.

She has previously been in several tech startups. She has been a Software Engineer and a Researcher with Google Inc., Mountain View, CA, since 2013. She has worked in various fields, including natural language understanding, search and recommendations, and machine learning for climate and sustainability.

Dr. Goyal is the first woman CS graduate from any of the IITs.

Tharun Sankar received the B.A. degree in computer science from Cornell University, Ithaca, NY, USA, in 2020.

He has been a Software Engineer with Google Research, Mountain View, CA, USA, since 2020. His research work has spanned fields such as computer vision, audio processing, computational sustainability, and natural language processing.

Matthias Ihme received the Ph.D. degree from Stanford University, Stanford, CA, USA, in 2008.

He has served on the Faculty in the Department of Aerospace Engineering, University of Michigan, Ann Arbor, MI, USA. He is currently an Assistant Professor of mechanical engineering with Stanford University. His primary research interests are the prediction and modeling of turbulent reacting flows, scalar mixing, and subgrid-scale modeling for large eddy simulation.

Dr. Ihme was a recipient of the NSF CAREER Award, the AFOSR Young Investigator Award, the ONR Young Investigator Award, and the NASA Early Career Faculty Award.

Yi-Fan Chen received the Ph.D. degree in applied physics from Cornell University, Ithaca, NY, USA, in 2005.

He has been a Software Engineer and a Researcher with Google Inc., Mountain View, CA, USA, since 2012. He has worked in various fields, including nonlinear and ultrafast optics, photolithography, video processing, recommendation systems, and high-performance computing.