5-2022

Predicting the Likelihood and Scale of Wildfires in California using Meteorological and Vegetation Data

Matthew Walters
University of Arkansas, Fayetteville

Predicting the Likelihood and Scale of Wildfires in California using Meteorological and Vegetation Data

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science Industrial Engineering

by

Matthew Walters
University of Arkansas
Bachelor of Science in Industrial Engineering, 2021

May 2022
University of Arkansas

This thesis is approved for recommendation to the Graduate Council

Chase Rainwater, Ph.D.
Thesis Director:

Xiao Liu, Ph.D.
Committee member

Jack Cothren, Ph.D.
Committee member

Abstract

Wildfires have devastating ecological, environmental, economical, and public health impacts through the deterioration of water and air quality, CO_2 emissions, property damage, and lung illnesses. The early detection and prevention of wildfires allow for the minimization of these risks. The use of Artificial Intelligence (AI) in wildfire detection and prediction has been highly researched as a tool to assist firefighters in stopping wildfires in its early stages. The three common wildfire prediction categories include image and video detection, behavior prediction, and susceptibility prediction. Data such as climate, weather, vegetation, satellite images, and historical wildfire data is most commonly used. Many approaches such as Support Vector Machines (SVM), Basic Neural Networks (BNN), Recurrent Neural Networks (RNN), Long Short-Term Memory Networks (LSTM), and Convolutional Neural Networks (CNN) have been highly used in wildfire prediction. The goal of this research is to discover the best combination of data and prediction methodology that most accurately predicts a locations likelihood and scale of a wildfire occurring in any given month to assist in the resource allocation and planning of fighting wildfires.

Table of Contents

1 Introduction

Although wildfires are a natural occurring event in a healthy ecosystem, they have detrimental effects to property, human health, and the environment [1]. Since 1983, the number of wildfires each year in the U.S. has remained relatively constant at around 70,000 annually. However, the intensity of the wildfires have increased since 1983 from roughly 3 million acres burned to around 7 million acres burned annually past 2010 [2]. California is responsible for a large majority of U.S. wildfires. In 2020, California wildfires were responsible for 4.1 million acres of the 10 million total acres in the U.S. [3]. For this reason, this research will focus on California.

With the increase of data related to wildfires, there has been an increase in research related to the prediction and detection of forest fires using artificial intelligence [4]. One of the most important motivating factors behind the application of artificial intelligence in fighting wildfires is locating the most high-risk areas, so that firefighters and other resources can be deployed to those areas as soon as possible [5]. Many datasets related to climate, weather, vegetation, satellite, and historical wildfires have shown promising results in predicting wildfires. New data sources have recently been created that have yet to be utilized in the existing research.

The goal of this research is to predict the locations that are most vulnerable to wildfires to know where resources should be deployed to fight wildfires while it is in its early stages. This will be accomplished by (1) identifying the best features to include, (2) exploring various prediction methodologies, and (3) optimizing the accuracy of the prediction model. This research will explore various combinations of features and artificial intelligence techniques to find the best approach that maximizes the accuracy of predicting the most vulnerable locations to wildfires.

Chapter 2 begins with a discussion of wildfires including how they ignite and spread, the negative effects of wildfires, and how wildfires are fought and prevented. Next it continues with a literature review of the three main fields of machine learning in wildfire prediction: image and video recognition of wildfires, behavior prediction of wildfires, and susceptibility prediction of wildfires. Chapter 3 discusses the data and the prediction methodology. It begins by discussing the datasets used and the data processing required to extract the features. Then the prediction methodology is discussed. Chapter 4 explains the metrics used to measure the models performance, and then the top performing hyper-parameters for each metric is displayed and discussed. Finally, chapter 5 concludes the paper by providing a summary and an outlook for the future work in this area of study.

2 Literature Review

This chapter focuses on the review of wildfires and the methodologies used in the prediction of wildfires. The first section focuses on how wildfires ignite and spread, the effects of wildfires, and how wildfires are fought and prevented. The second section explores the previous work related to the use of machine learning on wildfires.

2.1 Background on Wildfires

There are various ignition sources that cause a fire to start, and there are various conditions that allow a fire to spread more rapidly. Wildfires present some environmental, ecological, and health concerns. There are various techniques used to put out an ongoing fire as well as techniques to contain and prevent the spread of a future fire.

2.1.1 How Wildfires Ignite and Spread

All wildfires begin with an ignition source. This ignition source can either be human or naturally occurring. Some human sources include campfires, smoking, power line failures, sparks from railroads, and arson. Some naturally occurring sources include lightning strikes, lava, and meteors [6]. There are certain conditions that allow a fire to start and spread more rapidly. High temperature, low humidity, little rainfall, dry vegetation, and fast winds are all conditions that contribute to violent and fast spreading wildfires. Furthermore, certain plants, trees, and shrubs contain oils that burn more quickly and intensely. The topography is also a factor, for fires move more quickly uphill. This research aims to utilize these variables when predicting if a location is currently vulnerable to a wildfire [7].

2.1.2 Negative Effects of Wildfires

Although wildfires play a key role in shaping the ecosystem by allowing for a renewal of plants, it causes detrimental environmental, ecological, and health problems. Wildfires emit CO_2 into the atmosphere, and it is estimated that wildfires make up for 5 to 10 percent of annual CO_2 emissions each year [8]. Wildfires also affect the physical, chemical, and biological quality of streams, rivers, lakes, and reservoirs deteriorating the water quality making it undrinkable and uninhabitable for wildlife; and these changes can remain present for years and decades after the fire [9].

Though wildfires are a natural process that is integral to the life history of plants and animals in the ecosystem [10]. Nevertheless, severe wildfires can undermine the native biodiversity by destroying the native vegetation, introducing invasive species, and eliminating essential wildlife. Scientists are still studying the effects wildfires have on ecosystems, plant, and animal life [11].

The smoke produced from Wildfires contain pollutants such as particle pollution, which is a mix of very tiny solid and liquid particles suspended in air that can cause damage to lungs [12]. According to the Center for Disease Control and Prevention (CDC) smoke exposure increases respiratory and cardiovascular hospitalizations; emergency department visits; medication dispensations for asthma, bronchitis, chest pain, chronic obstructive pulmonary disease (commonly known as COPD), and respiratory infections; and medical visits for lung illnesses. It has also been associated with hundreds of thousands of deaths annually [13].

2.1.3 How Wildfires are Fought and Prevented

After a wildfire has been detected, it is declared “active,” and firefighters work to “contain” the fire. During the containment phase, firefighters surround the fire with a physical barrier utilizing rivers, obstacles, trenches, and controlled burns. After containment, firefighters work to “control” the fire. During this phase they will reinforce the barriers, remove fuel that could help the spread, and cool certain hot spots. A fire is considered “out” when no hot spot is detected within the containment area for at least 48 hours [7]

2.2 Machine Learning Applications in Wildfire Prediction

The literature reveals three different applications in machine learning for wildfires: detecting fires through video and images, predicting the behavior of wildfires such as spread and scale, and predicting the occurrence of wildfires. Table 2.1 contains a summary of the current literature related to the machine learning applications in wildfire prediction. The motivation behind these three prediction strategies is to assist firefighters in containing a wildfire in its early stages before it gets out of control.

2.2.1 Image and Video Recognition of Wildfires

The existing literature regarding the detection of fires through images and videos research how cameras, unmanned aerial vehicles (UAV), and satellite images can be used to detect fires in an effort to locate an extinguish fires quickly. Zhentian et al. [14] proposes attaching a camera and a microcomputer onto a UAV. The camera will send live footage to the microcomputer that will run the raw footage through a trained, lightweight CNN responsible for detecting wildfires. The CNN is able to recognize wildfires with around 83% accuracy. Chanthiya et al. [15] trained a Support Vector Machine (SVM) with LANDSAT satellite images to detect fires. Features such as land surface temperature, fire intensity, water vapor, and top of atmosphere temperature are extracted from the LANDSAT images. The SVM was able to outperform other techniques with an accuracy of 99.21%.

2.2.2 Behavior Prediction of Wildfires

Two types of behavior of a wildfire to predict are the directional spread of a wildfire and the scale of a forest fire. Liang et al. [16] uses meteorological factors, a back-propagation neural network (BPNN), a recurrent neural network (RNN), and a long short-term memory (LSTM) network to predict the scale of wildfires in Alberta, Canada. They found that LSTM exhibited the highest accuracy with 90.9%. They discovered that it is feasible to predict the scale of a forest wildfire at the beginning of its occurrence by using meteorological data. Perumal et al. [17] uses both a Gated Recurrent Unit (GRU) and a Long Short-Term Memory (LSTM) network to determine whether a wildfire will continue to burn and given that it does, predict which one of the eight cardinal directions the wildfire will spread. They discovered the GRU performs better for longer time series than the LSTM. The research showed promise in predicting the direction the wildfire will spread, but are unable to assess if the wildfire continues to burn.

2.2.3 Occurrence and Susceptibility Prediction of Wildfires

The goal of occurrence and susceptibility prediction of wildfires is to identify the locations that are currently most vulnerable to forest fires. As previously mentioned in Section 2.1.1, there are certain conditions that allow wildfires to start and spread more rapidly. The literature that exists in this area of study attempts to train deep learning algorithms that can discover these conditions and make predictions on the susceptibility of wildfires in certain locations throughout time. Zhang et al. [18] uses 14 meteorological and geographical features to predict how susceptible a location is to wildfires for a given year in the Yunnan Province in China. A convolutional neural network (CNN) is used to predict the probability that a forest fire would occur within a 500 meter x 500 meter patch of land. This model scored a 0.86 AUC, outperforming other approaches such as random forests, support vector machines, multilayer perceptron neural network, and kernel logistic regression. Natekar et al. [19] uses an LSTM network to predict the latitude and longitude coordinates of a wildfire in India. This research utilizes satelite observations from the Visible Infrared Imaging Radiometer Suite (VIIRS) to extract the date and location of wildfires as well as other meteorological data. The LSTM model could predict the occurrence of wildfires with a 94.77% accuracy.

2.2.4 Future of Deep Learning for Wildfire Prevention

The ideal fire detection and prevention system of the future will incorporate all three of the wildfire prediction categories. Meteorological data will be used to locate the most vulnerable areas most susceptible to wildfires to warn firefighters to be prepared. Satellite and UAV images will be used to detect a wildfire while it is still in its early stages. Behavioral models will be used to predict the spread of a wildfire to assist firefighters in stopping the spread. This research focuses on the susceptibility and scale prediction of wildfires by using satellite and meteorological data.

Author	Data Source	Prediction Task	Methodology
Jiao et al. [14]	Images	Image	CNN
Chanthiya et al. [15]	Satellite	Image	SVM
Liang et al. [16]	Meteorological	Behavior	LSTM
Perumal et al. [17]	Satellite	Behavior	RNN LSTM GRU
Zhang et al. [18]	Meteorological	Susceptibility	CNN
Natekar et al. [19]	Satellite Meteorlogical	Location	LSTM
Han et al. [20]	Video	Image Video	CNN RNN
Avula et al. [21]	Video	Video	CNN
Arteaga et al. [22]	Images	Image	CNN
Rahul et al. [23]	Images	Image	CNN
Chen et al. [24]	Images	Image	CNN
Maeda et al. [25]	Satellite	Susceptibility	ANN

Table 2.1: Wildfire Prediction Literature

3 Data Processing and Prediction Methodology

This chapter will begin by discussing the three datasets used in this research and the data processing required to transform the data into usable inputs and labels. Then the initial prediction methodology will be introduced such as the labeling technique, the loss function, and the baseline model.

3.1 Data and Data Processing

This section begins by first introducing the three open datasets from the National Aeronautics and Space Administration (NASA) and the California Department of Forestry and Fire Protection. Then the data processing that is required to transform the raw data into usable features is discussed.

3.1.1 Description of Data

The first dataset, supported by NASA is called Daymet which provides estimates of daily weather and climatology variables throughout North America [26]. Daymet utilizes daily meteorological observations from weather stations to produced estimates of seven daily weather parameters on a 1 km x 1km gridded surface from 1980 to 2020 [27] [28]. The seven weather parameters include minimum and maximum temperature, precipitation, vapor pressure, radiation, snow water equivalent, and day length. Figure 3.1 contains a choropleth map of the maximum temperature in California on January 1, 2020.

The second dataset, supported by NASA, is called MOD13Q1, and it utilizes a satellite based sensor called Moderate Resolution Imaging Spectroradiometer (MODIS) to determine the vegetation at a 250m resolution [29].The data ranges from 2001 to 2020, and it is generated every 16 days. MOD13Q1 uses Normalized Difference Vegetation Index (NDVI) as a measurement for vegetation. NDVI quantifies vegetation by measuring the difference

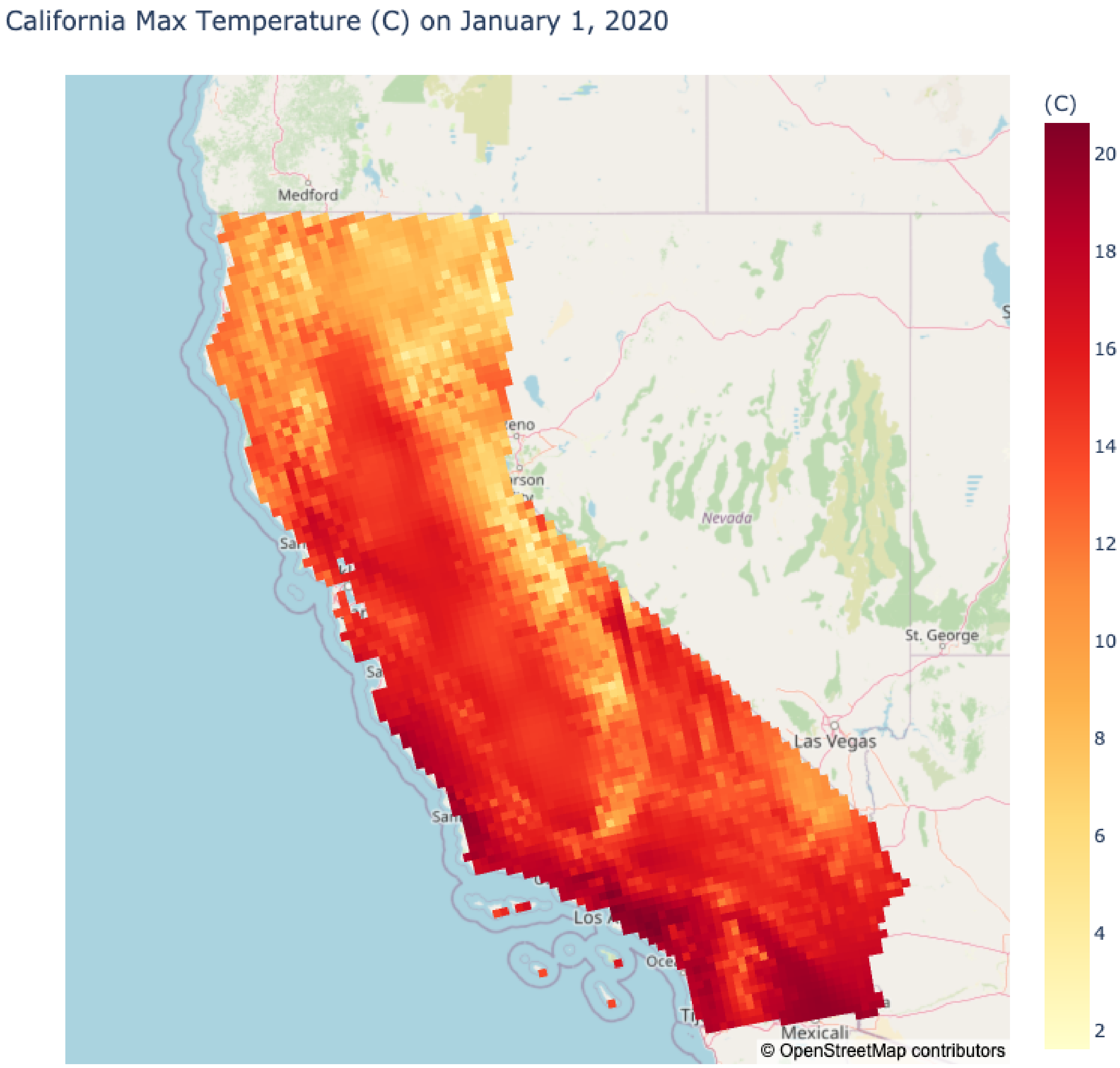

Figure 3.1: Example of Daymet data for California on January 1, 2020

between near-infrared (which vegetation strongly reflects) and red light (which vegetation absorbs) [30]. Figure 3.2 contains choropleth map of vegetation in California for January 2020.

The third dataset, provided by California Department of Forestry and Fire Protection is called CAL FIRE, and it contains historical and current data on wildfires. They provide information such as the acres destroyed, the year of the fire, fire discovery date time, fire containment date time, and the perimeter location of the fire which is stored as a polygon shape with longitude and latitude coordinates. The fire discovery date will act as the start date and the fire containment date will act as the end date of the wildfires. The data spans from 1950 to 2020. Figure 3.3 contains the area locations and acres burned of all California wildfires that took place in 2020.

3.1.2 Data Processing

As previously mentioned, the goal of this research is to predict the occurrence and scale of a wildfire for a given location and time period. Therefore, the data will need to be aggregated by time and location. The goal of the model is to make predictions on a locations likelihood and scale of a wildfire occurring during a specific month. From 1980 to 2020, Daymet contains 41 years worth of data, and ranging from 2001 to 2020, MOD13Q1 has 20 years worth of data. To aggregate the data by location, California will first be divided into grids. The grids are 10km x 10km grids which results in 3,776 total grids. Each grid for each year will contain its corresponding daily weather data for all 365 days in the year from Daymet as well as the which will act as the features for the model. Each grid will also contain the fire scale and occurrence data from CAL FIRE which will act as the labels.

Daymet has a RESTful API that allows a user to query the data by providing specific dates and the location range. The API request returns a spatio-temporal dataset containing the monthly seven weather parameters within the time and location range. Further data processing is required to keep only the California locations and to format the data properly. Finally, each of the values within the seven weather parameters are normalized between 0

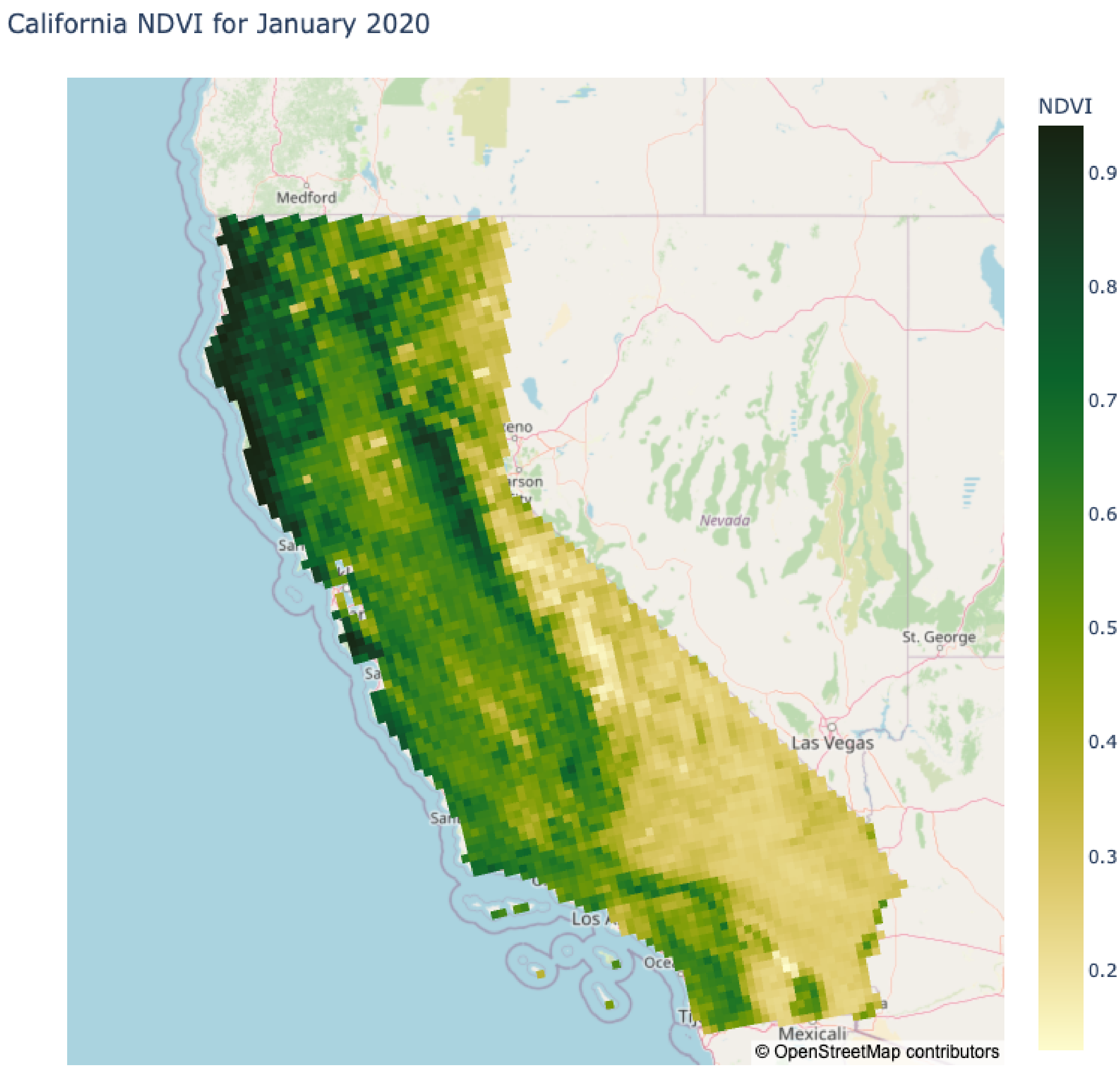

Figure 3.2: Example of NDVI for California in January of 2020

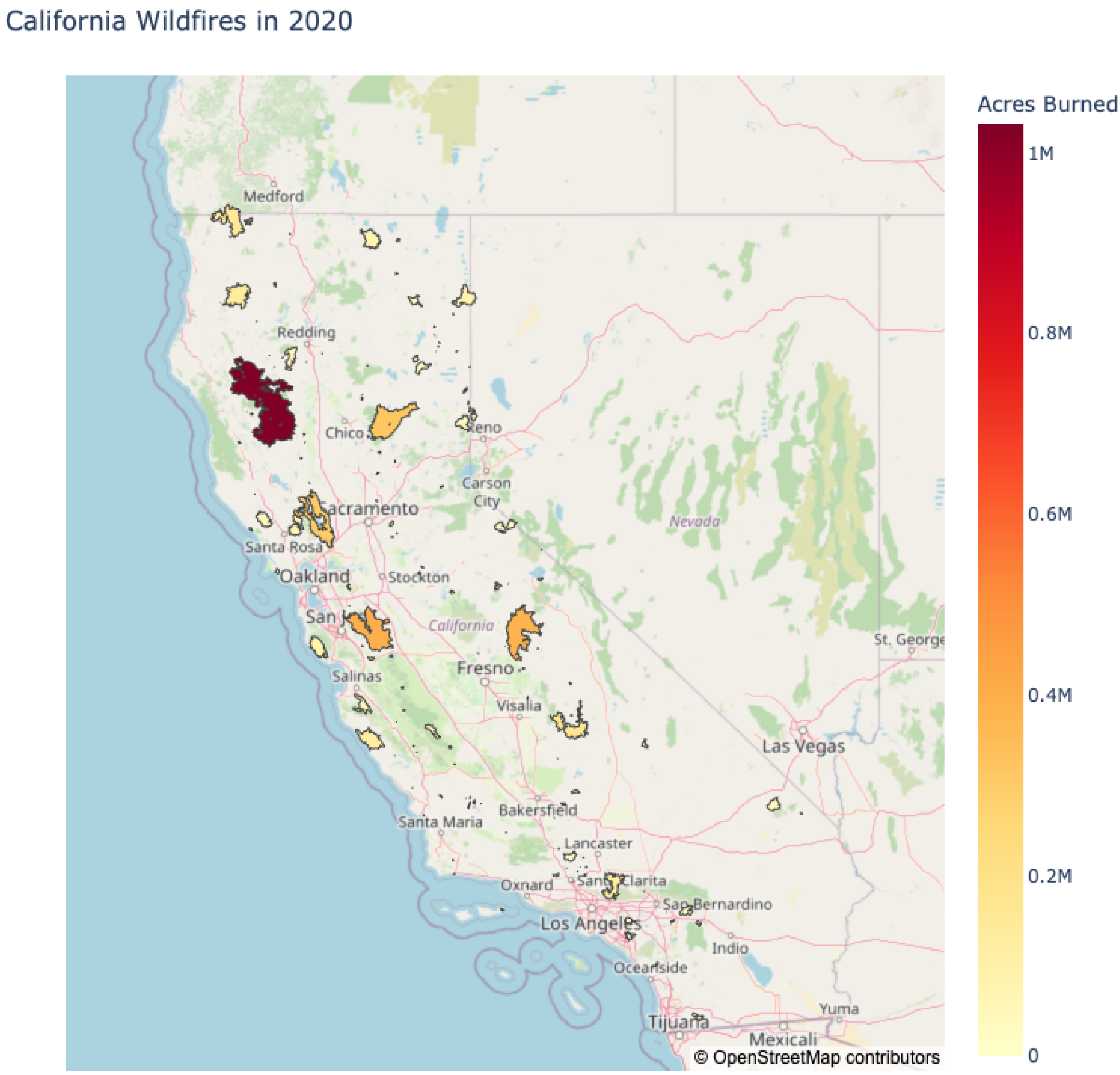

Figure 3.3: Locations and Scale of California Wildfires in 2020

and 1. These values will be used as the features for the model.

A similar approach is used to process the MOD13Q1 data. With data being generated every 16 days, the NDVI values generated right before the start of a month are used to represent the vegetation for that month. For example, the vegetation data generated by MOD13Q1 on January 17, 2020 will be used as features for February 2020. The reasoning for this approach is to capture the vegetation status of a location before a wildfire occurs during the month of interest. After the vegetation values for each month from 2001 to 2020 are determined, then the 250m x 250m values are aggregated to fit the 10km x 10km grids using the mean of the values within each grid. Then finally, the values are normalized between 0 and 1.

CAL FIRE also has a similar API that allows the user to query the most updated information from their server. Using the API, a dataset containing wildfires from 1950 to 2020 is created with the five categories of fire year, start date, end date, acres burned, and location perimeter of wildfires. Depending on the occurrence and scale, six categorical labels between 0 and 5 are used:

- 0 - No wildfire
- 1 - Less than 100 acres burned
- 2 - Between 100 and 1,000 acres burned
- 3 - Between 1,000 and 10,000 acres burned
- 4 - Between 10,000 and 100,000 acres burned
- 5 - More than 100,000 acres burned

The wildfire locations must then be aggregated into the 10km x 10km grids. Figure 3.4 illustrates this grid for 2020 wildfires after aggregation and categorization. The category of the wildfire within each grid will be used as the label for the classification model. The wildfire history of a grid is also used as a feature within the model. Both short-term and long-term histories are captured by using a 1, 5, 10, and 20 year wildfire history. These four

”history” features capture the largest wildfire category (0-5) that occurred during the past 1, 5, 10, or 20 years. These four features are then one-hot encoded thus creating 24 new features (4 x 6).

3.2 Prediction Methodology

This section describes the methodology used to make monthly predictions on the likelihood and scale of wildfires using a basic neural network.

3.2.1 Basic Model

The goal of the model is to determine if a wildfire will occur within a 10km x 10km grid during a specific month, and if it predicts an occurrence, then the model will predict the scale of the wildfire. This is accomplished via a classification model using the 0-5 labels as discussed in section 3.1.2. There are six outputs representing each of the six categorical labels. The inputs include the latitude, longitude, month, maximum temperature, minimum temperature, vapor pressure, snow water equivalent, precipitation, vegetation, one year fire history, five year fire history, ten year fire history, and twenty year fire history. The training sample contains data from 2001 - 2015, and the test sample contains 2016 - 2020 data. Since the categories are not well balanced with a majority of instances containing a label of 0 (no wildfire occurring), oversampling is used to avoid a biased model that mostly predicts 0. Random oversampling is applied to the training data to ensure the categories are balanced. Since the six classes are numerically ordered, and ordinal regression layer is used as the final layer. When labels are numerically ranked, ordinal regression is advantageous over the standard classification approach. Ordinal Regression accounts for the numerical ranking of the labels whereas the standard classification approach treats each class independently [31]. Dropout and regularization are both used to prevent over fitting the model. The hyper-parameters of the basic neural network will be tuned to optimize the models performance. Section 4.2 will discuss the tuning of the hyper-parameters.

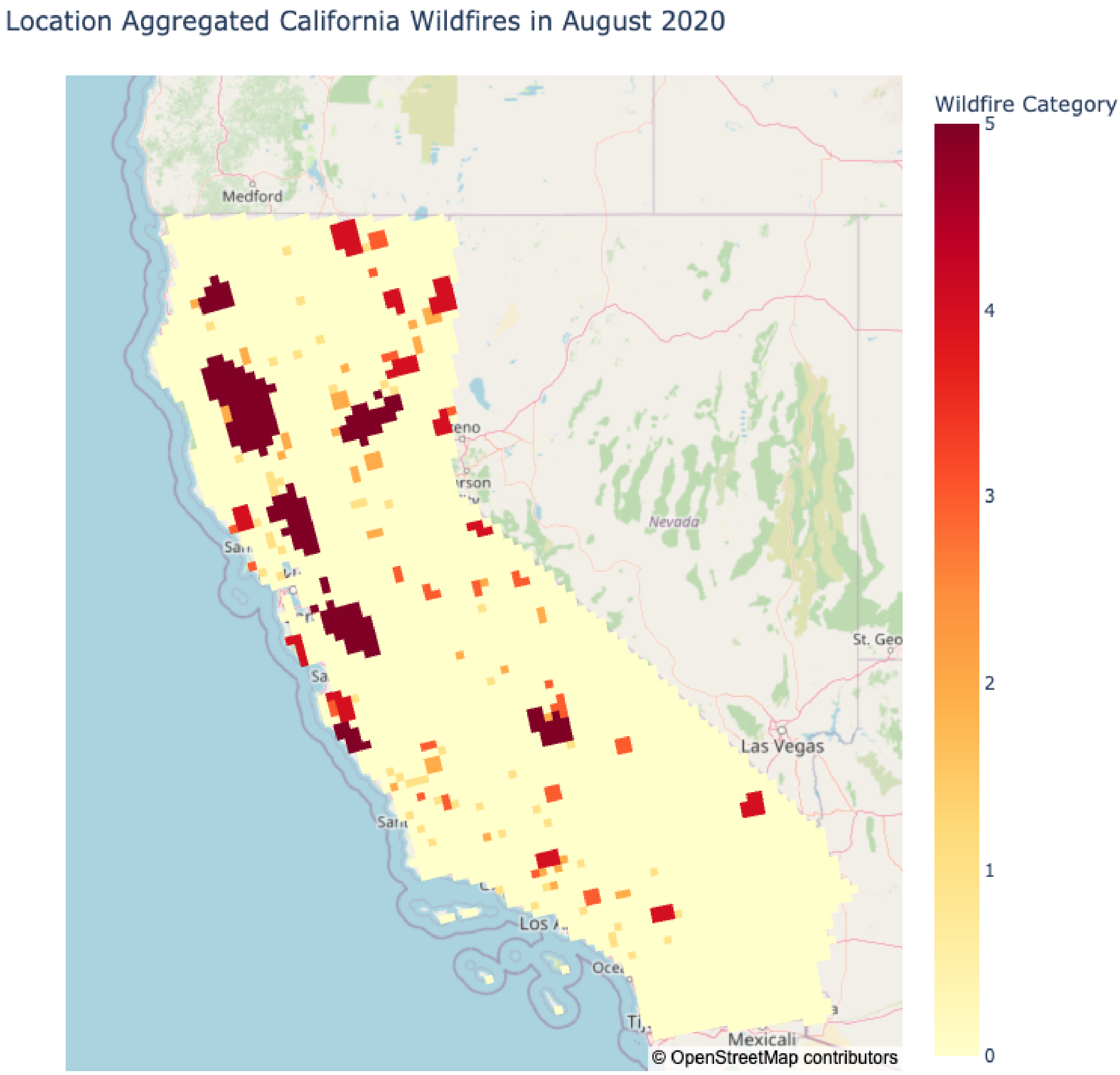

Figure 3.4: Aggregated Location and Category of California Wildfires in 2020

4 Results and Discussion

This section first discusses the metrics used to measure the performance of the models ability to predict the likelihood and scale of a wildfire occurring at each location during each month. Then, the grid search used to tune the hyper-parameters of the model will be discussed. Finally, the results will be displayed.

4.1 Metrics

As previously mentioned, the two primary goals of the model is to accurately predict the likelihood and scale of a wildfire. Therefore, metrics that capture both of these performances are necessary. The metrics used for this research are accuracy, occurrence recall and precision, large fire occurrence recall and precision, scale accuracy, and relaxed scale accuracy. Occurrence recall and precision finds the recall and precision of detecting the occurrence of wildfires ignoring the correct scale of the fire (1-5). Large scale fire occurrence recall and precision finds the precision and recall of predicting large wildfires (3,4, or 5) correctly. For example, if a wildfire has a label of 3, 4 or 5, then a prediction of 3, 4, or 5 is considered correct. Scale accuracy finds the accuracy of predicting the scale of a fire that occurs correctly. Relaxed scale accuracy finds the accuracy of the models ability to predict the correct scale closely. For example if a wildfire has a label of 4, then a prediction of either 3 or 5 is considered correct. Since wildfires are a costly and dangerous event, false negatives are much more costly than false positives, therefore, maximizing recall should be prioritized over precision.

4.2 Hyper-Parameters and Results

There are several hyper-parameters that are tuned to optimize the performance of the model. A grid search is performed during tuning. The hyper-parameters tuned are the number of hidden layers, number of neurons, number of epochs, the batch size, the learning rate, dropout value, and regularization value. The following values were used for the grid search:

- learning rate - 0.005, 0.05, 0.1
- number of epochs - 50, 100, 200
- batch size - 5000, 10000, 50000
- number of hidden layers - 2, 3, 4
- number of neurons per layer - 50, 100, 200
- l2 regularization value - 0, 0.0001, 0.0005, 0.001
- dropout value - 0, 0.1, 0.15, 0.2, 0.3

Table 4.1 displays the results for the top five performers in accuracy, occurrence recall, large fire occurrence recall, scale accuracy, and relaxed scale accuracy.

4.3 Discussion

Looking at the occurrence recall results, this model shows promise in its ability to identify locations that are susceptible to wildfires with the top five models identifying locations that had wildfires 97.1%, 95%, 94.8%, 94.7%, and 94.6% of the time. The model also performs well at identifying large fires (greater than 1,000 acres) with top five large fire occurrence recall models identifying occurring large fires 86.2%, 81.7%, 77.7%, 75.5%, and 74.9% of the time. The model has difficulty predicting the correct scale of an occurring wildfire indicated by the low performances in scale accuracy and relaxed scale accuracy. The top five performing models in the scale accuracy category only predict the correct scale of an occuring fire 27.4%, 26.4%, 26.4%, 26.2%, and 25.8% of the time. The top five performing models in the relaxed scale accuracy category only predict one scale away from the correct scale 63.8%, 62.6%, 61.2%, 61.1%, and 60.8% of the time. All of the models tend to over predict the occurrence of wildfires indicated by the low occurrence precision and large fire

Hyper-Parameters	A	OR/P	LFOR/P	SA	RSA
(0.005, 200, 5000, 3, 200, 0, 0)	**92.2**	24.6/8.8	9.2/7.1	5.2	12.0
(0.005, 200, 50000, 3, 200, 0, 0)	**90.5**	31.3/8.6	14.5/6.2	7.1	16.4
(0.005, 50, 5000, 3, 200, 0, 0)	**89.2**	36.3/8.4	16.5/6.3	8.2	18.8
(0.050, 200, 5000, 3, 200, 0, 0)	**87.6**	43.3/8.5	23.8/6.9	9.8	24.9
(0.050, 200, 50000, 3, 200, 0, 0)	**87.5**	44.0/8.6	26.2/6.6	10.0	24.6
(0.10, 50, 5000, 4, 100, 0.1, 0)	43.4	**97.1**/3.8	68.1/3.1	19.3	**61.2**
(0.05, 100, 5000, 3, 200, 0, 0.001)	51.5	**95.0**/4.3	55.7/3.4	18.6	58.8
(0.05, 100, 5000, 3, 100, 0, 0.001)	53.9	**94.8**/4.6	**86.2**/3.1	20.6	57.7
(0.05, 50, 5000, 3, 200, 0.1, 0)	55.2	**94.7**/4.7	66.9/3.2	24.0	60.1
(0.05, 50, 5000, 3, 50, 0.1, 0)	55.3	**94.6**/4.7	63.8/3.2	23.1	58.3
(0.10, 100, 5000, 3, 100, 0.20, 0.001)	57.3	90.1/4.7	**81.7**/3.2	14.7	46.6
(0.10, 200, 10000, 4, 100, 0, 0)	55.9	93.0/4.7	**77.7**/3.2	18.6	56.5
(0.05, 200, 5000, 3, 50, 0, 0)	54.4	94.5/4.6	**75.5**/3.3	23.7	**63.8**
(0.10, 50, 10000, 3, 100, 0, 0)	60.3	91.6/5.1	**74.9**/3.5	22.0	58.7
(0.10, 100, 5000, 3, 100, 0.15, 0)	60.7	88.7/5.0	27.4/4.4	**27.4**	50.5
(0.10, 200, 10000, 3, 50, 0, 0)	61.6	86.8/5.0	45.7/4.1	**26.4**	52.7
(0.05, 50, 5000, 4, 50, 0, 0)	58.4	92.2/4.9	52.4/3.1	**26.4**	54.4
(0.10, 200, 10000, 4, 100, 0.10, 0)	62.1	89.3/5.2	58.3/3.9	**26.2**	52.1
(0.05, 50, 10000, 3, 100 0, 0)	59.8	91.7/5.0	57.6/4.2	**25.8**	59.3
(0.10, 50, 10000, 3, 50, 0.10, 0)	57.6	92.8/4.8	68.4/3.6	20.6	**62.6**
(0.10, 200, 5000, 4, 100, 0.10, 0)	55.9	92.2/4.6	63.7/3.7	19.2	**61.1**
(0.05, 100, 5000, 3, 100, 0.15, 0)	58.3	91.7/4.9	65.6/3.7	19.0	**60.8**

Table 4.1: This table displays the performances of the model with various hyper-parameters. Each column contains the various metrics: accuracy (A), occurrence recall/precision (OR/P), large fire occurrence recall/precision (LFOR/P), scale accuracy (SA), and relaxed scale accuracy (RSA) are displayed. The format of the hyper-parameter column is (learning rate, epochs, batch size, hidden layers, neurons, dropout, regularization). The metric values are displayed as percentages. Bold values indicate a top five performer in its performance category.

occurrence precision values which are all below 10%. This means of all the locations that the model predicts a fire occurring, only less than 10% of the locations actually contained a fire.

This brings up the debate of whether it is better to have a model that has a higher recall or precision performance. Ideally, the perfect model would minimize both the number of false positives and false negatives. If a deployed model has a false negative and fails to predict a fire that occurs, then the cost of damage is much higher than if the model predicts a false positive, for the cost of being unprepared for a wildfire that occurs is greater than the cost of preparing for a wildfire that does not occur. Therefore, recall of predicting wildfires is a more important performance metric than precision. It also might be the case that low precision values are to be expected due to the nature of how wildfires ignite. As previously mentioned a wildfire requires an ignition source which usually comes from sources such as lightning strikes, meteors, and human activities. These sources are random and difficult to predict. So it might be the case that the model predicts that a location is currently susceptible to a wildfire because the data shows that it has high temperatures, low humidity, little rainfall, dry vegetation, etc., but if no ignition source occurs, then a wildfire will not occur. Thus, many of the false positives could be attributed to locations having the perfect climate and weather conditions for wildfires, but lacking the ignition source.

5 Conclusion

Looking ahead at the future implementation of artificial intelligence in assisting the fight of wildfires, the ability to predict a locations likelihood of a wildfire occurring is crucial for fighting and minimizing the damage of wildfires. If a location is known to be in high risk for a wildfire, then resources can be sent to that area such as firefighters to mitigate the risk or UAV's to survey the area. The future goal of research in this field is to have an artificial intelligence system that continuously processes live and historical data to determine a locations susceptibility to wildfires. This research explores this goal by attempting to predict the likelihood and scale of wildfires in California. This research showed promise in its ability to identify locations that are susceptible to fires indicated by the high occurrence recall values. However, the model has difficulty predicting the correct scale of an occurring wildfire indicated by the low performances in scale accuracy and relaxed scale accuracy. Furthermore, it tends to over predict the occurrence of wildfires that never occur indicated by the low occurrence precision and large fire occurrence precision values.

The future research in this field should look into how the size of a location impacts the prediction performance. This research used a location size of 10km x 10km, but perhaps a smaller or larger location size would be advantageous. New features such as elevation and slope should be explored. Since many wildfires are caused by human activity, perhaps human behavior in certain location can be captured as features. New model architectures such as recurrent neural networks (RNN) and convolutional neural networks (CNN) should be explored. A CNN would be great for capturing the vegetation density of surrounding locations to better account for the potential spread and scale of wildfires. Overall, this research shows potential in predicting a locations susceptibility to wildfires, and there is plenty of progress to be made in the future.

Bibliography

[1] National Park Services. (2016) Dangers of wildland fire. [Online]. Available: https://www.nps.gov/articles/wildland-fire-dangers-of-fire.htm

[2] National Interagency Fire Center. (2020) Total wildland fires and acres (1983-2020). [Online]. Available: https://www.nifc.gov/fire-information/statistics/wildfires

[3] Cal Fire. (2021) Stats and events. [Online]. Available: https://www.fire.ca.gov/stats-events/

[4] J. Cowan. (2021) How artificial intelligence is fighting wildfires. [Online]. Available: https://www.nytimes.com/2021/07/15/us/wildfires-artificial-intelligence.html

[5] P. Norman. (2021) Anticipating wildfires with machine learning tools. [Online]. Available: https://towardsdatascience.com/anticipating-wildfires-with-machine-learning-tools-5eb43b450d36

[6] U.S. Department of the Interior Indian Affairs. Understanding human vs. naturally occurring wildfires. [Online]. Available: https://www.bia.gov/bia/ots/dfwfm/bwfm/wildfire-prevention-and-education/home-bureau-indian-affairs-bia-trust-services-division-forestry-and-wildland-fire-management-branch

[7] K. Patel. (2021) What you need to know about how wildfires spread. [Online]. Available: https://www.washingtonpost.com/weather/2021/07/28/wildfires-spread-faq-west-explained/https://www.washingtonpost.com/weather/2021/07/28/wildfires-spread-faq-west-explained/

[8] B. Berwyn. (2018) How wildfires can affect climate change (and vice versa). [Online]. Available: https://insideclimatenews.org/news/23082018/extreme-wildfires-climate-change-global-warming-air-pollution-fire-management-black-carbon-co2/

[9] J. Mendenhall. (2019) Wildfire's impact on our environment. [Online]. Available: https://deq.utah.gov/communication/news/wildfires-impact-on-our-environment

[10] U.S. Forest Service. (2020) Fire effects on environment. [Online]. Available: \url{https://www.fs.usda.gov/pnw/page/fire-effects-environment#Fire\%20Ecology}

[11] California Department of Fish and Wildlife. (2021) Science: Wildfire impacts. [Online]. Available: https://wildlife.ca.gov/Science-Institute/Wildfire-Impacts

[12] American Lung Association. (2016) How wildfires affect our health. [Online]. Available: https://www.lung.org/blog/how-wildfires-affect-health

[13] Center for Disease Control and Prevention. (2020) Wildfires. [Online]. Available: https://www.cdc.gov/climateandhealth/effects/wildfires.htm

[14] Z. Jiao, Y. Zhang, J. Xin, L. Mu, Y. Yi, H. Liu, and D. Liu, "A deep learning based forest fire detection approach using uav and yolov3," in *2019 1st International Conference on Industrial Artificial Intelligence (IAI)*, 2019, pp. 1–5.

[15] P. Chanthiya and V. Kalaivani, "Forest fire detection on landsat images using support vector machine," *Concurrency and Computation: Practice and Experience*, vol. 33, no. 16, p. e6280, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.6280

[16] H. Liang, M. Zhang, and H. Wang, "A neural network model for wildfire scale prediction using meteorological factors," *IEEE Access*, vol. 7, pp. 176 746–176 755, 2019.

[17] R. Perumal and T. L. v. Zyl, "Comparison of recurrent neural network architectures for wildfire spread modelling," in *2020 International SAUPEC/RobMech/PRASA Conference*, 2020, pp. 1–6.

[18] G. Zhang, M. Wang, and K. Liu, "Forest fire susceptibility modeling using a convolutional neural network for yunnan province of china," *International Journal of Disaster Risk Science*, vol. 10, pp. 386–403, 2019.

[19] S. Natekar, S. Patil, A. Nair, and S. Roychowdhury, "Forest fire prediction using lstm," in *2021 2nd International Conference for Emerging Technology (INCET)*, 2021, pp. 1–5.

[20] J. Han, G. Kim, C. Lee, Y. Han, U. Hwang, and S. Kim, "Predictive models of fire via deep learning exploiting colorific variation," in *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2019, pp. 579–581.

[21] S. B. Avula, S. J. Badri, and G. Reddy P, "A novel forest fire detection system using fuzzy entropy optimized thresholding and stn-based cnn," in *2020 International Conference on COMmunication Systems NETworkS (COMSNETS)*, 2020, pp. 750–755.

[22] B. Arteaga, M. Diaz, and M. Jojoa, "Deep learning applied to forest fire detection," in *2020 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, 2020, pp. 1–6.

[23] M. Rahul, K. Shiva Saketh, A. Sanjeet, and N. Srinivas Naik, "Early detection of forest fire using deep learning," in *2020 IEEE REGION 10 CONFERENCE (TENCON)*, 2020, pp. 1136–1140.

[24] Y. Chen, Y. Zhang, J. Xin, G. Wang, L. Mu, Y. Yi, H. Liu, and D. Liu, "Uav image-based forest fire detection approach using convolutional neural network," in *2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 2019, pp. 2118–2123.

[25] E. Maeda, A. Formaggio, Y. Shimabukuro, G. Arcoverde, and M. Hansen, "Predicting forest fire in the brazilian amazon using modis imagery and artificial neural networks," *International Journal of Applied Earth Observation and Geoinformation*, vol. 11, pp. 265–272, 08 2009.

[26] M. Thornton, R. Shrestha, Y. Wei, P. Thornton, S. Kao, and B. Wilson, "Daymet: Daily surface weather data on a 1-km grid for north america, version 4," 2020. [Online]. Available: https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1840

[27] P. E. Thornton, S. W. Running, and M. A. White, "Generating surfaces of daily meteorological variables over large regions of complex terrain," *Journal of Hydrology*, vol. 190, no. 3, pp. 214–251, 1997, aggregate Description of Land-Atmosphere Interactions. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022169496031289

[28] P. E. Thornton, R. Shrestha, M. Thornton, S.-C. Kao, Y. Wei, and B. E. Wilson, "Gridded daily weather data for north america with comprehensive uncertainty quantification," *Scientific Data*, vol. 8, no. 190, 2021.

[29] K. Didan, "Mod13q1 modis/terra vegetation indices 16-day l3 global 250m sin grid v006 [data set]," 2015. [Online]. Available: https://doi.org/10.5067/MODIS/MOD13Q1.006

[30] GIS Geography. (2021) What is ndvi (normalized difference vegetation index)? [Online]. Available: https://gisgeography.com/ndvi-normalized-difference-vegetation-index/

[31] R. Kotwani. (2020) Deep ordinal logistic regression. [Online]. Available: https://towardsdatascience.com/deep-ordinal-logistic-regression-1afd0645e591