

University of Nevada, Reno

# **Spatiotemporal Machine Learning for Wildfire Spread Behavior Prediction**

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

by

Ahren Daniel Roth

Dr. Frederick C. Harris Jr., Advisor

May, 2024

© by Ahren Roth  
All Rights Reserved



THE GRADUATE SCHOOL

We recommend that the thesis  
prepared under our supervision by

**Ahren Daniel Roth**

entitled

**Spatiotemporal Machine Learning for Wildfire Spread Behavior Prediction**

be accepted in partial fulfillment of the  
requirements for the degree of

MASTER OF SCIENCE

Dr. Frederick C. Harris, Jr.,  
*Advisor*

Dr. Alireza Tavakkoli  
*Committee Member*

Dr. Jonathan Greenberg,  
*Graduate School Representative*

Markus Kemmelmeier, Ph.D., Dean,  
*Graduate School*

May, 2024

## Abstract

Wildfire is one of the most destructive ecological processes of our natural world. It is an integral part of our ecosystem, and the life cycle of many species depend on it for propagation. In order to manage fire, we must understand its behavior and properly model the condition of the surrounding ecosystem.

Predicting wildfire behavior is essential to effectively managing the wildland environment. Wildland management includes suppressing fires where necessary, promoting them where advantageous, all while protecting people, property, and resources. Fire behavior is influenced by the interplay of many factors. Topography, fuel type and load, wind, humidity, and temperature all effect the direction and rate of growth of wildland fires. Wildfires can create their own weather, amplifying the effects prior fire behavior have on future fire behavior. Spatiotemporal machine learning techniques may be utilized to model complex real-world fire behavior dynamics and produce accurate predictions. Anticipating wildfire behavior is the first step to effectively managing its effects.

Leveraging the ability of deep learning models to infer future fire behavior from past fire behavior is a novel approach to wildfire behavior prediction. Spatiotemporal modeling has proven effective at accurately predicting wildfire spread behavior days and weeks in advance. Current performance is limited by the availability of high-resolution, high frequency data. As remote-sensing data improves, and the availability of computational resources grows, temporal-spatial modeling of wildfire spread behavior will contribute more to effective wildland management practices.

## Dedication

I dedicate this thesis to all the hard-working human beings that make this world a great place to be. Thank you for inspiring me to action!

## Acknowledgments

I would like to thank my committee Dr. Frederick C. Harris, Jr., Dr. Alireza Tavakkoli, and Dr. Jonathan Greenberg for their invaluable guidance throughout my graduate studies.

# Contents

<b>Abstract</b>	<b>i</b>
<b>Dedication</b>	<b>ii</b>
<b>Acknowledgments</b>	<b>iii</b>
<b>List of Tables</b>	<b>vi</b>
<b>List of Figures</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Background and Related Work</b>	<b>3</b>
2.1 Wildfire . . . . .	3
2.2 Wildland Fuel . . . . .	3
2.2.1 Wildland Fuel Model . . . . .	5
2.2.2 Wildland Fuel Group . . . . .	5
2.2.3 Wildland Fuel Data . . . . .	6
2.2.4 Semi-Supervised Learning for Wildland Fuel Mapping . . . . .	8
2.3 Wildfire Weather . . . . .	12
2.4 Wildfire Behavior . . . . .	13
2.4.1 Wildfire Behavior Characteristics . . . . .	13
2.4.2 Wildfire Behavior Modeling . . . . .	15
2.4.3 Wildfire Behavior Data . . . . .	16
2.5 Relevant Machine Learning . . . . .	17
2.5.1 Artificial Neural Networks (ANN) . . . . .	18
2.5.2 Principle Component Analysis (PCA) . . . . .	19
2.5.3 Autoencoders . . . . .	20
2.5.4 Clustering . . . . .	23
2.6 Image Comparison . . . . .	23
<b>3 A Spatiotemporal Approach to Wildfire Spread Behavior Prediction</b>	<b>26</b>
3.1 Wildfire Spread Behavior Detection . . . . .	26
3.2 Wildland Environmental Conditions . . . . .	27
3.3 Spatiotemporal Machine Learning . . . . .	27

<b>4</b>	<b>Wildfire Data Set and Wildfire Spread Behavior Prediction Model</b>	<b>29</b>
4.1	Wildfire Data Set . . . . .	29
4.1.1	Fuel Map Data . . . . .	29
4.1.2	Slope Angle Data . . . . .	31
4.1.3	Thermal Imaging Data . . . . .	32
4.1.4	Weather Data . . . . .	33
4.1.5	Constructing a Wildfire Event Time Instance Sequence . . . . .	34
4.2	Wildfire Spread Behavior Prediction Model . . . . .	35
4.2.1	Constructing the Model with TensorFlow . . . . .	35
4.2.2	Training the Model . . . . .	35
4.2.3	Making a Prediction with the Model . . . . .	36
<b>5</b>	<b>Experiments</b>	<b>39</b>
5.1	Simulated Data . . . . .	39
5.1.1	Simple Flaming Front . . . . .	39
5.1.2	Fuel Map . . . . .	39
5.1.3	Accelerating Factors . . . . .	40
5.1.4	Multiple-Step Predictions . . . . .	40
5.2	Real-World Data . . . . .	41
5.2.1	Single-Step Predictions . . . . .	44
5.2.2	Multi-Step Predictions . . . . .	44
<b>6</b>	<b>Results and Analysis</b>	<b>46</b>
6.1	Simulated Data . . . . .	46
6.2	Real-World Data . . . . .	46
6.3	Dixie Fire Case Study . . . . .	49
6.3.1	Initial Growth Phase . . . . .	50
6.3.2	Later Growth Phase . . . . .	51
<b>7</b>	<b>Conclusions and Future Work</b>	<b>56</b>
7.1	Conclusions . . . . .	56
7.2	Future Work . . . . .	57
	<b>References</b>	<b>58</b>

# List of Tables

2.1	Semi-supervised fuel mapping cluster inclusion and exclusion scores. A higher inclusion score indicates a more prevalent fuel type representation within a cluster. A lower exclusion score indicates increased uniqueness between cluster representations. . . . .	11
6.1	Simulated training data prediction comparison results . . . . .	46
6.2	Real-world training data single time-step prediction comparison results	47
6.3	Real-world training data multiple time-step prediction comparison results	49

# List of Figures

2.1	Caldor Fire as it threatens the city of South Lake Tahoe in the distance	4
2.2	Wildland fuel examples: mixed shrub and grassland [3]	4
2.3	13 fuel model loading descriptions [3]	6
2.4	Chart of fuel load and size by fuel group[3]	7
2.5	Graphs of fuel depth and orientation by fuel group[3]	7
2.6	Fuel type representations from NAIP data [24]	8
2.7	LANDFIRE fuel map of the Northwestern United States representing the 13 Anderson Fire Behavior Fuel Models [3, 22]	9
2.8	Original NAIP image region (left) and semi-supervised fuel map representation of Prosser reservoir region (right). Brush-Brown, Sand-Beige, Pine Forest-Green, Water-Blue, Unknown Type-Grey, Dry Grass-Gold [24]	10
2.9	Semi-supervised fuel mapping pipeline [24]	11
2.10	Pyrocumulus clouds [12, 17]	12
2.11	The Fire Behavior Triangle [19]	14
2.12	Flaming Front Ellipse [26]	14
2.13	The Fire Modeling Pentagon [26]	16
2.14	A comparison of daily fire spread mapped by 1km Aqua/MODIS (left), 750m VIIRS (center) and 375m VIIRS (right) data at the Taim Ecological Reserve in southern Brazil. The color-coded legend provides day and time of data acquisition. [17]	17
2.15	Topology of a fully connected artificial neural network[6]	18
2.16	Topology of a convolutional neural network[5]	19
2.17	Topology of a long-short term memory node[13]	20
2.18	Example of PCA and dimensionality reduction[4]	21
2.19	Topology of a fully-connected autoencoder[25]	22
2.20	Comaprison of (A) GMM probablistic soft boundaries vs (B) K-means hard boundaries[32]	23
2.21	Example of RGB image conversion to histogram[2]	25
3.1	Wildfire Spread Behavior Prediction Model machine learning pipeline	28

4.1	Screenshot from the LANDFIRE Data Distribution site [23] . . . . .	30
4.2	Screenshot of a binary light fuels map from the ArcGIS Pro mapping tool [9] . . . . .	31
4.3	Screenshot of thermal imaging data from the FIRMS US and Canada map [18] . . . . .	32
4.4	The National Weather Service Red Flag Weather Matrix [30] . . . . .	33
4.5	Wildfire Spread Behavior Prediction Model . . . . .	36
4.6	Confidence value threshold example . . . . .	37
5.1	Simulated simple flaming front input and prediction comparison with ground truth . . . . .	40
5.2	Simulated flaming front with a fuel map input and prediction comparison with ground truth . . . . .	41
5.3	Simulation with fuel map and accelerator input and prediction comparison with ground truth . . . . .	42
5.4	Single and multiple step prediction comparison . . . . .	43
5.5	Single time-step hot-spot prediction vs ground truth . . . . .	44
5.6	Single time-step burn area prediction vs ground truth . . . . .	45
5.7	Multiple time-step hot-spot prediction vs ground truth . . . . .	45
5.8	Multiple time-step burn area prediction vs ground truth . . . . .	45
6.1	Simulated training data loss plot . . . . .	47
6.2	Real-world training data loss plot . . . . .	48
6.3	Average MSE/SSIM value vs number of time steps for hotspot prediction . . . . .	49
6.4	Average MSE/SSIM value vs number of time steps for burnt area prediction . . . . .	50
6.5	Satellite imagery of the impacted region prior to the Dixie Fire with the initial detection data overlaid [14] . . . . .	51
6.6	Predicted Dixie fire hot-spot detection map from July 17th to July 25th . . . . .	52
6.7	Dixie fire ground truth hot-spot detection map from July 17th to July 25th . . . . .	53
6.8	Predicted Dixie fire hot-spot detection map from August 7th to August 11th . . . . .	54
6.9	Dixie fire ground truth hot-spot detection map from August 7th to August 11th . . . . .	55

# Chapter 1

## Introduction

Wildland firefighting is best accomplished well before a fire even starts. Responsible forest management, by developing and analyzing accurate forest models, is key to minimizing the ecological and economic impacts of wildfire[11]. Fire is a natural process that many environments are well adapted to. Excessive fire suppression can lead to excessive fuel loads, which may enable larger and more destructive fires to develop. These catastrophic fires can do irreversible damage to environments well adapted to more frequent and less destructive wildfires. Ecosystems made vulnerable by global warming are most at risk from the potential hazards posed by the unnaturally large and destructive fires made possible by the compounded effects of climate change and ineffective forest management[1].

Data relevant to wildfire behavior prediction is readily available in standardized formats across North America. Fuel maps, topography and atmospheric conditions are some of the most important factors in the fire behavior equation. Satellite based radiometer data provides a 2-dimensional representation of wildland fire behavior on a macroscopic level. Accurately modeling the wildland environment with the behavior characteristics of past fires, is the basis upon which wildfire behavior prediction may be built.

Predicting wildland fire behavior is essential to strategizing fire control and suppression, as well as setting priorities for effective forest management. Traditional methods of wildland fire prediction include physics based and empirical models[28]. Physics models utilize basic known behaviors on either the macroscopic or molecular

level to compute fire propagation within a system. Empirical models rely on inference from past behavior to predict future behavior. Traditional empirical models are produced through statistical methods that return localized probabilities of fire propagation over a period of time. Using modern machine learning models, and the vast amounts of data available relevant to wildfire behavior prediction, it is possible to improve upon traditional empirical models.

Spatiotemporal techniques leverage the power to build effective models in respect of space and time dimensions[27]. Convolutional Neural Networks (CNN) are often the spatial component of these models; they are useful for extracting relevant features from input images. Long-Short Term Memory (LSTM) networks may be utilized for the temporal component, which is used for making predictions on sequence data. Together, these models allow for the prediction of wildland fire behavior with greater granularity than traditional empirical methods. These models answer not only if and when, but how wildfire will propagate through an environment.

This thesis details wildfire behavior prediction utilizing spatiotemporal machine learning techniques, trained on real-world data. Following the Introduction in Chapter 1, Chapter 2: Background and Related Work covers relevant aspects of modeling the wildland environment, wildfire behavior, and machine learning. Chapter 3: A Spatiotemporal Approach to Wildfire Spread Behavior Prediction, presents the spatiotemporal machine learning model and training regimen. Construction of the Wildfire Data Set and Wildfire Spread Behavior Prediction Model is covered in Chapter 4, while Chapter 5 details the Experiments. Chapter 6 contains the Results and Analysis including a case study of the Dixie fire, and finally Chapter 7 includes the Conclusions and Future Work suggestions.

# Chapter 2

## Background and Related Work

### 2.1 Wildfire

A fire burning uncontrolled in a predominately natural environment is known as a wildfire[26]. Wildfire is a natural and essential part of many ecosystems. Wildfires can burn in and be influenced by urban environments, but the primary fuel loading and means of propagation is by natural materials. Wildfire behavior is characterized by the contributing factors of the environment. Region, fuel loading, topography, moisture content and fire weather influence fire intensity, direction, and rate of spread of the wildfire. Wildfires have extreme effects on the environments and economies where they burn, which may have generational impacts on the flora, fauna, and communities that reside there. The Caldor fire, shown in Figure 2.1, burnt 221,835 acres, destroying more than a thousand structures and displacing families in the process [10]. Anticipating and predicting wildfire behavior is essential to minimizing negative impacts while retaining and encouraging the benefits. Fire behavior in the wildland environment is both dynamic and complex, making accurate predictions a difficult problem to solve.

### 2.2 Wildland Fuel

Wildland fuels are an essential element in the fire triangle. These fuels consist of grasses, shrubs, timber, and slash[3]. The images in Figure 2.2 depict wildland with shrub and grassland fuel types. When combined with oxygen and heat, wildland



Figure 2.1: Caldor Fire as it threatens the city of South Lake Tahoe in the distance fuels may ignite. Provided favorable conditions, a fire may propagate rapidly and destructively.



Figure 2.2: Wildland fuel examples: mixed shrub and grassland [3]

**Fuel Size:** Fuel size correlates directly with the ratio of mass to surface area. Fuels of larger size have a higher mass to surface area ratio, and generally dry out and burn slower than those with a lower ratio. Small fuels have a very low ratio and are often have faster spread rates.

**Fuel Orientation:** Fuels in the wildland environment are commonly vertically oriented. Most grasses, shrubs and trees grow upward towards the sunlight. Horizontally oriented fuels are often the result of natural or man-made events, such as a previous season's heavy snow fall, or logging slash.

**Fuel Moisture Content:** The moisture content of wildland fuels is an important factor to combustion. Dry fuels are more likely to combust and fires in dry fuel move much faster than those in fuels that contain more moisture. Fuel moisture content is equal to the ratio of the wet fuel mass divided by the dried mass[15].

### 2.2.1 Wildland Fuel Model

Several classification methods exist for grouping like fuels and defining their combustion characteristics. The advent of mathematical models for predicting wildland fire behavior has led to the development of fuel models that define a standardized set of properties that segment the wildland fuel distribution. The chart in Figure 2.3 provides the loading descriptions for the 13 Anderson fire behavior fuel models. Wildland fire behavior differences amongst fuel models is driven by two primary factors, fuel load and particle size distribution[3]. These factors are relevant to predicting whether a fire will ignite, the rate of spread, and its intensity.

### 2.2.2 Wildland Fuel Group

Fuel groups collect wildland fuel models with similar physical characteristics. These groups include: grasses, shrubs, timber litter, and logging slash. Grasses and shrubs are considered light and flashy fuels, while timber litter and logging slash are considered heavy fuels. Fuel orientation determines the rate of fuel load increase with depth. Vertically oriented fuels, grasses and shrubs, increase in depth rapidly with load. Horizontally oriented fuels, timber litter and logging slash, slowly increase in depth with increased load[3]. The chart in Figure 2.4 depicts the load distribution of the four Anderson fuel groups by size and proportion of the constituent parts. The

Fuel model	Typical fuel complex	Fuel loading				Fuel bed depth	Moisture of extinction dead fuels
		1 hour	10 hours	100 hours	Live		
		-----Tons/acre-----				<i>Feet</i>	<i>Percent</i>
	<b>Grass and grass-dominated</b>						
1	Short grass (1 foot)	0.74	0.00	0.00	0.00	1.0	12
2	Timber (grass and understory)	2.00	1.00	.50	.50	1.0	15
3	Tall grass (2.5 feet)	3.01	.00	.00	.00	2.5	25
	<b>Chaparral and shrub fields</b>						
4	Chaparral (6 feet)	5.01	4.01	2.00	5.01	6.0	20
5	Brush (2 feet)	1.00	.50	.00	2.00	2.0	20
6	Dormant brush, hardwood slash	1.50	2.50	2.00	.00	2.5	25
7	Southern rough	1.13	1.87	1.50	.37	2.5	40
	<b>Timber litter</b>						
8	Closed timber litter	1.50	1.00	2.50	0.00	0.2	30
9	Hardwood litter	2.92	.41	.15	.00	.2	25
10	Timber (litter and understory)	3.01	2.00	5.01	2.00	1.0	25
	<b>Slash</b>						
11	Light logging slash	1.50	4.51	5.51	0.00	1.0	15
12	Medium logging slash	4.01	14.03	16.53	.00	2.3	20
13	Heavy logging slash	7.01	23.04	28.05	.00	3.0	25

Figure 2.3: 13 fuel model loading descriptions [3]

graph in Figure 2.5 provides a graphical representation of the relationships between fuel depth and orientation for each fuel group. A wildland fuel group may be used to generalize wildland fire behavior as an average of the fuel models that it is composed of.

### 2.2.3 Wildland Fuel Data

Data relevant to wildland fuels are available from a number of sources and in a variety of representations. The LANDFIRE[22] program has produced fuel maps for the whole of the continental United States. High resolution imagery is also available through the National Agricultural Imagery Program.

**National Agricultural Imagery Program:** The National Agricultural Imagery Program[24] captures and provides digital ortho-photography on a 3-year update cycle. Current and archived images are available via the Earth Explorer user interface and associate Bulk Download tool. The images are large, approximately 12000 x 12000 px in size with an approximate resolution of 40 cm x 40 cm per pixel. The images are RGB and some include a 4th infrared channel. The images in Figure 2.6

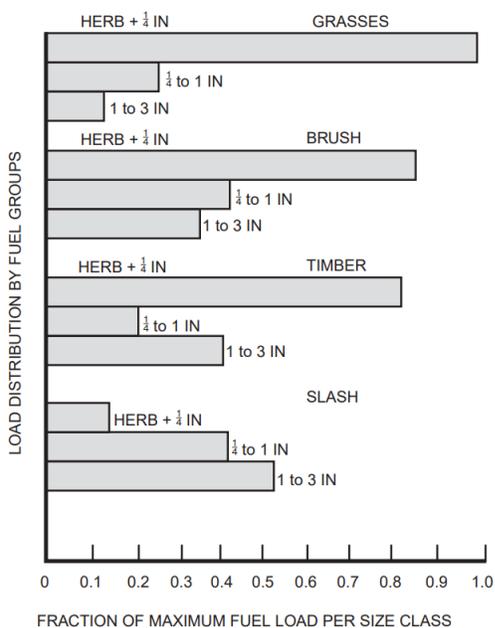


Figure 2.4: Chart of fuel load and size by fuel group[3]

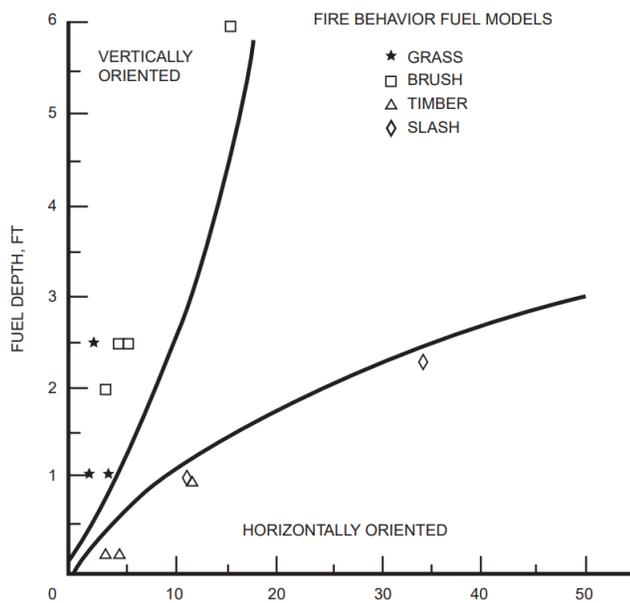


Figure 2.5: Graphs of fuel depth and orientation by fuel group[3]

represent a few of the fuel types distinguishable from the NAIP data set.



Figure 2.6: Fuel type representations from NAIP data [24]

**LANDFIRE:** LANDFIRE[22] is a suite of spatial data layers relevant to land management for wildfire risk reduction. The layers cover the whole of the United States and its surrounding territories. LANDFIRE consists of 20+ data products including fuel, disturbance, and topographic maps. The 13 Anderson Fire Behavior Fuel Models for the Northwestern region of the United States are shown in Figure 2.7. Spatial data is offered at a native 30-meter resolution, but suggested application scale is generally much larger. LANDFIRE data may be augmented with local knowledge to improve the pertinence to a particular application. The LANDFIRE program was established to provide consistent data, focused on effective fire management in the wildland environment, across the whole of the United States.

### 2.2.4 Semi-Supervised Learning for Wildland Fuel Mapping

Fuel types vary according to region and are strongly affected by atmospheric conditions. Grass may dry out and burn a week after a long period of rainfall, but larger

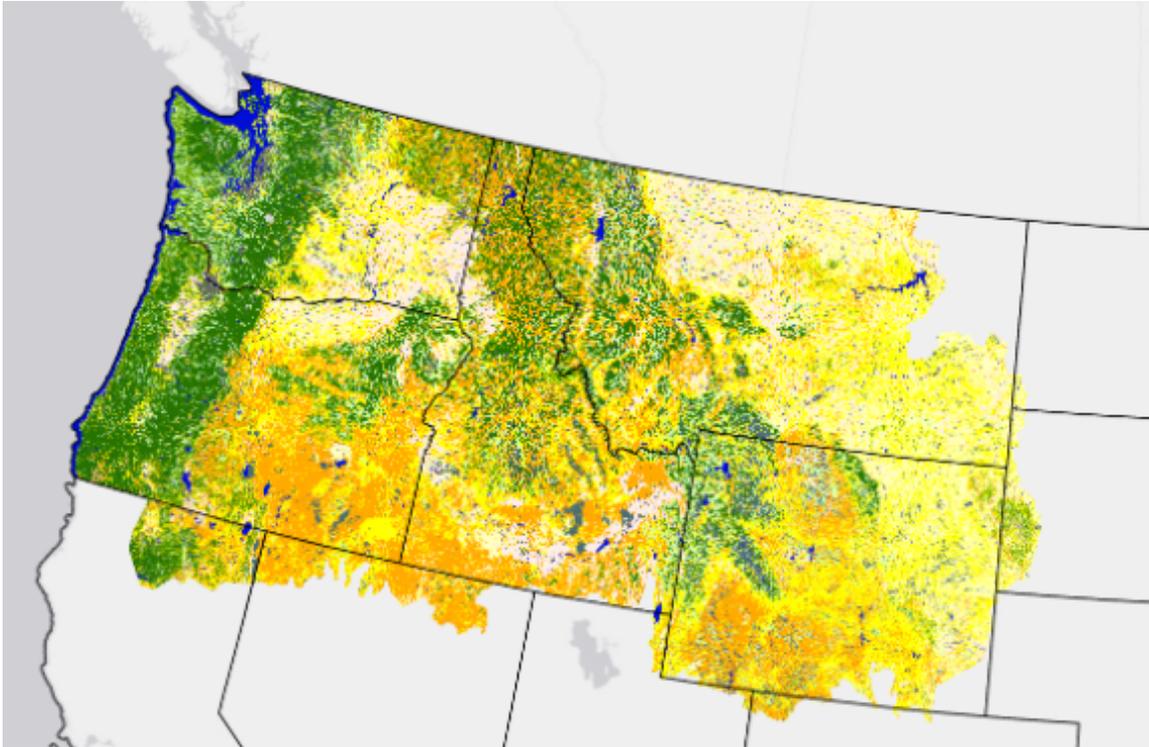


Figure 2.7: LANDFIRE fuel map of the Northwestern United States representing the 13 Anderson Fire Behavior Fuel Models [3, 22]

branches and stumps may take much longer to be capable of sustaining fire under normal conditions. Fuel maps are an important construct for predicting the ability of a fire to spread in a specific region. The availability of high-resolution satellite imagery, along with the application of machine learning methods, allows for the construction of high accuracy fuel maps that encompass a broad array of fuel types. Figure 2.8 provides an example input and resultant output from a semi-supervised fuel mapping model. In reading order, the fuel types may describes as: brush, dense pine forest, grass, sand, sparse pine forest, and water.

**Semi-Supervised Fuel Mapping Method:** Fuel type mapping utilizing semi-supervised methods has several advantages. Due to the diverse, expansive, and dynamic environment in which wildfire exists, it is important for a mapping system to be quick, thorough, and flexible to change. Autoencoders allow for the composition

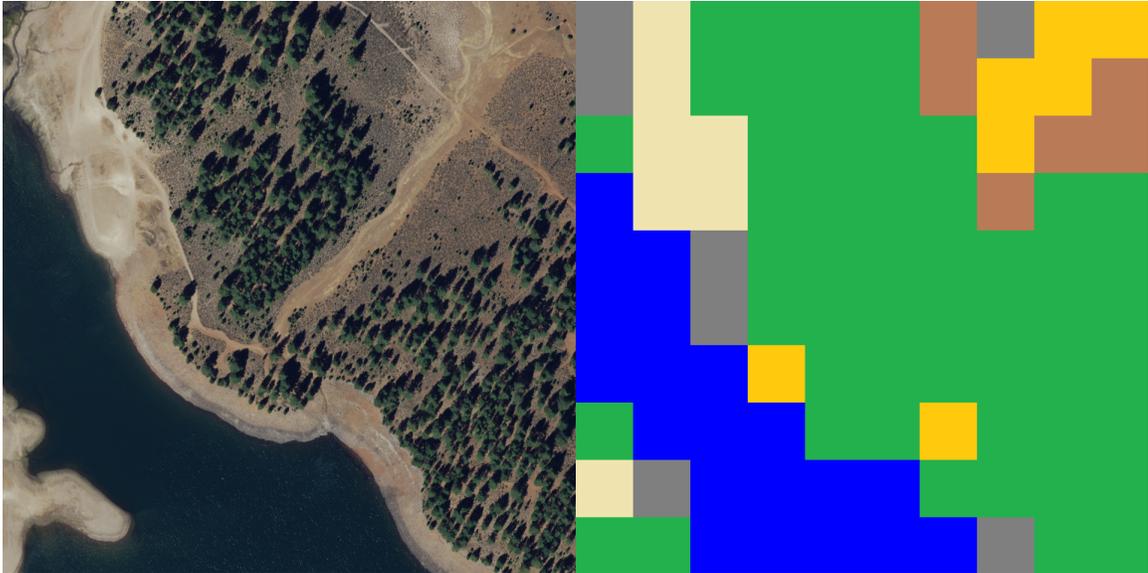


Figure 2.8: Original NAIP image region (left) and semi-supervised fuel map representation of Prosser reservoir region (right). Brush-Brown, Sand-Beige, Pine Forest-Green, Water-Blue, Unknown Type-Grey, Dry Grass-Gold [24]

of reduced dimension encodings that are composed of the most salient features relevant fuel type map construction. Gaussian mixture models are an effective means of clustering the encodings to group like fuel types together. The resultant clusters are representative of a particular fuel type that is discernible utilizing the input to the model. Figure 2.9 graphically depicts the proposed semi-supervised fuel mapping method.

**Semi-Supervised Fuel Mapping Findings:** Autoencoders are capable of creating code representations of images that are differentiable by fuel type using Gaussian Mixture Models. Utilizing a semi-supervised approach for fuel type mapping greatly reduces the amount of time and manpower necessary to produce new maps and update existing ones. This approach allows for the rapid creation of high-resolution fuel maps that cover vast areas. Dissimilar fuel types were found to be more easily separable than fuel types that are close in appearance. The features the autencoder creates a reduced dimensionality representation from are not necessarily the most salient features for wildland fuel type classification. It was found that the majority of

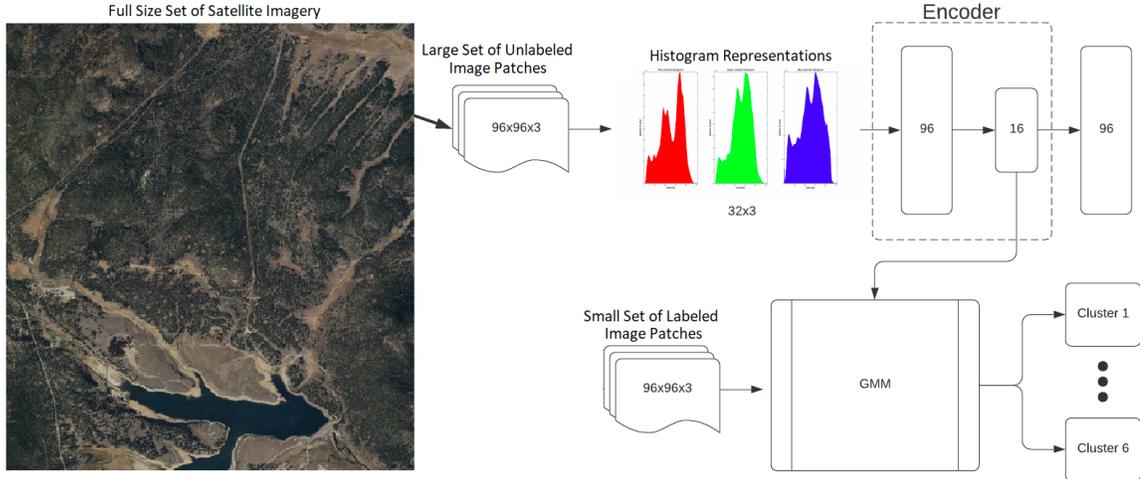


Figure 2.9: Semi-supervised fuel mapping pipeline [24]

the code representation was allocated to artifact localization within the image patch, which is not relevant to fuel type classification. Utilizing the histogram representation as an input to the encoder resolved this issue but reduced the ability of network to differentiate fuel types with similar color profiles. Table 2.1 provides the results from several experiments utilizing various auto-encoder architecture. Future work will focus on modifying a convolutional autoencoder pipeline, to reduce the presence of feature localization within the code representation.

Table 2.1: Semi-supervised fuel mapping cluster inclusion and exclusion scores. A higher inclusion score indicates a more prevalent fuel type representation within a cluster. A lower exclusion score indicates increased uniqueness between cluster representations.

MODEL	INCLUSION	EXCLUSION
FULLY CONNECTED(FC) 32	71.6%	33.3%
FC 64	80.0%	16.7%
FC 64 WITH NOISE	80.3%	50.0%
FC 64 WITH REGULARIZER	76.7%	50.0%
CONVOLUTIONAL	78.3%	33.3%
HISTOGRAM REPRESENTATION	93.3%	16.7%

## 2.3 Wildfire Weather

Weather plays a huge role in the likelihood of occurrence and impact of wildfire [21]. Storm clouds can develop lightning strikes to ignite fires. High winds and low humidity create optimal conditions for wildfire spread. Wildfires influence the weather; small fires can create strong up-slope breezes accelerating propagation. Large wildfires can evaporate large amounts of water and transport it vertically to form pyrocumulus clouds. Two examples of pyrocumulus clouds from different vantage points are shown in Figure 2.10. These clouds are capable of producing lightning strikes miles ahead of the fire from which they originated.



Figure 2.10: Pyrocumulus clouds [12, 17]

**Lightning:** Cloud to ground lightning strikes are the most common natural cause of wildfire ignition. Lightning is the result of an electrical discharge when opposing charges build up between two bodies. In the case of cloud to ground lightning, ice formation within a cloud builds charge until the electrical resistance of the air between cloud and ground is overcome. At which point, the excess charge is discharged to ground and the electrical resistance creates heat that can spark a fire. Lightning complexes may form thousands of strikes over a broad area, in a short period of time. Dry lightning is the occurrence of lightning without accompanying precipitation, and when paired with favorable surface conditions, is an extremely effective mechanism for starting large destructive wildfires.

**Wind:** Wind is capable of transporting heat and burning material well beyond the flaming front of a wildfire. Convective currents driven by wind can accelerate fuel drying, making materials in the fire path more susceptible to rapid combustion. Many of the most destructive wildfires in history have been quick, wind driven fires that burned large swaths of land in a very short period of time.

**Humidity:** Humidity is a measure of the water vapor concentration in the air. Relative humidity factors in the holding capacity of the air at a given temperature and is expressed as a percentage. Low humidity can be used as an indicator for favorable fire conditions.

**Precipitation:** Precipitation is the condensation of water vapor that falls from the clouds. It comes in the form of rain, sleet, snow, and hail. Precipitation is often preceded by a rise in relative humidity and has generally negative effects on wildfire spread.

**Temperature:** Temperature is a measure of molecular thermal vibration. As temperatures rise, materials can vaporize and are more prone to flaming combustion.

## 2.4 Wildfire Behavior

Wildland fire behavior describes the dynamic characteristics and effects of wildfire. Wildfire behavior is governed by fuels, weather, and topography. Together, these factors form the Fire Behavior Triangle seen in Figure 2.11.

### 2.4.1 Wildfire Behavior Characteristics

The quantifiable characteristics that define wildfire behavior are flame front rate of spread, flame length, fireline intensity and heat released per unit area. These characteristics are useful measures for objectively anticipating fire growth and impact[26]. A diagram of the various spread regions of a wildfire is depicted in Figure 2.12.

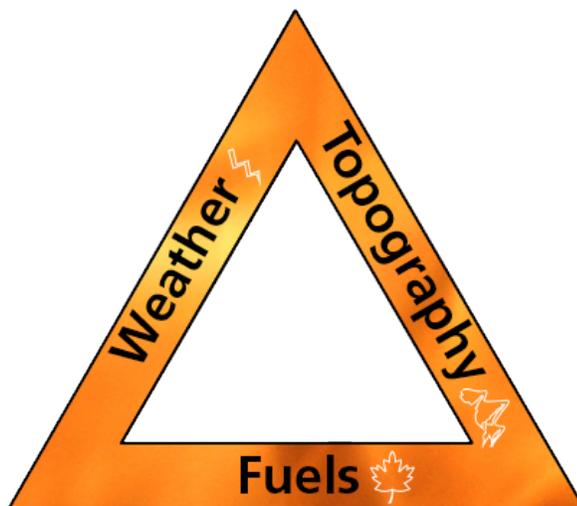


Figure 2.11: The Fire Behavior Triangle [19]

**Flame Front Rate of Spread (ROS):** The flaming front of a fire is at the leading edge of the fire, made up of flaming combustion into un-burned fuels. The flame front completely encircles the fire and is composed of subsections including: head, flank, hank, and rear. ROS for each section is often defined as a function of the ROS at the head of the fire.

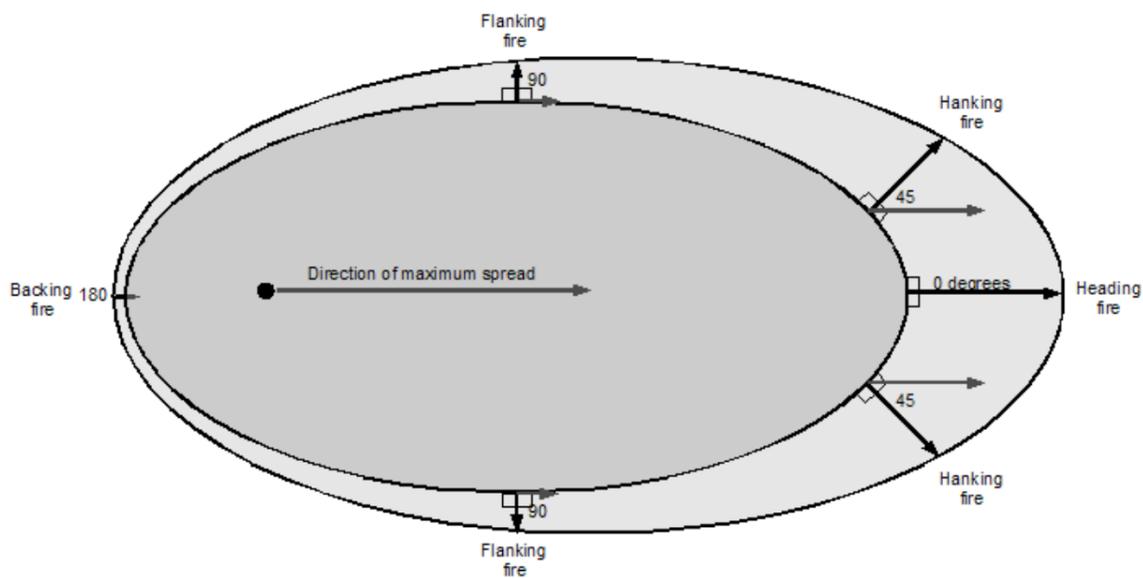


Figure 2.12: Flaming Front Ellipse [26]

**Flame Length (FL):** Flame length is a measure of the visible flaming combustion. Alternatively, flame height may be used.

**Fireline Intensity (FLI):** The amount of heat released per unit length of the flaming front.

**Heat Released Per Unit Area (HPA):** The amount of heat released per unit area respective of flaming front depth.

## 2.4.2 Wildfire Behavior Modeling

Wildfires are dynamic and complex; their behavior being driven by a vast array of confounded factors. The five major influences on wildfire behavior modeling are depicted in the fire modeling pentagon shown in Figure 2.13 [26]. Modeling wildfire behavior is advantageous to individuals, organizations, and communities that reside near, or rely on the wildland. Accurate wildfire models are an integral component of effective forest management and wildfire incident response. Many approaches to modeling wildfire behavior exist, but they can primarily be separated into two classes: empirical and physical based modeling, with some methods composed of a hybridization of the two.

**Empirical Modeling:** Models that rely on data from prior wildland fire events are known as empirical models. These models consist of a set of inputs that can be mapped to a set of outputs to recreate the particulars of a wildfire event through experience, without explicitly defining the relationships of inputs to outputs and the nature of the mapping. Empirical data is often used in fire spread equations like those developed by Rothermel and others[3]. Lab and control experiments were used to measure and quantify experiential data relevant to fire spread. The factors derived from this data are inputs into a set of equations developed for calculating flaming front rate of spread (ROS), heat per unit area (HPA), flame length (FL) and fire line intensity (FLI).

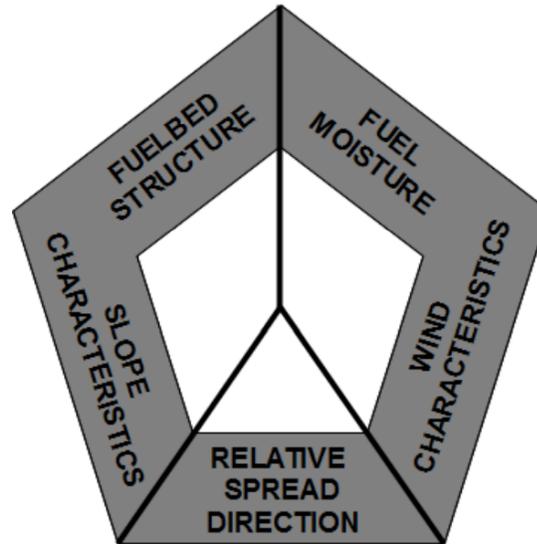


Figure 2.13: The Fire Modeling Pentagon [26]

**Physics-Based Modeling:** Physics based models make use of fundamental physical principles to calculate heat transfer and its effects on surrounding materials and the atmosphere. These models allow for greater granularity and more deterministic behavior than empirical methods, at the cost of more input requirements and greater computational complexity than that of empirical methods. For these reasons, most physics-based models are not applicable in the field due to their inability to produce predictions faster than real-time.

### 2.4.3 Wildfire Behavior Data

Data relevant to wildfire behavior has traditionally been condensed into equations and models for simulation purposes. Satellite-based spectrometer data [17] may be used to approximate wildfire spread behavior, as seen in Figure 2.14.

**MODIS:** Moderate Resolution Imaging Spectroradiometer (MODIS) instrumentation is aboard the Terra and Aqua satellites. The two of the satellites collectively view the entire Earth's surface every 1 to 2 days. They acquire data in 36 spectral bands. A subset of these bands are useful for detecting radiation associated with wildfires. The MODIS instrumentation produces data relevant to wildfire detection

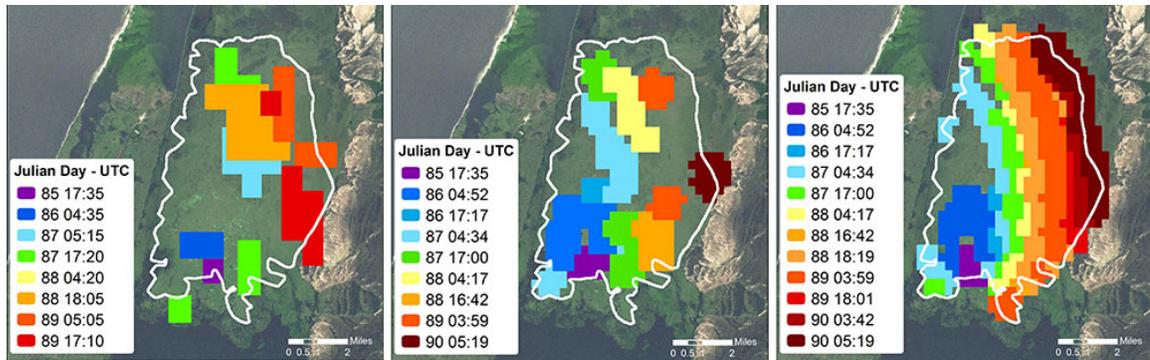


Figure 2.14: A comparison of daily fire spread mapped by 1km Aqua/MODIS (left), 750m VIIRS (center) and 375m VIIRS (right) data at the Taim Ecological Reserve in southern Brazil. The color-coded legend provides day and time of data acquisition. [17]

at a maximum resolution of 1 km.

**VIIRS:** Visible Infrared Imaging Radiometer Suite (VIIRS) instrumentation is aboard two satellites, the Suomi National Polar-orbiting Partnership (Suomi NPP) and the National Oceanic and Atmospheric Administration-20 (NOAA-20). These satellites orbit enables them to observe the entire Earth’s surface twice each day. The instrumentation collects data on 22 spectral bands at a maximum resolution of 375 meters. A subset of these bands contain information relevant to wildfire detection.

## 2.5 Relevant Machine Learning

Machine learning is the application of algorithms on data to create and refine models that perform a task without being explicitly programmed. There are many different machine learning algorithms and models, each with characteristics that make them more or less suitable for specific tasks. Machine learning techniques relevant to wildfire spread behavior prediction, and semi-supervised fuel mapping are covered in this section.

### 2.5.1 Artificial Neural Networks (ANN)

Artificial neural networks are loosely modeled on the biological brain and its processes. They consist of one or more layers, each containing a set of nodes. An example ANN is shown in Figure 2.15. Training data is passed through the network of layers, and performance feedback updates are returned through the process of back-propagation. As the network is trained on more data, it becomes better at the task.

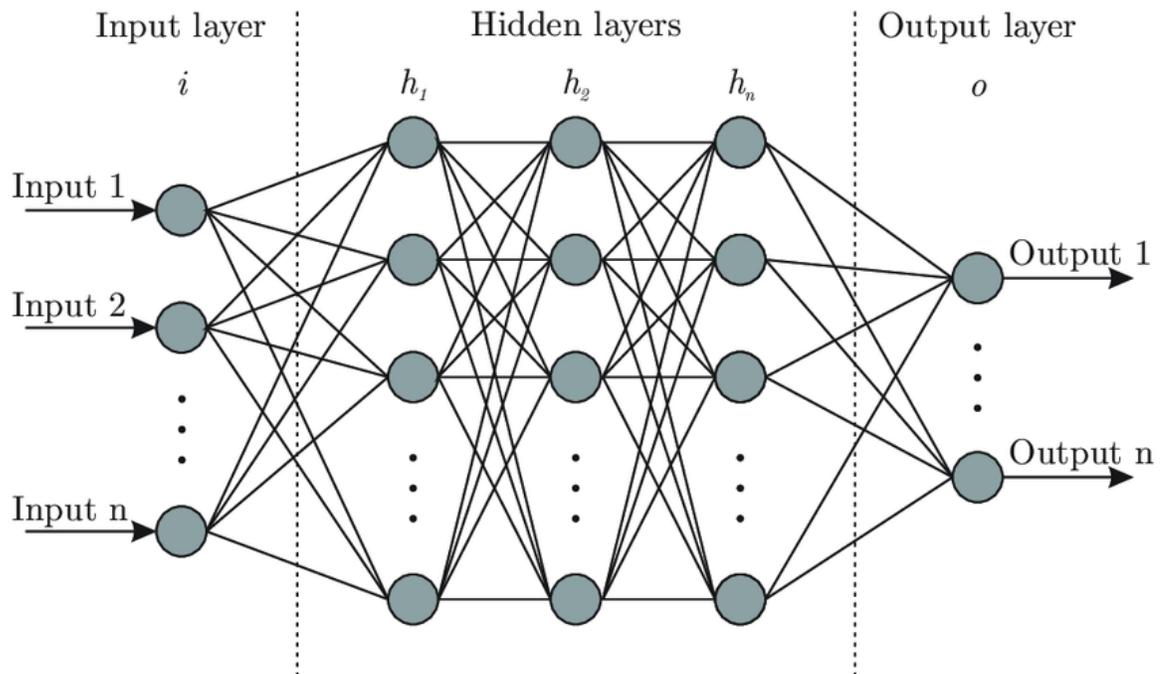


Figure 2.15: Topology of a fully connected artificial neural network[6]

**Convolutional Neural Network (CNN):** Convolutions enable a neural network to work on data of higher dimension. Spatial features within the data are generalized allowing CNNs to perform better than traditional NNs on multidimensional data, such as RGB images. An example CNN architecture is depicted in Figure 2.16.

**Recurrent Neural Network (RNN):** Neural networks with nodes capable of creating a cycle are known as RNNs. These cycles enable the networks to store

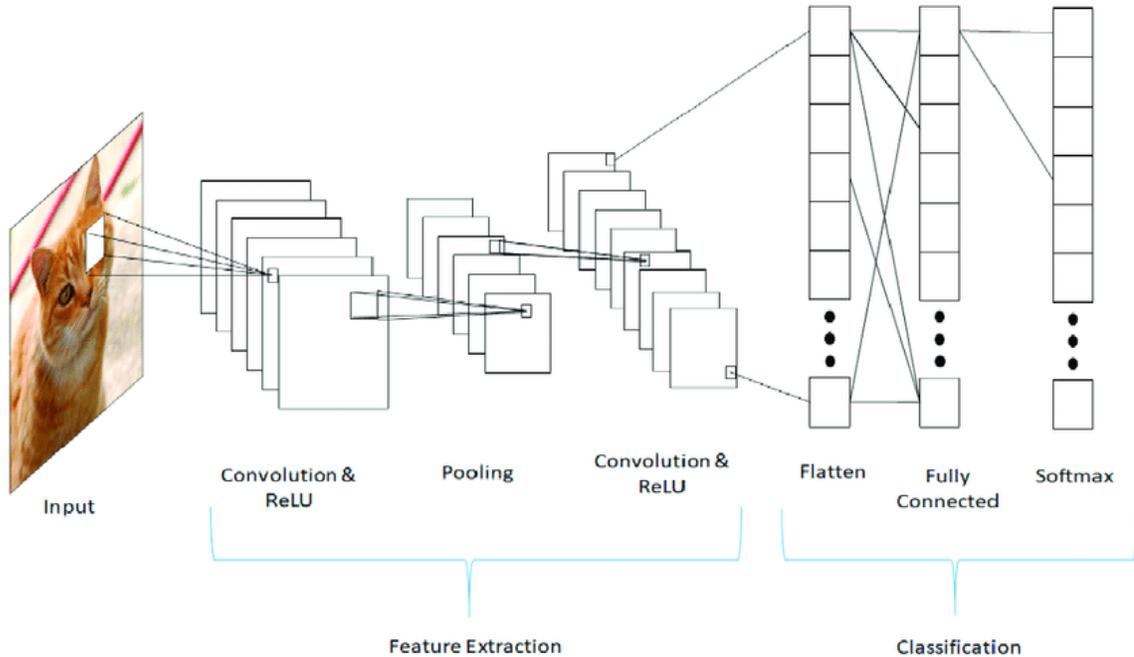


Figure 2.16: Topology of a convolutional neural network[5]

temporal components of the data.

**Long Short Term Memory (LSTM):** RNN performance can be poor on long sequences of data where vanishing gradients become a problem. LSTMs address this through a system of cell states and gates, shown in Figure 2.17, which allows the network to selectively remember important aspects of the training data.

**ConvLSTM:** ConvLSTM models are designed for spatiotemporal machine learning tasks. Traditional LSTM architecture is augmented with convolutions for multi-dimensional inputs and connections, making them particularly well suited for prediction tasks on series of images [8].

## 2.5.2 Principle Component Analysis (PCA)

PCA is the application of statistical methods to reduce the dimensionality of a dataset. It is often applied to data sets with many features and few samples. During PCA, features may be combined and are converted to principle components, which form

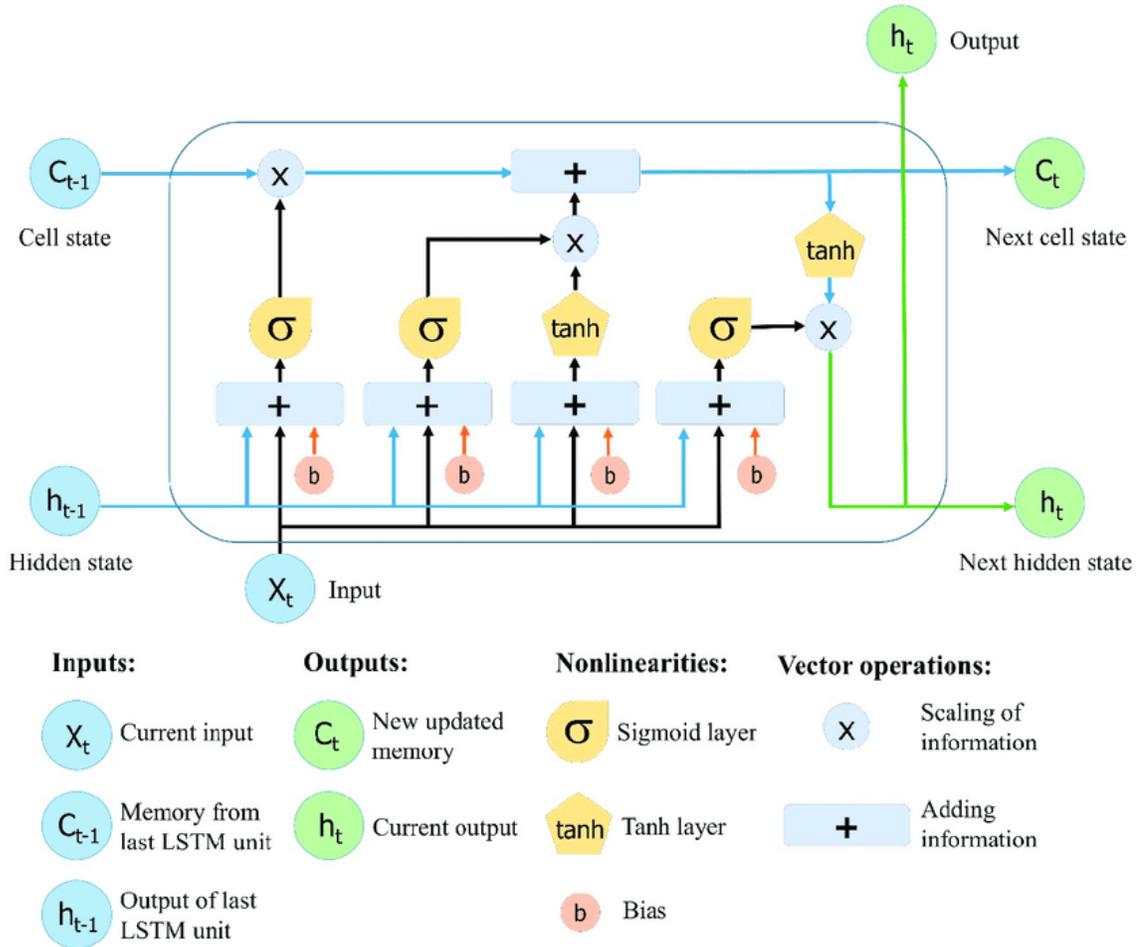


Figure 2.17: Topology of a long-short term memory node[13]

a new axis system that maximizes the variance in the data, while minimizing the number of significant axis. The process flow of PCA and dimensionality reduction is depicted in Figure 2.18.

### 2.5.3 Autoencoders

An autoencoder is a pair of neural networks that are used to create a code representation of an input and then create an approximation of the original input form the code representation. They are commonly used in data compression applications where the code representation is of reduced size to the input. Training the autoencoder requires the implementation of an encoder-decoder pair. Images are passed to

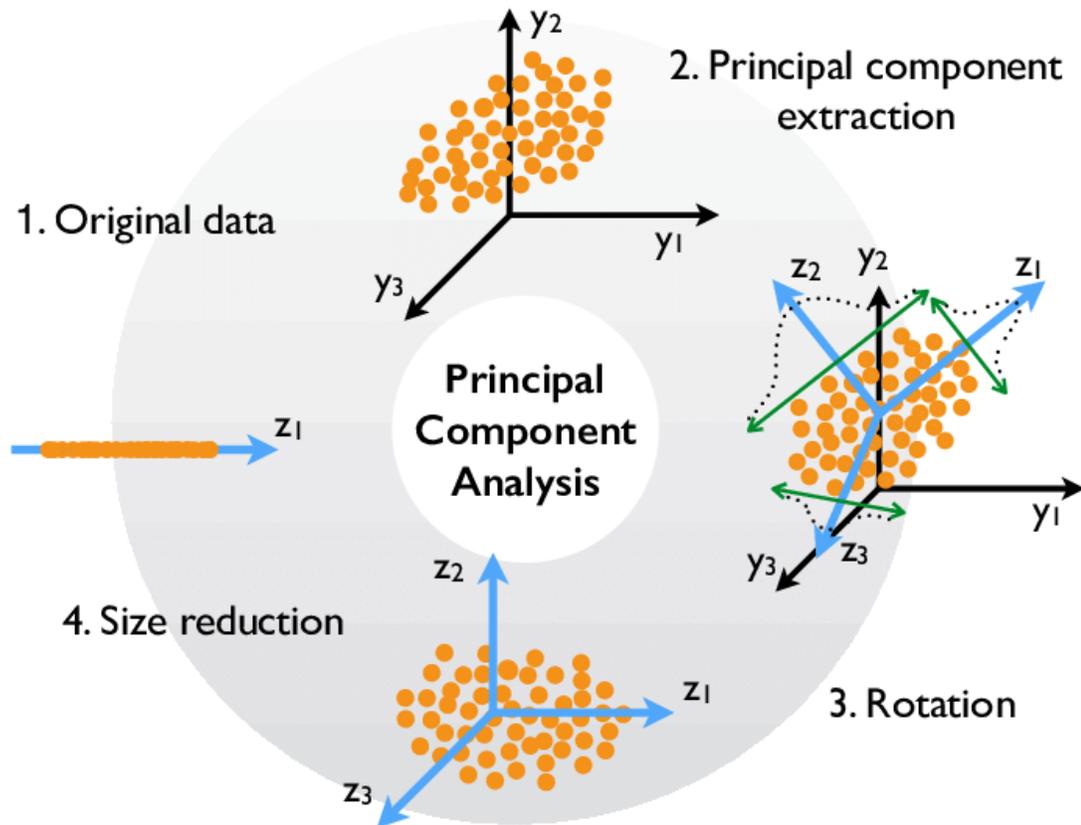


Figure 2.18: Example of PCA and dimensionality reduction[4]

the encoder which generates and outputs the code. The code is then passed to the decoder, with which it attempts to regenerate the original input, as seen in Figure 2.19. A cost function is applied to the combined differences in the output and passed back through the neural networks of the autoencoder for training.

**Fully Connected:** A fully connected autoencoder takes an input and flattens it into a 1-dimensional vector. The vector is then passed through a series of fully connected layers to produce the code. The code is passed through the fully connected layers of the decoder and then to a reshape layer that reconstructs the output to its original dimensions of the input.

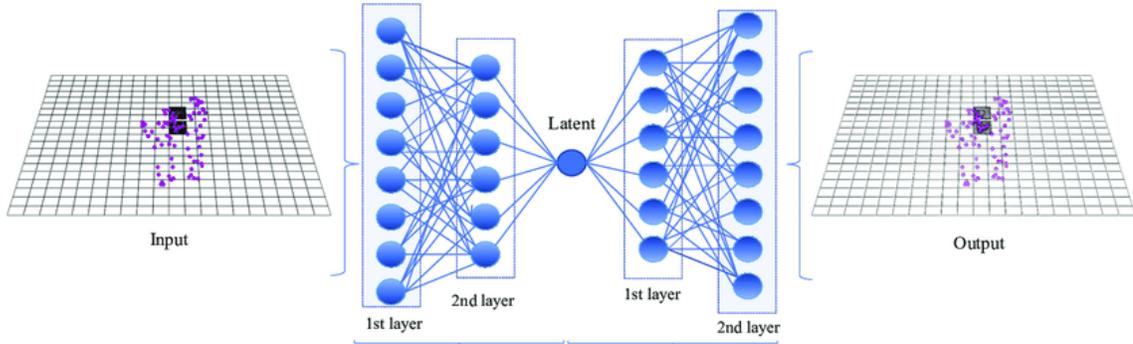


Figure 2.19: Topology of a fully-connected autoencoder[25]

**Convolutional:** Convolutional layers enable an autoencoder to include spatial information within the encodings. The early layers encode simple features such as edges, while the latter layers may be composed of larger structures such as trees.

**Noise:** The inclusion of noise promotes the autoencoder to include the most salient features relevant to the problem within the encoding, while excluding the features that are not. The addition of noise to the input is a means of facilitating this concept during training. Noise is produced via a random gaussian distribution and added to the input. During training the noisy input is passed through the layers of the encoder, which creates a code representation, then reconstructed via the decoder and the loss is computed by comparison with the noise free version of the input. In this fashion, the autoencoder is trained to pick up on the most prevalent features within the image.

**Regularizer:** A regularizer is applied to prevent the autoencoder from overfitting the data. For autoencoder applications that utilize a reduced dimension code representation, the stochastic metric has far less dimensions than that of the input image. A regularizer is utilized to enforce a sparse representation of the network to accentuate the differences between the code representations.

## 2.5.4 Clustering

Clustering is a means of grouping things with like features together based upon a metric. The results of two examples of clustering, GMM and K-means, are shown in Figure 2.20.

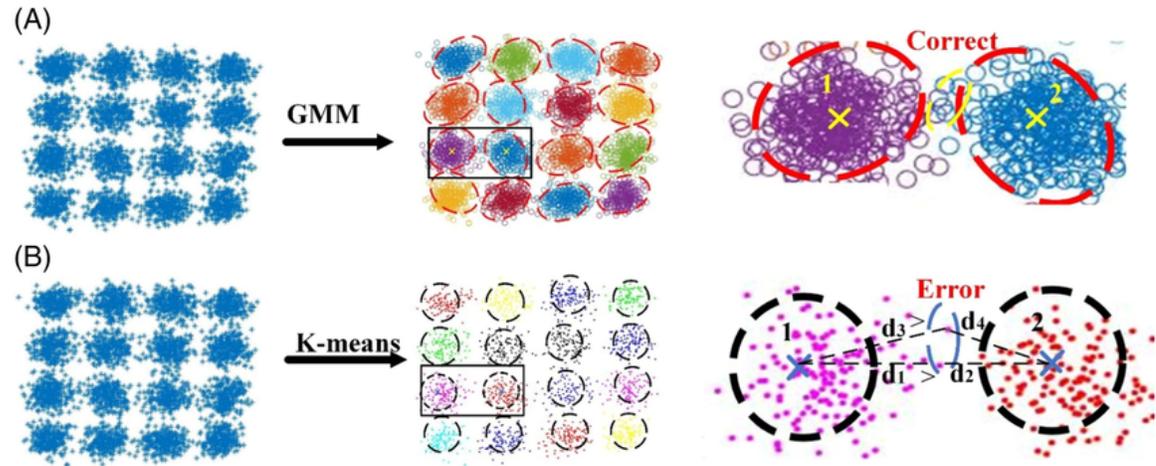


Figure 2.20: Comparison of (A) GMM probabilistic soft boundaries vs (B) K-means hard boundaries[32]

**K-Means:** K-Means is a simple clustering algorithm that attempts to separate data into  $K$  clusters with hard boundaries.

**Gaussian Mixture Models (GMM):** A Gaussian Mixture Model provides a probabilistic model for representing clusters. The soft boundaries of the clusters allow for overlap and the dimensions of the clusters may be elongated to better represent the data.

## 2.6 Image Comparison

Image comparison can be used to quantify the likeness or dis-likeness between two images. Mean squared error (MSE) and structural similarity index measure (SSIM) are two common approaches.

**Mean Squared Error (MSE):** MSE is a pixel-by-pixel method that averages the square of the differences between two images. The MSE equation is seen in Equation 2.1.

$$1/D \sum_{i=1}^D (x_i - y_i)^2 \quad (2.1)$$

**Structural Similarity Index Measure (SSIM):** The structural similarity index measure was introduced as an objective image quality index. It models image distortion as the combination of three factors: loss of correlation, luminance distortion, and contrast distortion [31]. While SSIM is more similar to the human visual system than MSE, it is not an effective substitute for applications that require models based directly on the human visual system [20]. The equation for SSIM is provided in Equation 2.2.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (2.2)$$

**Histogram:** A histogram representation is used to strip spatial information from the pixels in an image. The resolution of a histogram is configured by the bin size. Pixels of similar intensity are grouped and counted within a single bin. A histogram is constructed for each color channel of an input image. Figure 2.21 shows two sample images and their resultant histogram representations.

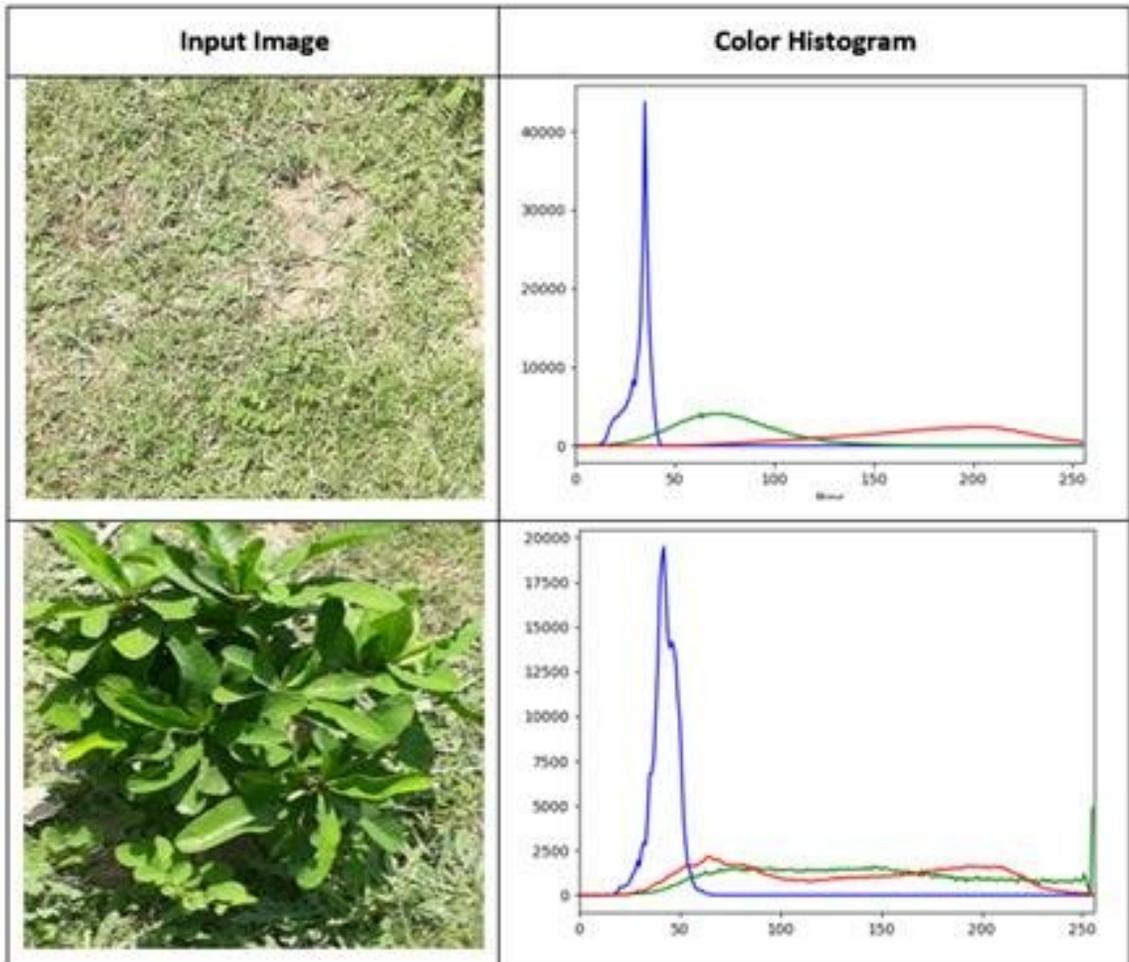


Figure 2.21: Example of RGB image conversion to histogram[2]

## Chapter 3

# A Spatiotemporal Approach to Wildfire Spread Behavior Prediction

Wildfires can be extremely destructive to the environment and communities exposed to them. Predicting wildfire behavior may help save lives and lessen the risk of exposure for people and wildlife. Traditional methods of wildfire prediction rely on simulations produced by calculating a series of equations. These equations were derived through experiments in the lab and on the field. Another method of predicting wildfire spread behavior is possible by modeling spectrometer data with environmental conditions. Direct spatiotemporal modeling of real-world wildfire behavior is faster and more adaptive than traditional approaches. Advances in sensor data acquisition have opened the door to this novel approach, and will be explored here.

### 3.1 Wildfire Spread Behavior Detection

Satellite based spectrometer instrumentation is capable of detecting and locating wildfire activity. As the wildfire spreads, new hot-spots are logged in the spectrometer data and maps of the fire boundaries, active areas, and burnt regions may be constructed. A time series of these maps may be produced for each wildfire event and used to train a spatiotemporal model for future wildfire behavior prediction.

**VIIRS 375-Meter Thermal Anomaly Data Limitations:** Wildfire spread in light fuels at fast rates may not be detected, because a region of the fire may combust and cool down in the time between scans. Spatial resolution of 375-meter is relatively coarse for mapping applications. Due to these constraints, medium to large fires, with consistent, slow to moderate spread rate are primarily considered for wildfire spread behavior prediction by the proposed method.

## 3.2 Wildland Environmental Conditions

In order to provide the machine learning algorithms with context for wildfire spread, the time series data may be augmented with relevant environmental data. Fuel maps, topographical information, and weather conditions are the primary drivers of wildfire spread behavior. Inclusion of these factors as input features may improve model performance by providing cause for the observed effect of wildfire spread.

**LANDFIRE Fuel Map Product Data Limitations:** LANDFIRE provides fuels and topographical data across the United States with a 30-meter spatial resolution that is updated once every several years. Due to the low update rate of the LANDFIRE products, fuel maps may not accurately represent the current condition of wildland fuels.

**Weather Data Limitations:** Weather station data is generalized across large regions of the studied area. Localized conditions such as terrain specific wind patterns and precise precipitation quantities are not considered in this study.

## 3.3 Spatiotemporal Machine Learning

ConvLSTMs are machine learning models specifically designed to take spatial relationships into consideration with sequence data. Two-dimensional arrays may be constructed, each representing pertinent information about cells in a map grid. Each two-dimensional array may contain different information about the map space; such

as, detected hot-spots, fuel types, slope angle, or wind. Utilizing multiple simple two-dimensional arrays, rather than one complex two-dimensional array, reduces the data type requirements of the model, improving computational performance. A set of these rectified two-dimensional arrays of the same size may be stacked by augmenting the container array with a third dimension. This three-dimensional stack of time instance conditions may be stacked in series using an additional fourth dimension for each sample. Finally a five-dimensional array of multiple samples is input to the ConvLSTM model for training. Predictions may be produced from the trained model utilizing an unseen four-dimensional test sequence, as depicted in Figure 3.1.

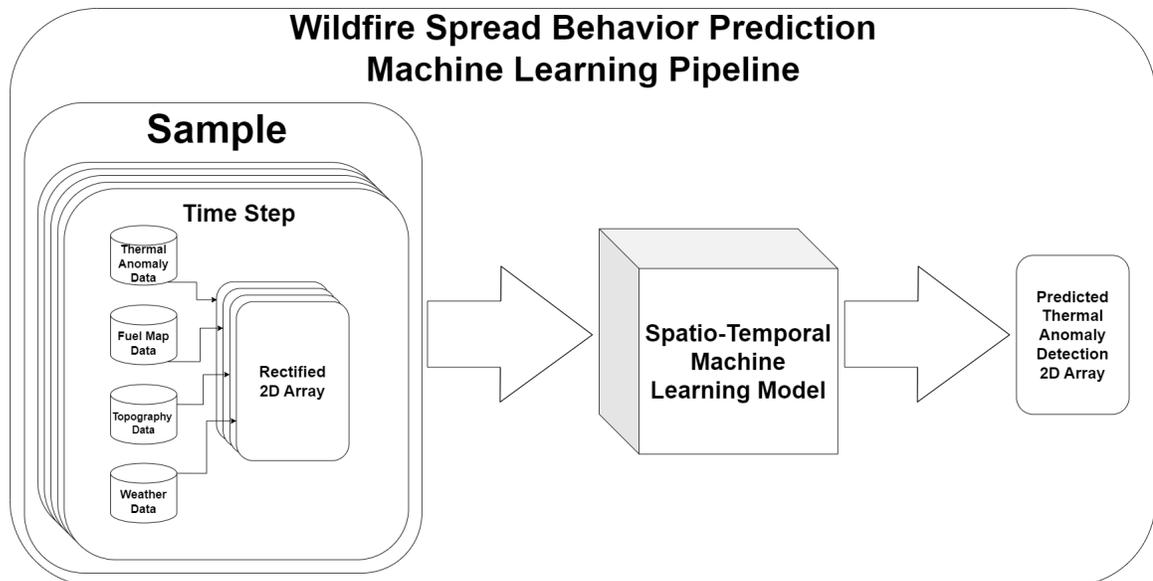


Figure 3.1: Wildfire Spread Behavior Prediction Model machine learning pipeline

# Chapter 4

## Wildfire Data Set and Wildfire Spread Behavior Prediction Model

### 4.1 Wildfire Data Set

In order to train a model capable of simulating wildfire in the spatiotemporal domain, a data set must be constructed. The features of the data set contain influential factors relevant to wildfire behavior. The data set consists of multiple series of two-dimensional arrays, with channels comprised of spatial, temporal, and spatiotemporal data. Each series is composed of sequential time instances that are representative of a wildfire event. Mapping software is utilized to consolidate data types and rectify the two-dimensional arrays.

#### 4.1.1 Fuel Map Data

Fuel map data may be sourced from the USGS LANDFIRE Data Distribution site[23]. Several types of fuel maps are available. For this application, the 2022 distribution of the 13 Anderson Fire Behavior Fuel Models data is selected[3].

#### Downloading Data from the USGS LANDFIRE Data Distribution Site:

1. Navigate to the USGS LANDFIRE Data Distribution site at <https://landfire.gov/viewer/>
2. Select a geographic region with the Data Download Tool, as seen in Figure 4.1

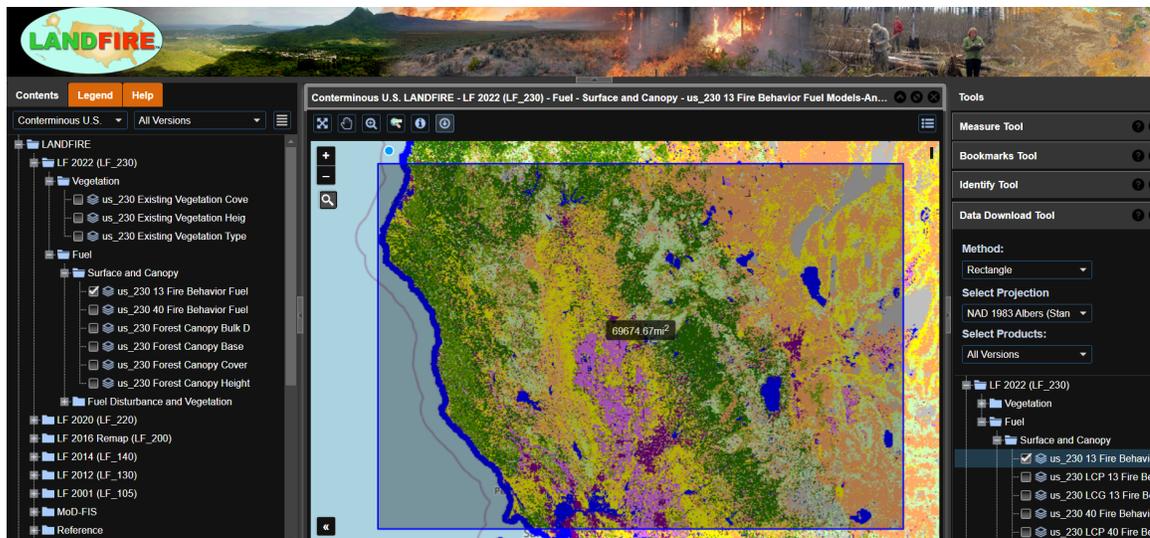


Figure 4.1: Screenshot from the LANDFIRE Data Distribution site [23]

3. Select the products to download from the toolbox
4. Enter an email address and click the Download button

**Importing Fuel Map Data into ArcGIS Pro:** For this application we will add a light fuels and a heavy fuels map to the container map. 13 Anderson Fire Behavior Fuel Models 1 - 7 are included in the light fuels map, and 8 - 12 are included in the heavy fuels map. An example of the binary representation of a light fuels map created in ArcGIS Pro is shown in Figure 4.2. All non-fuel types are excluded from both fuel maps.

1. Open an existing, or create a new container map
2. Make a folder connection with each fuel map to import
3. Add the fuel map(s) to the current container map
4. Select the fuel type color value to consolidate like fuel types

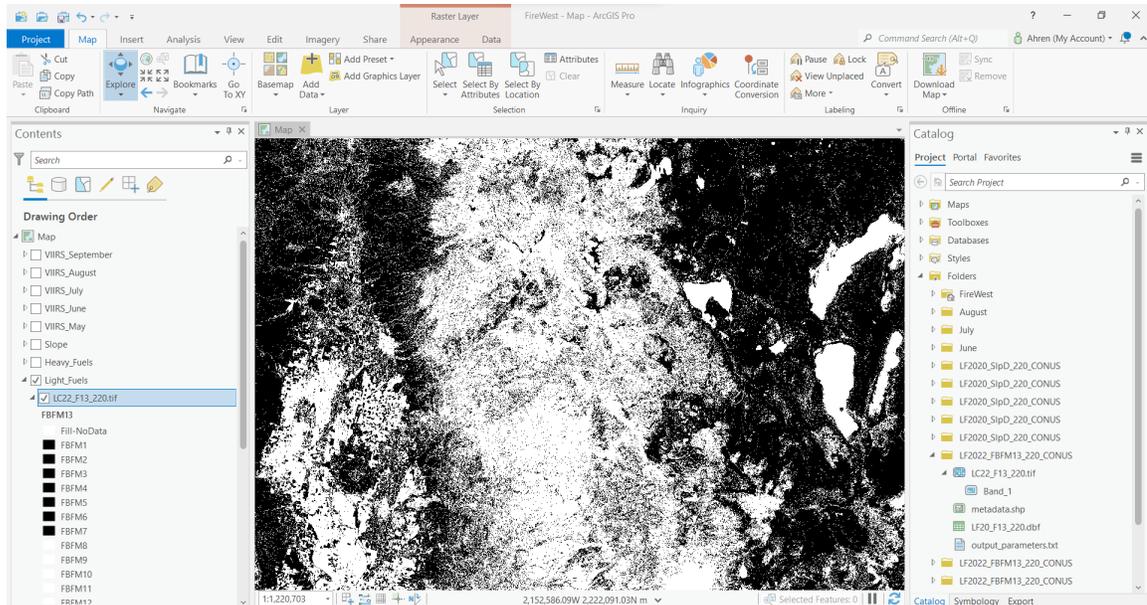


Figure 4.2: Screenshot of a binary light fuels map from the ArcGIS Pro mapping tool [9]

### 4.1.2 Slope Angle Data

Slope angle data may be sourced from the USGS LANDFIRE Data Distribution site [23]. Slope angle data is selected in the degrees format for this application.

**Importing Slope Angle Data into ArcGIS Pro:** For this application we will add a slope angle map that indicates areas with a slope greater than 15 degrees.

1. Open an existing, or create a new container map in ArcGIS Pro
2. Make a folder connection with each slope angle map to import
3. Add the slope angle map(s) to the current container map
4. Set the color value threshold to 15 to indicate areas with slope greater than 15 degrees

### 4.1.3 Thermal Imaging Data

Thermal Imaging data may be viewed with the NASA Fire Information and Resource Management System (FIRMS) online map [18], as shown in Figure 4.3. Data beyond seven days in age may be requested through the FIRMS Archive Download tool [16]. For this application, data from the VIIRS sources are selected.



Figure 4.3: Screenshot of thermal imaging data from the FIRMS US and Canada map [18]

#### Downloading Data from the NASA Fire Information for Resource Management System (FIRMS) Archive:

1. Navigate to the FIRMS Archive <https://firms.modaps.eosdis.nasa.gov/download/>
2. Specify satellite data source
3. Specify format and time range
4. Enter an email address and submit the request

#### Importing Thermal Imaging Data into ArcGIS Pro:

1. Parse the thermal imaging data into individual files by date into .csv format

2. Open an existing, or create a new container map in ArcGIS Pro
3. Make a folder connection with the parsed thermal imaging data
4. Add each .csv to the current ArcGIS map
5. Format the data point marker to approximate the pixel size of the data

#### 4.1.4 Weather Data

Archived weather data including: daytime and nighttime temperatures, dew point, and wind speed is sourced from Weather Spark [7]. Weather data is generally available from the closest airport to the area of interest. Daytime and nighttime relative humidity is calculated using the equation in Figure 4.1. The Red Flag Weather Matrix in Figure 4.4 is used to determine if a Red Flag Warning, or Extreme Fire Weather Warning was in effect.

Red Flag Weather Matrix		Sustained Winds				
		<6 mph	6--11 mph	12-20 mph	21-29 mph	30+ mph
H u m i d i t y	Daytime Min <29-42% and/or Night Max 60-80%					RFW
	Daytime Min <19-28% and/or Night Max 46-60%				RFW	RFW
	Daytime Min <9-18% and/or Night Max 31-45%			RFW	RFW	RFW
	Daytime Min <9% and/or Night Max <31%		RFW	RFW	RFW	PDS RFW
		Low	Medium	High	Very High	Extreme

Figure 4.4: The National Weather Service Red Flag Weather Matrix [30]

**Calculating Relative Humidity from Temperature and Dew Point:** Relative humidity values are derived from dew point and temperature measurements sourced from Weather Spark using the equation in Equation 4.1.

$$RelativeHumidity = 100 * \frac{e^{(17.625*DP)/(243.04+DP)}}{e^{(17.625*T)/(243.04+T)}} \quad (4.1)$$

#### 4.1.5 Constructing a Wildfire Event Time Instance Sequence

Time instance data for wildfire events is compiled into sequences for training the spatiotemporal model. A single viewpoint and scale is chosen for each wildfire event. All wildfire events use the same scale. Bitmap images of size 100-pixels x 100-pixels are created for each layer of the map, including: thermal imaging hot-spot detection, burnt area, light and heavy fuels, slope, red flag and extreme fire weather warning areas. The bitmap images are converted to a bit depth of one.

1. Select a viewpoint and scale for the wildfire event sequence and bookmark it
2. Select the slope layer and deselect all others
3. Select export type as bitmap and image size as 100-pixels x 100-pixels
4. Export the image
5. Repeat steps 2 through 4 for the fuel and fire weather maps
6. Create the hot-spot images by selecting the thermal imaging layer for the initial day of the wildfire event sequence and deselect all others
7. Select export type as bitmap and image size as 100-pixels x 100-pixels
8. Export the image
9. Repeat steps 6 through 8 for each day of the wildfire event time instance sequence

10. To create the burnt area images, repeat steps 6 through 9, but do not deselect any of the prior days thermal imaging layers throughout the sequence
11. Convert each of the images generated above to a bit depth of 1

## 4.2 Wildfire Spread Behavior Prediction Model

A spatiotemporal machine learning model is constructed to predict wildfire spread behavior. The model takes sub-sequences of the Wildfire Data Set as input, and returns a sequence one time-step in advance of the input sequence as output. The model is composed of multiple layers that add depth, trainable parameters, and subsequently improve the ability of the model to infer complex behaviors, and generalize to unseen data.

### 4.2.1 Constructing the Model with TensorFlow

The Tensor Flow Sequential object is used to construct the model [29]. The model consists of an input layer, three ConvLSTM2D layers, with BatchNormalization between them, and a Conv3D layer for output. The plot in Figure 4.5 depicts the model architecture. A 5-dimensional tensor is input for training the model, which is dimensionally composed of: samples, time steps, channels, columns, and rows.

### 4.2.2 Training the Model

The Wildfire Spread Behavior Prediction Model is trained on samples from the Wildfire Data Set. Each sample consists of a defined length sequence of time instances from the set. The time instances include thermal imaging, topographical, fuels, and weather data. The differences between the time steps in the sequences define the representation of wildfire behavior.

**Parsing the Data Set into Samples:** Samples are constructed by parsing the data set into time series of length defined by the model. The channels of each time

## Wildfire Spread Behavior Prediction Model

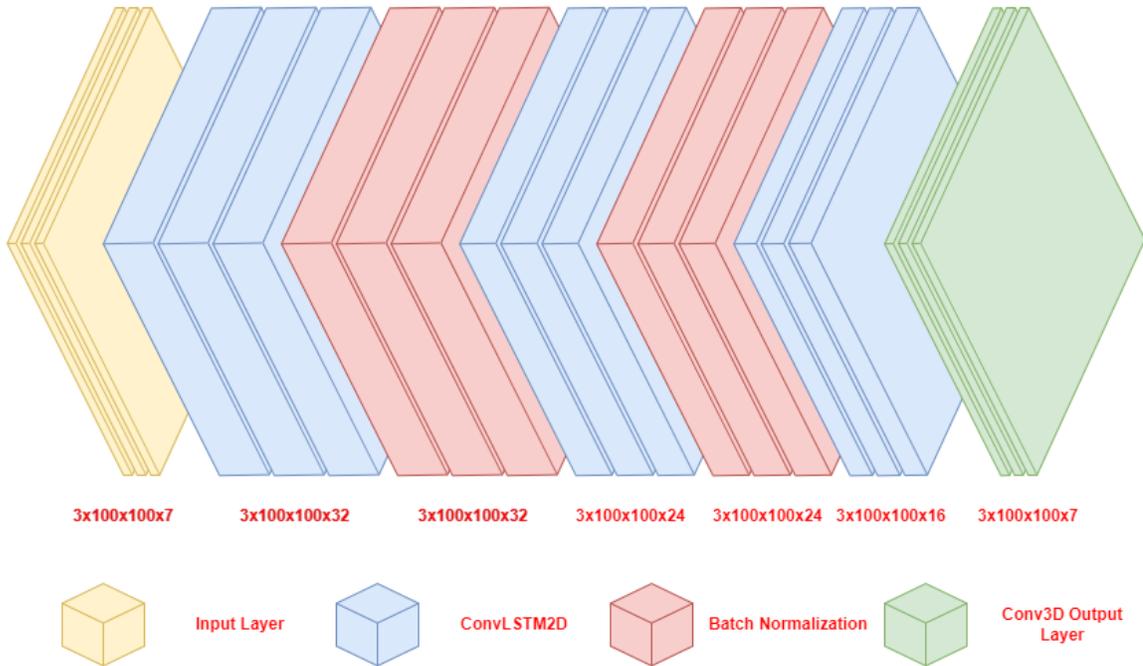


Figure 4.5: Wildfire Spread Behavior Prediction Model

series are composed of rectified layers from each time step of the data set. Samples one time step in advance of the input are used to train the output of the model.

### 4.2.3 Making a Prediction with the Model

The Wildfire Spread Behavior Prediction Model is trained to predict one time step in advance of the input. The input for a prediction consists of a defined length sequence of time instances, similar to one of the samples used to train the model. The output of the model has the same dimensional shape as the input, but is one time-step in advance. The last time instance in the output sequence contains the model prediction. Hot-spot and burnt area prediction arrays are isolated from the remainder of the output array.

**Hot-Spot Prediction:** The hot-spot prediction is the inferred array representation of the thermal imaging data, one time-step in advance of the input. Hotspots indicate

areas where a wildfire is active, and spread behavior may be present.

**Burnt Area Prediction:** The cumulative burnt area is predicted by including all prior hot spot detection in the subsequent inputs. Depending on the end use, burnt area may be more effective than hot-spot prediction at conveying pertinent information to wildfire spread over time.

**Confidence Value Threshold:** The output of the model includes a two-dimensional representation of a thermal imaging map for the next step in the time sequence, that serves as a hot-spot prediction. Each pixel of the hot-spot prediction is an 8-bit encoding of a confidence value that the area on the map it represents will be an active fire area. For input into the model these 8 bit values must be converted to binary. A threshold value is used to force anything below to zero and anything above to one. Burnt area predictions are handled with the same method. As seen in the sequence example in Figure 4.6, depending on the value of the threshold, predictions beyond one time step may differ significantly.

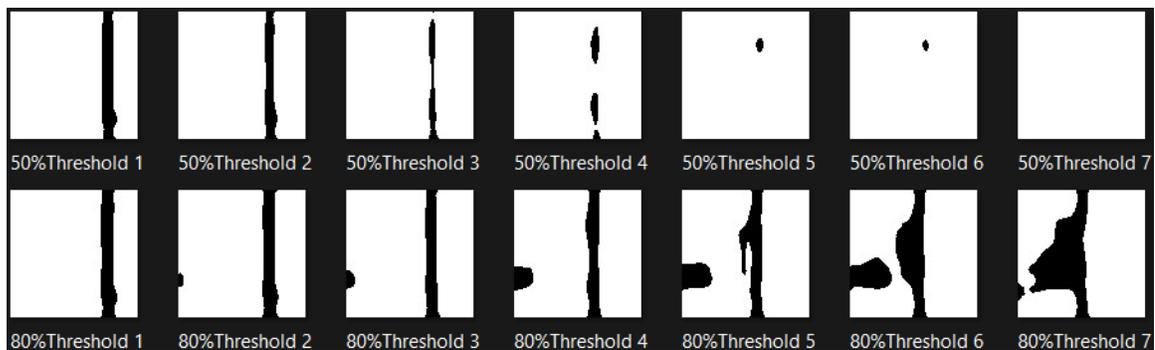


Figure 4.6: Confidence value threshold example

**Predictions Beyond One Time Step:** Predictions beyond one time step are produced by including data from prior predictions as input for subsequent predictions. Model predictions are converted from confidence values to binary, prior to being included in the input sequence. Input sequences for predictions less than the model

sequence length of time-steps in advance are composed of actual data for the earlier time instances in the sequence, and predicted data for the latter time instances. Predictions made for time-steps beyond the model input sequence length are composed entirely of predicted data. This method allows for predictions to be made an unlimited number of time steps in advance of the original input.

# Chapter 5

## Experiments

### 5.1 Simulated Data

Simulated data was constructed, that exhibited simple behaviors, to gauge the ability of the ML models to make accurate predictions in specific scenarios. The models were initially trained on this data to obtain information that was useful to refining the models and their inputs for improved performance.

#### 5.1.1 Simple Flaming Front

A data set was constructed to simulate a flaming front burning across a level field at a constant rate. This scenario was used to gauge the ability of the model to predict fire behavior in a consistent fuel type, without being influenced by external factors. The model input and a ground truth comparison of a single prediction may be seen in Figure 5.1. Dark pixels in the time step images are representative of detected hot spots in the radio-spectrometer data and predictions.

#### 5.1.2 Fuel Map

The addition of a simple binary fuel map to the flaming front experiment provided a degree of complexity for the model to simulate. The model input and ground truth comparison of a single prediction may be seen in Figure 5.2. Light pixels in the fuel map indicate the presence of combustible material in that location, dark pixels represent the absence of fuel.

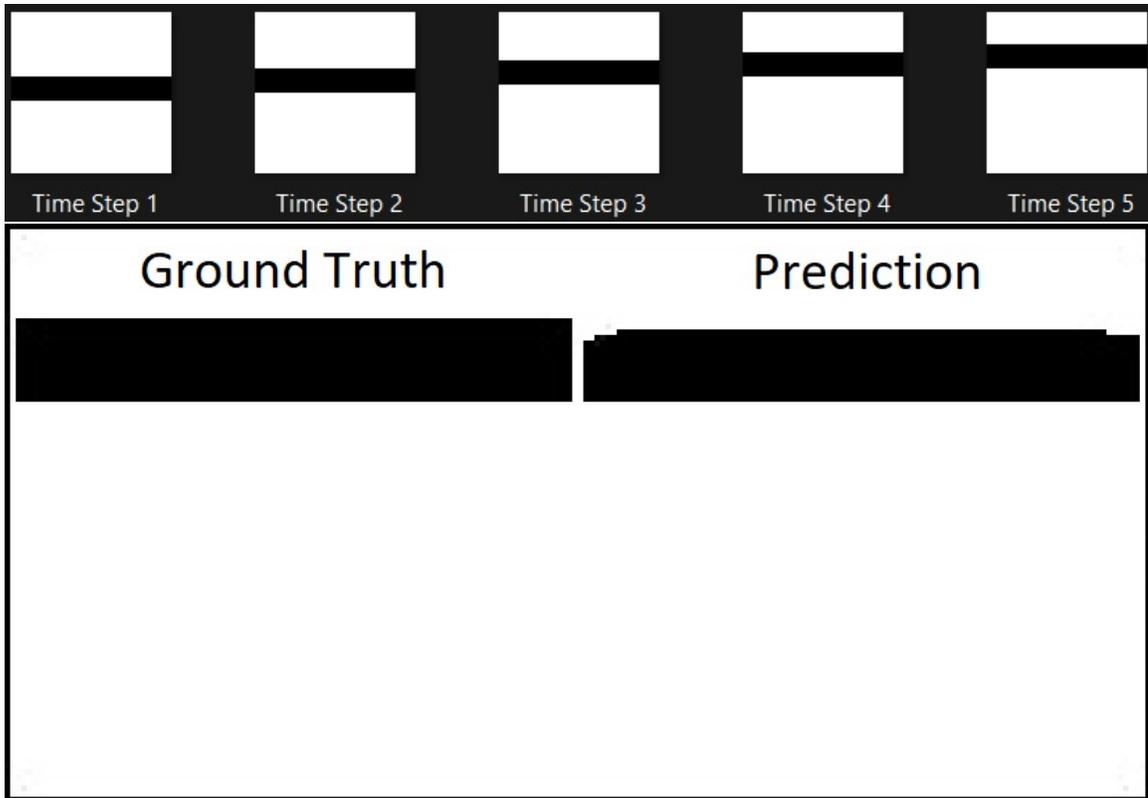


Figure 5.1: Simulated simple flaming front input and prediction comparison with ground truth

### 5.1.3 Accelerating Factors

Light flashy fuel types, low fuel moisture content, and steep slopes, as well as hot, dry, and windy atmospheric conditions are all factors that accelerate fire spread. An accelerating factor map was included as an input to the model to simulate these conditions, and gauge the models ability to accurately predict fire growth through spatially variable conditions. The initial input to the model, as well as a side by side ground truth comparison of a sequence of single step comparisons, is shown in Figure 5.3.

### 5.1.4 Multiple-Step Predictions

Predictions beyond one time step from the model input were produced by including prior predictions in the input to the subsequent predictions. In Figure 5.4, the top

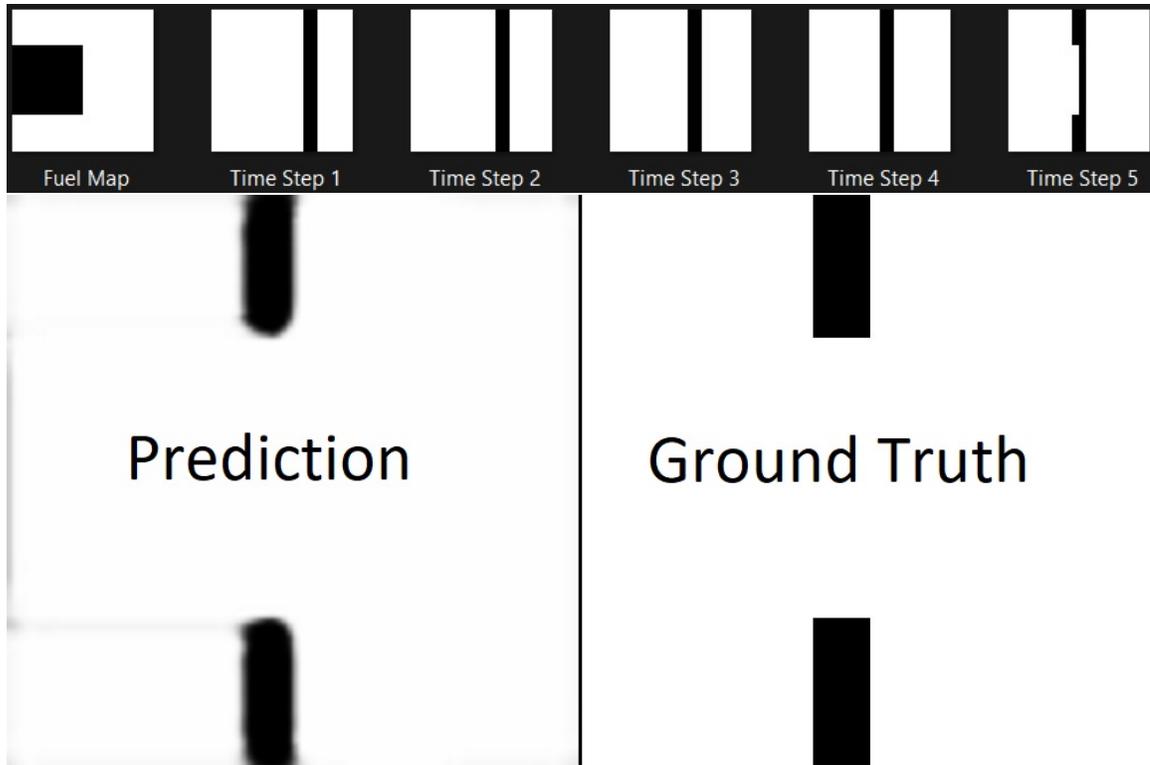


Figure 5.2: Simulated flaming front with a fuel map input and prediction comparison with ground truth

row includes the fuel map and a series of 5 time steps that were used to make the initial prediction, Prediction 1. Each subsequent prediction was made by replacing the original input steps with the available prior predictions.

## 5.2 Real-World Data

A data set was constructed from real-world measurements and wildland conditions that are associated with influencing wildfire behavior. Two-dimensional spatial components were aggregated into higher-dimensional arrays and compiled together to form time series. A spatiotemporal machine learning model was trained on the data set, a portion of which was reserved for validation and testing.

**Real-World Data Set Specifics:** The real-world data set was constructed from 10 wildfire events that occurred on the West Coast of the United States during the

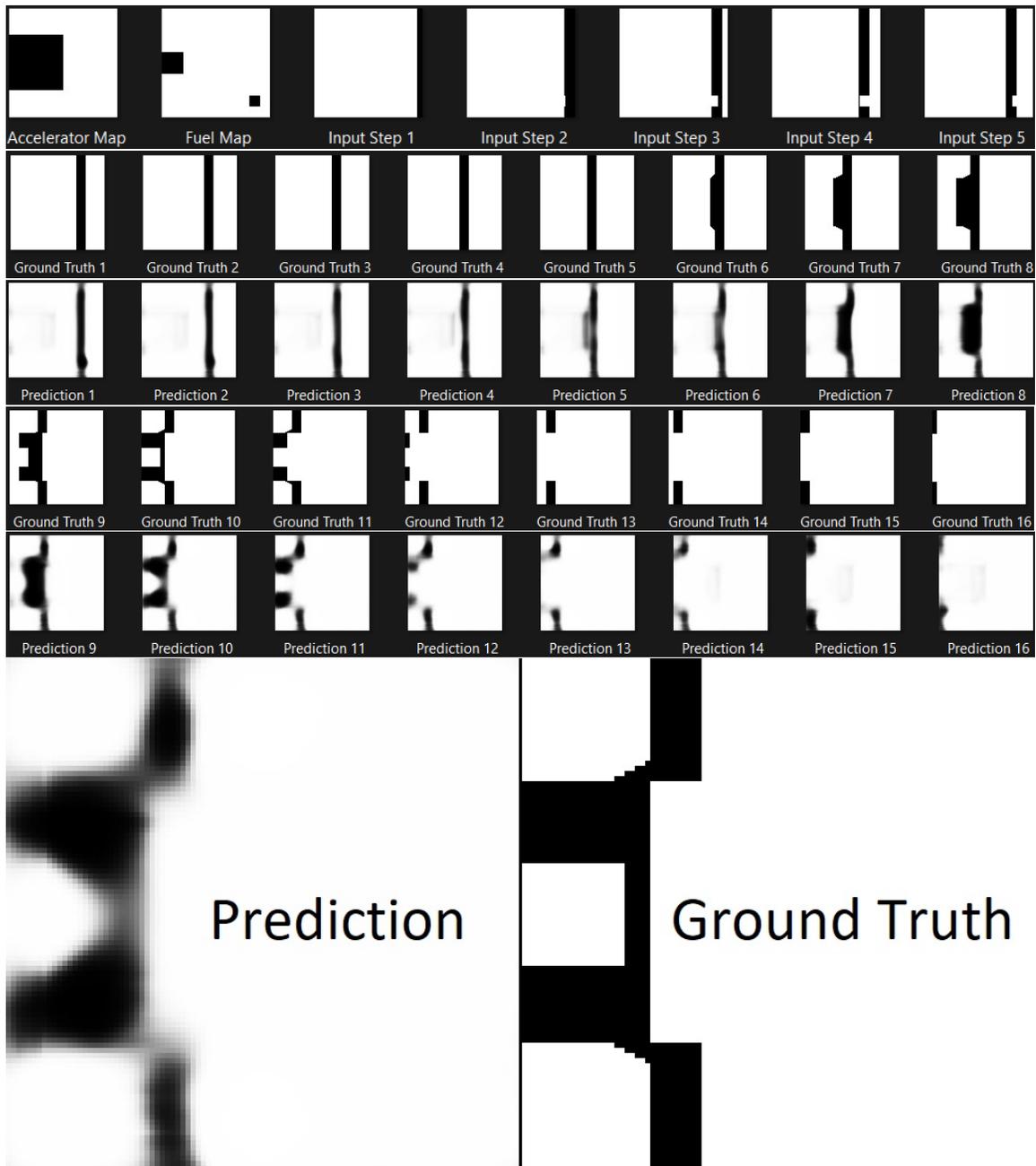


Figure 5.3: Simulation with fuel map and accelerator input and prediction comparison with ground truth

2021 fire season. The wild fire events ranged in duration from 9 to 54 days in length. A set of time instance images were produced for each day of each wildfire event. The set of images for each time instance contained:

- Detected thermal imaging hot-spot



Figure 5.4: Single and multiple step prediction comparison

- Wildfire event burnt area
- Presence of Anderson 13 light fuel models 1 - 7

- Presence of Anderson 13 heavy fuel models 8 - 12
- Presence of slope angles beyond 15 degrees
- Occurrence of Red Flag Warning day
- Occurrence of Extreme Fire Weather day

**Training, Validation, and Test Samples:** Sample time series length of three was chosen, and 250 training samples were constructed from the real-world data set. Two wildfire events were excluded from the training data, reserving 55 samples for validation and testing.

### 5.2.1 Single-Step Predictions

A single prediction was produced for each sub-series of the portion of the data set that was set aside for testing. Accuracy of the predictions were gauged by calculating the MSE and SSIM, for both the hot-spot and burnt area predictions. The ground truth comparison for single-step hot-spot predictions can be seen in Figure 5.5, and single-step burnt area predictions in Figure 5.6.

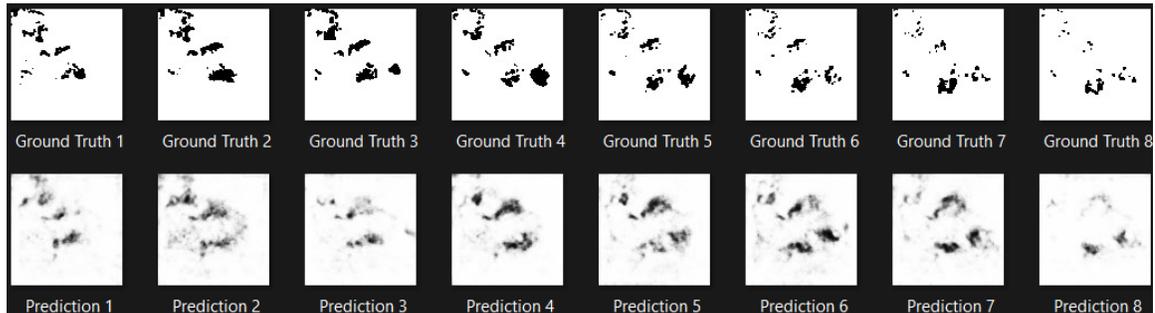


Figure 5.5: Single time-step hot-spot prediction vs ground truth

### 5.2.2 Multi-Step Predictions

Multiple predictions were produced for each sub-series of the portion of the data set that was set aside for testing. Predictions beyond one time-step were made by including predicted data as input for subsequent predictions. The ground truth comparison

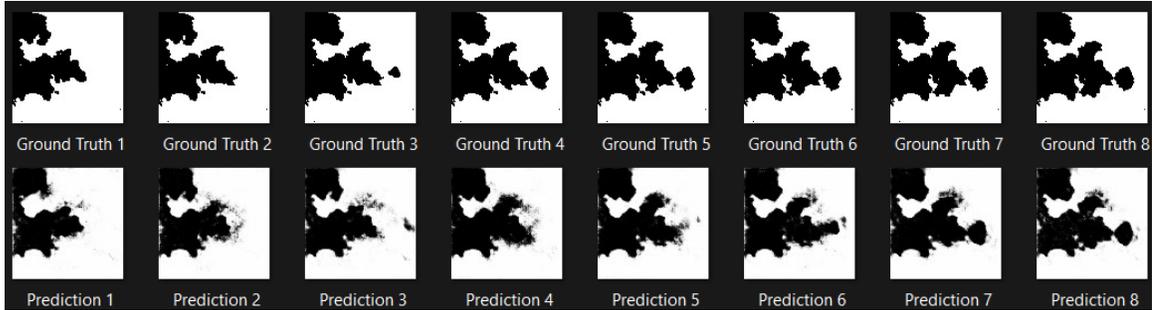


Figure 5.6: Single time-step burn area prediction vs ground truth

for multiple-step hot-spot predictions can be seen in Figure 5.7, and multiple-step burnt area predictions in Figure 5.8. Accuracy of the predictions were gauged by calculating the MSE and SSIM for both the hot-spot and burnt area predictions. A graph correlating accuracy of prediction by number of time-steps from real data was produced.

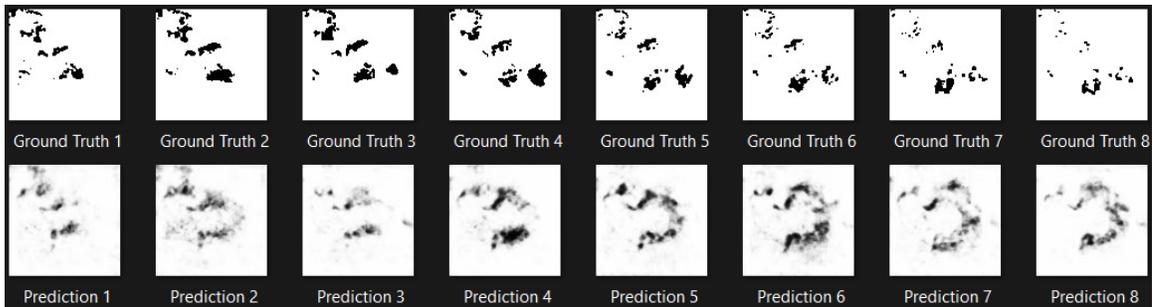


Figure 5.7: Multiple time-step hot-spot prediction vs ground truth

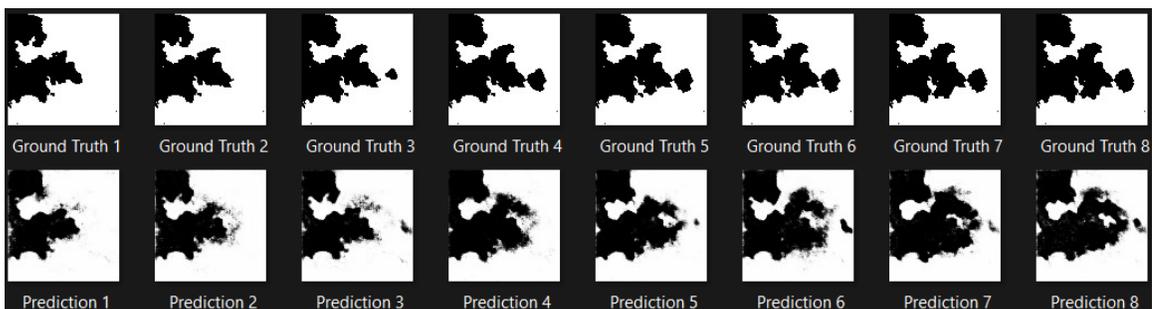


Figure 5.8: Multiple time-step burn area prediction vs ground truth

# Chapter 6

## Results and Analysis

### 6.1 Simulated Data

**Simulated Data Model Loss** Test data achieved similar model loss to the training data, due to the simple and predictable behaviors exhibited by the simulated data. The model training loss plot can be seen in Figure 6.1.

**Simulated Data Image Comparison** Image comparison results indicated that the models were capable of accurately inferring the simple behaviors exhibited by the simulated data. Multiple-step predictions produced more loss than single-step predictions, as can be seen in Table 6.1.

Experiment Name	MSE	MSE Range	SSIM	SSIM Range
Flaming Front	0.214	–	0.930	–
Fuel Map	.001	–	0.961	–
Accelerating Factor	0.033	0.052	0.891	0.169
Multiple Step	0.100	0.156	0.805	0.273

Table 6.1: Simulated training data prediction comparison results

### 6.2 Real-World Data

**Real-World Data Model Loss** Training and test data model loss were comparable throughout the model training process. At 500 epochs, test data model loss was approximately twice that of training data model loss, as can be seen in Figure 6.2.

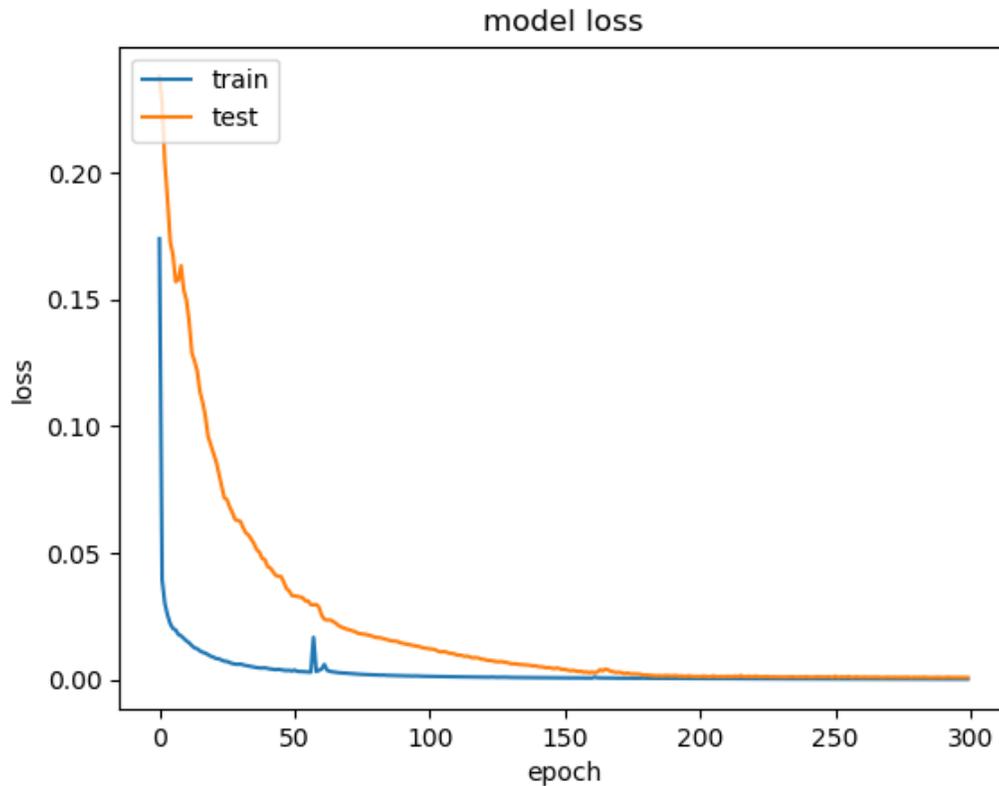


Figure 6.1: Simulated training data loss plot

The complex nature of wildfire behavior is attributed to the significant difference between test and training model loss.

### Single-Step Hot-Spot vs Burnt Area Prediction Performance Comparison

Image comparison results between hot-spot and burnt area predictions were near identical, as can be seen in Table 6.2. The model performed equally well at predicting both spread behavior formats.

Experiment Name	MSE	MSE Range	SSIM	SSIM Range
Heat Map Single-Step	0.053	0.099	0.801	0.273
Burnt Area Single-Step	0.050	0.093	0.802	0.236

Table 6.2: Real-world training data single time-step prediction comparison results

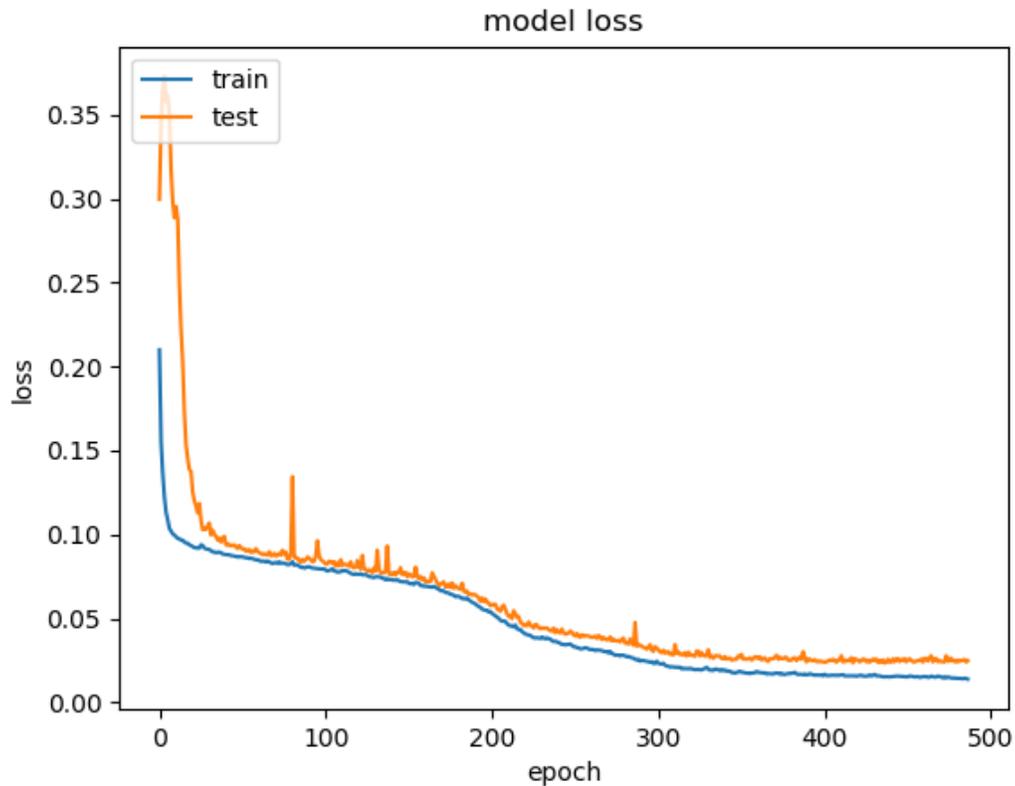


Figure 6.2: Real-world training data loss plot

**Multiple-Step Prediction Performance** Image comparison results for multiple-step predictions made by the real-world model indicated a consistent drop in performance as time steps from real data increased. Table 6.3 provides image comparison results for each prediction type, by the number of steps from ground truth model input. The Hotspot Prediction Comparison graph in Figure 6.3, when compared with the Burnt Area Prediction Comparison graph in Figure 6.4, shows that burnt area prediction image comparison performance drops off at a steeper rate. This may be attributed to the nature of the burnt area maps being comprised of the aggregation of the hot-spot prediction data, thus containing more differences.

Comparison Type	Step 1	2	3	4	5	6	7	8	9	10
Hot-spot MSE	0.054	0.057	.068	0.082	0.089	0.099	0.100	0.103	0.101	0.101
Hot-spot SSIM	0.801	0.793	0.771	0.735	0.721	0.697	0.679	0.668	0.667	0.661
Burnt Area MSE	0.049	0.052	0.062	0.100	0.109	0.126	0.155	0.165	0.176	0.200
Burnt Area SSIM	0.805	0.798	0.777	0.704	0.690	0.664	0.614	0.599	0.581	0.541

Table 6.3: Real-world training data multiple time-step prediction comparison results

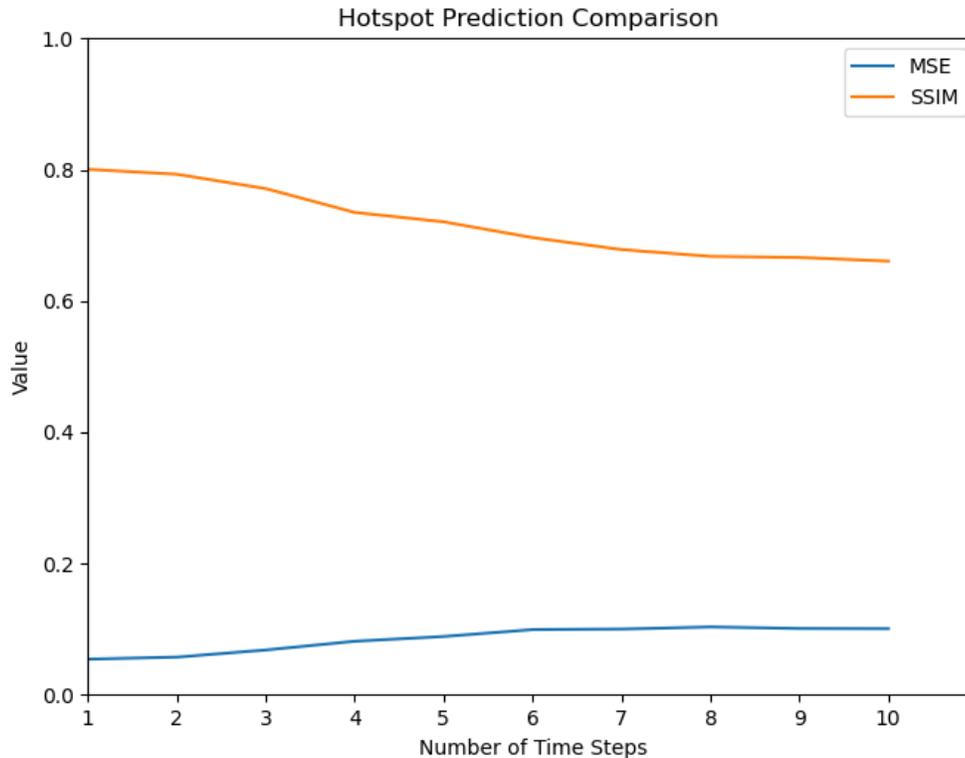


Figure 6.3: Average MSE/SSIM value vs number of time steps for hotspot prediction

### 6.3 Dixie Fire Case Study

The Dixie fire burned 963,309 acres in Northern California, starting on July 13th, 2021, and was fully contained on October 25th, 2021, a period of 104 days. One firefighter lost his life, and more than 1,300 structures were destroyed over the course of the burn[10]. It was the largest single wildfire ever recorded in California history, and the first known fire to burn over the Sierra crest. The impacted region prior to the Dixie fire and initial detection area may be seen in Figure 6.5.

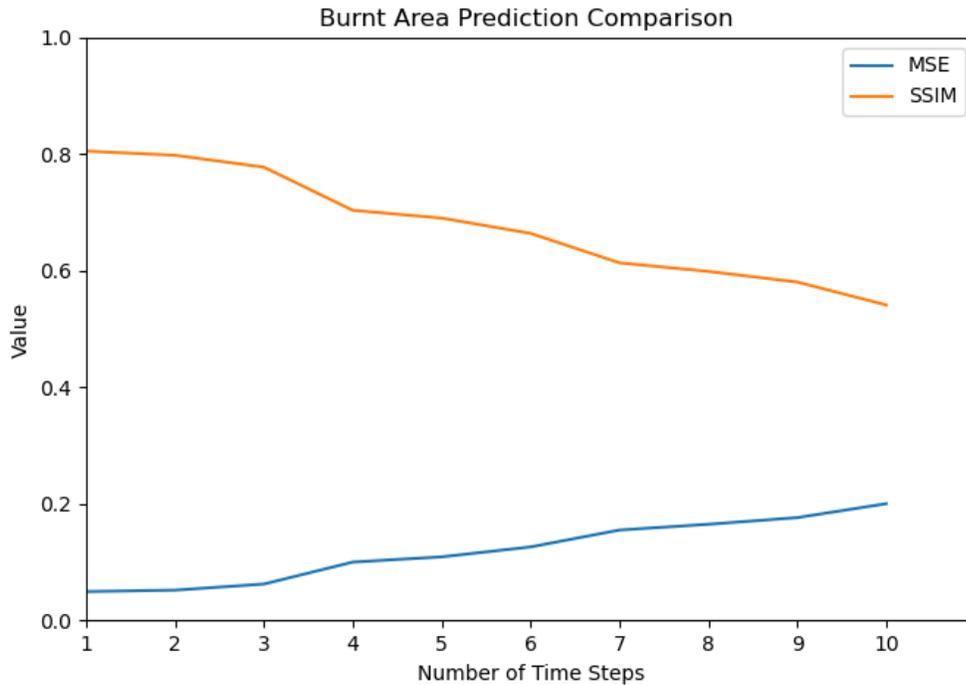


Figure 6.4: Average MSE/SSIM value vs number of time steps for burnt area prediction

### 6.3.1 Initial Growth Phase

The Dixie fire started on July 13th near the city of Paradise, CA. Over the next several days, it spread Northeast towards Lake Almanor. The Wildfire Spread Behavior Prediction Model was used to predict hot-spot detection areas in the region from July 17th to July 25th, the period from July 14th to July 16th was used as the initial input to the model. Figure 6.6 shows the predicted growth of the wildfire based on daily hot-spot predictions. Dark colors are chronologically earlier than light colors, as indicated by the legend. Figure 6.7 displays the ground truth hot-spot detection map in the same format and time range. In both figures, the wildfire can be seen progressing the the Northeast and slowing as it approaches the banks of Lake Almanor. Spotting may be seen in both the predicted and ground truth maps, as well as inhibited spread behavior in the primarily fuel barren area around Butt Valley Reservoir. The ground truth spread behavior outpaces the prediction in the easterly

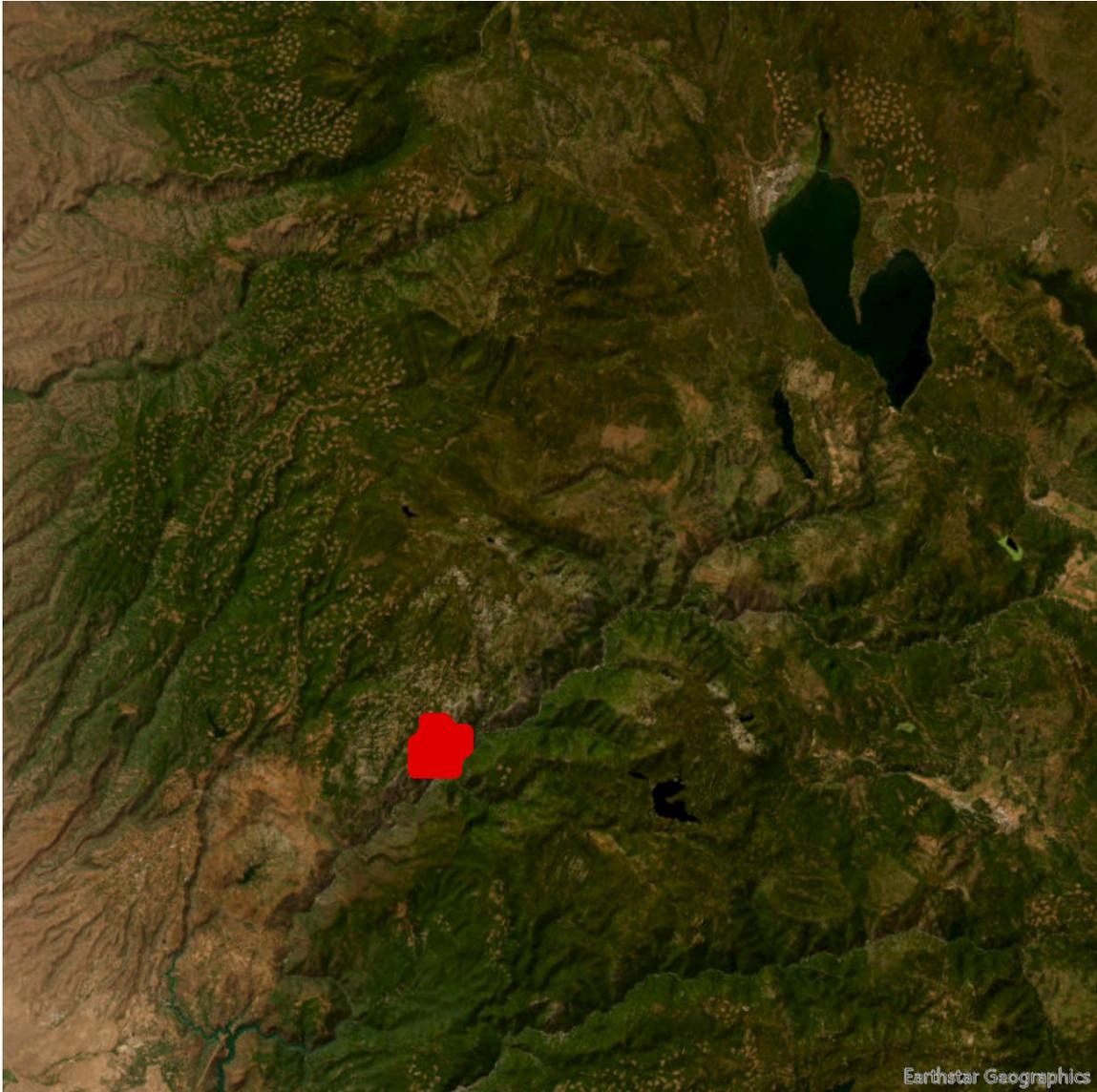


Figure 6.5: Satellite imagery of the impacted region prior to the Dixie Fire with the initial detection data overlaid [14]

direction during the latter stages of the predicted sequence.

### 6.3.2 Later Growth Phase

During the month of August, the Dixie fire had already burned hundreds of thousands of acres. It burned over the crest of the Sierra Nevada mountains and continued eastward, with rapidly expanding flanks to the North and South. The Wildfire Spread

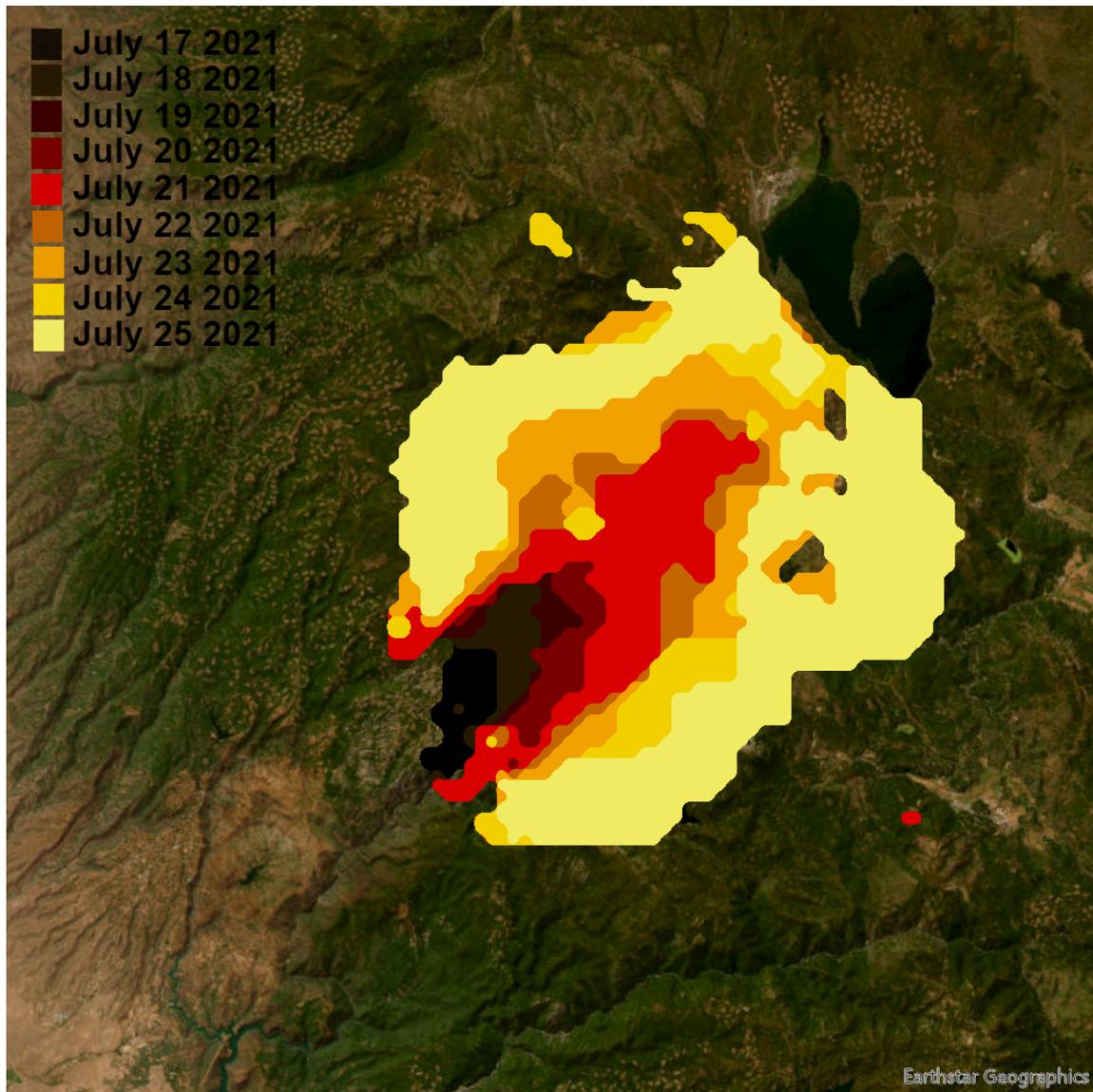


Figure 6.6: Predicted Dixie fire hot-spot detection map from July 17th to July 25th. The Behavior Prediction Model was used to predict hot-spot detection areas in the region from August 7th to August 11th, the period from August 4th to August 6th was used as the initial input to the model. Figure 6.8 shows the predicted growth of the wildfire based on daily hot-spot predictions. Dark colors are chronologically earlier than light colors, as indicated by the legend. Figure 6.9 displays the ground truth hot-spot detection map in the same format and time range. In both figures, the wildfire can be seen progressing East towards Susanville, CA and the surrounding areas. The

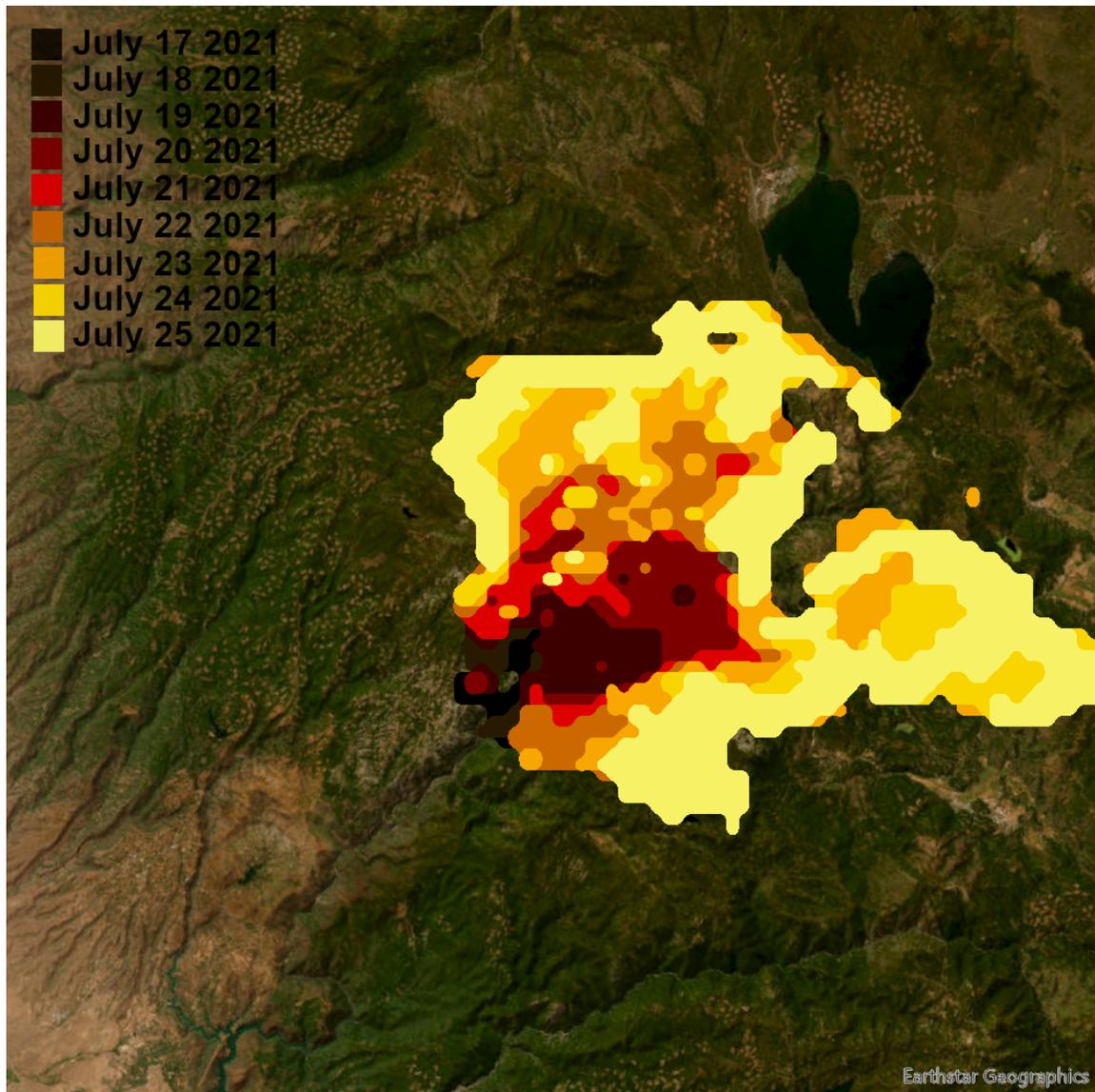


Figure 6.7: Dixie fire ground truth hot-spot detection map from July 17th to July 25th

model proves successful at predicting minimal hot-spot detections in areas devoid of fuels. The predicted hot-spot detection map suggests growth in excess of that observed by the ground truth data, possibly due to the lack of training data for fires that burn over the Sierra crest. Firefighting activities are also not available to the model, other than the observed impacts as recorded by the radiospectrometer data.

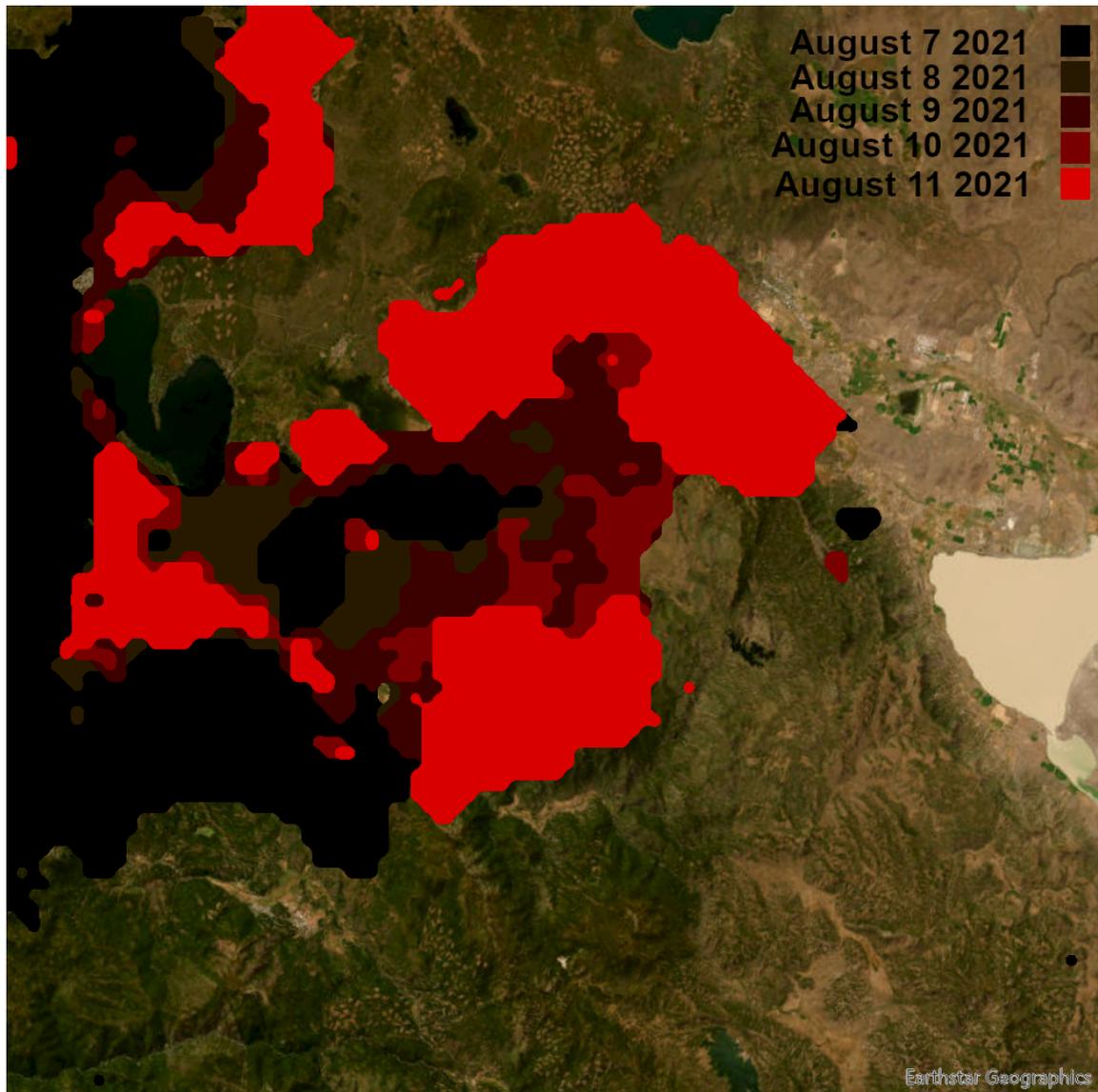


Figure 6.8: Predicted Dixie fire hot-spot detection map from August 7th to August 11th

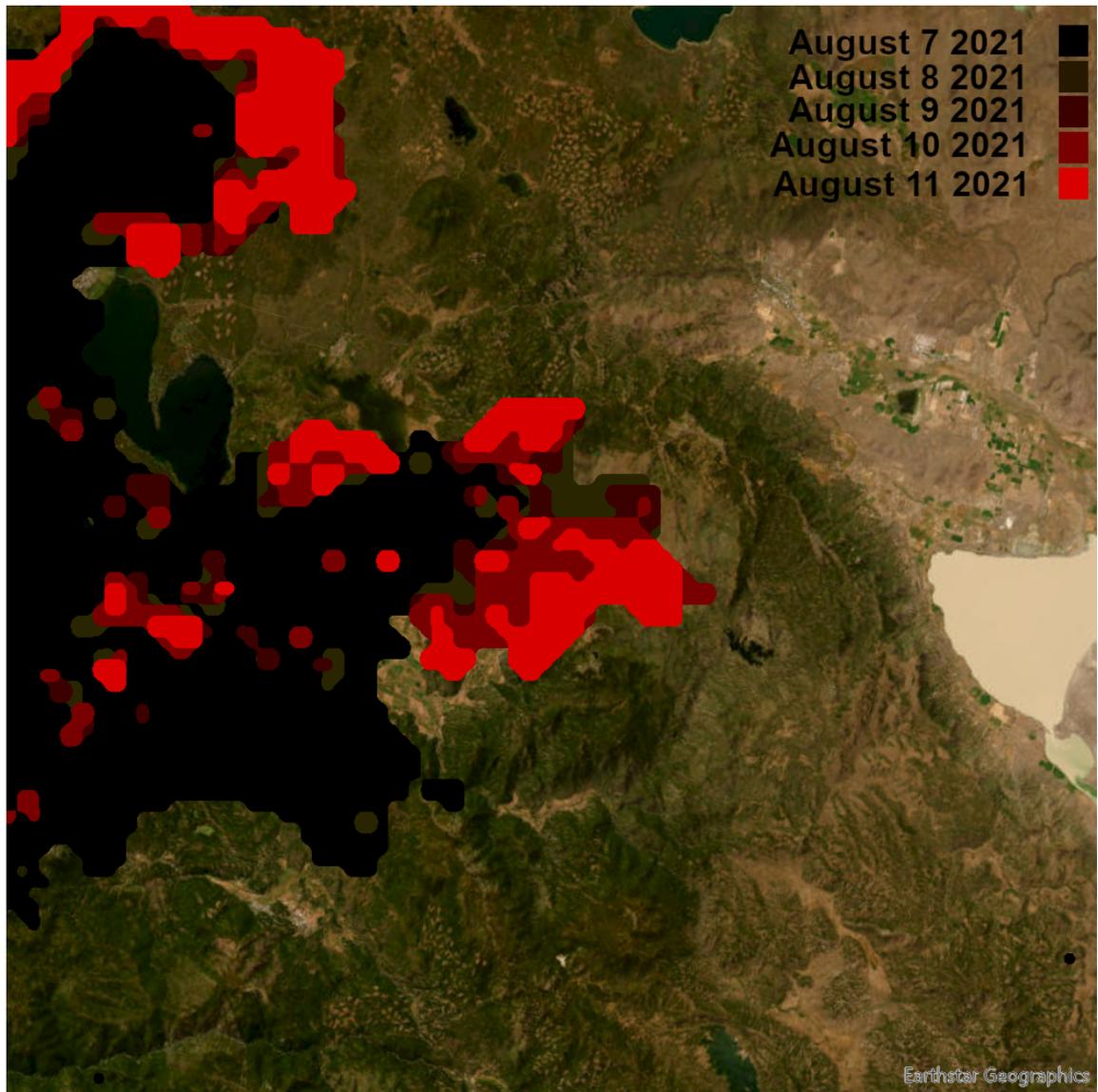


Figure 6.9: Dixie fire ground truth hot-spot detection map from August 7th to August 11th

# Chapter 7

## Conclusions and Future Work

### 7.1 Conclusions

Spatiotemporal machine learning models trained on real-world data are capable of effectively predicting wildfire behavior. The data set and model presented in this thesis, illustrate a novel approach to wildfire behavior prediction. Publicly available data from numerous sources about past wildfire events, was utilized to train a network of convolutional long-short term memory layers, to model and predict future wildfire behavior. Wildfire behavior was presented in two formats: thermal imaging hot-spot detection, and burnt area prediction. A method of predicting wildfire behavior well into the future was also presented. Predicted data from the model was rolled into subsequent iterations of the model as input, effectively enabling the model to make predictions an unlimited number of time-steps into the future. Two metrics for image comparison were used to measure the similarity of model predictions, to the ground-truth wildfire representations. As expected, image similarity decreased with an increase of time-steps predicted in advance. Though the model continued to produce realistic wildfire behavior predictions, even when relying solely on wildfire behavior data from preceding predictions, and the degradation in performance may be attributed to iterative stack-up error. A case study for the Dixie fire was also presented. It showcased the ability of the model to accurately predict wildfire spread in a real-world scenario. The low-resolution, and low-frequency of available data limited model performance, but the methods presented here will be applicable to

higher quality data as it becomes available.

## 7.2 Future Work

Spatiotemporal machine learning models have shown encouraging results in the domain of wildfire behavior prediction. Copious amounts of high-quality data are at the core of deep learning applications. Extending the Wildfire Data Set beyond the 2021 fire season may be the first step to improved model performance. The region of interest may also be expanded beyond the West Coast of the United States to North America, or even further. As more data is introduced, more features may need to be included to allow the model to localize predictions where necessary. Fuel types differ dramatically between regions, as do local weather patterns. Altering the model to utilize 8-bit, rather than binary, arrays will allow for modeling the wildland environment with greater precision and accuracy. The inclusion of generated wind maps composed of an array of vectors is of particular interest to this area of study. Including the time of thermal imaging acquisition in the model could yield improved results. Fire behavior often varies dramatically depending on the time of day, being more active in the midday. Including data relevant to firefighting activities may also be of benefit. Coordinates and estimated coverage areas of retardant drops and fire lines would add some missing context to the data. Presenting the predictions to the general public may also be of benefit, potentially affording residents of impacted areas more time to prepare for evacuation.

## References

- [1] Environmental Protection Agency. Climate change indicators: wildfires. November 1, 2023. URL: <https://www.epa.gov/climate-indicators/climate-change-indicators-wildfires> (visited on 01/15/2024).
- [2] Mansoor Alam, Muhammad Alam, Muhammad Roman, Muhammad Tufail, Umer Khan, and Muhammad Khan. Real-time machine-learning based crop/weed detection and classification for variable-rate spraying in precision agriculture. In *2020 7th International Conference on Electrical and Electronics Engineering (ICEEE)*, pages 273–280, April 2020. DOI: 10.1109/ICEEE49618.2020.9102505.
- [3] Hal E Anderson. Aids to determining fuel models for estimating fire behavior. In *Gen. Tech. Rep. INT-122*, 22p. Ogden, Utah: U.S.Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, 1982-04. URL: [https://www.fs.usda.gov/rm/pubs\\_int/int\\_gtr122.pdf](https://www.fs.usda.gov/rm/pubs_int/int_gtr122.pdf) (visited on 01/15/2024).
- [4] Aurelie Bellemans, Gianmarco Aversano, Axel Coussement, and Alessandro Parente. Feature extraction and reduced-order modelling of nitrogen plasma models using principal component analysis. *Computers & Chemical Engineering*, 115(12):504–514, 2018. ISSN: 0098-1354. DOI: 10.1016/j.compchemeng.2018.05.012. URL: <https://www.sciencedirect.com/science/article/pii/S0098135418304551>.
- [5] Daniel Berman, Anna Buczak, Jeffrey Chavis, and Cherita Corbett. A survey of deep learning methods for cyber security. *Information*, 10(4):122, April 2019. DOI: 10.3390/info10040122.
- [6] Facundo Bre, Juan Gimenez, and Víctor Fachinotti. Prediction of wind pressure coefficients on building surfaces using artificial neural networks. *Energy and Buildings*, 158:1429–1441, January 2018. DOI: 10.1016/j.enbuild.2017.11.045.
- [7] Cedar Lake Ventures, Inc. Weather spark: the weather year round anywhere on earth. URL: <https://weatherspark.com/> (visited on 01/15/2024).
- [8] Adnan Dehghani, Hamza Mohammad Zakir Hiyat Moazam, Fatemehsadat Mortazavizadeh, Vahid Ranjbar, Majid Mirzaei, Saber Mortezaei, Jing Lin Ng, and Amin Dehghani. Comparative evaluation of lstm, cnn, and convlstm for hourly short-term streamflow forecasting using deep learning approaches. *Ecological*

- Informatics*, 75:102119, 2023-07. ISSN: 1574-9541. DOI: <https://doi.org/10.1016/j.ecoinf.2023.102119>. URL: <https://www.sciencedirect.com/science/article/pii/S1574954123001486>.
- [9] esri. Arcgis pro. URL: <https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview> (visited on 01/15/2024).
- [10] CAL FIRE. California department of forestry and fire protection — cal fire. URL: <https://www.fire.ca.gov/> (visited on 01/15/2024).
- [11] Forests and Rangelands. The national cohesive wildland fire management strategy, 2013. URL: <https://www.forestsandrangelands.gov/strategy/thestrategy.shtml> (visited on 01/15/2024).
- [12] InciWeb. Inciweb. URL: <https://inciweb.wildfire.gov> (visited on 01/15/2024).
- [13] Xuan Hien Le, Hung Ho, Giha Lee, and Sungho Jung. Application of long short-term memory (lstm) neural network for flood forecasting. *Water*, 11(7):1387, July 2019. DOI: 10.3390/w11071387.
- [14] Earthstar Geographics LLC. Earthstar geographics llc. URL: <https://www.terracolor.net/> (visited on 01/15/2024).
- [15] Stuart Matthews. Dead fuel moisture research: 1991–2012. *International journal of wildland fire*, 23(1):78–92, September 2013. DOI: <https://doi.org/10.1071/WF13005>.
- [16] NASA. Archive download - nasa — lance — firms. URL: <https://firms.modaps.eosdis.nasa.gov/download/> (visited on 01/15/2024).
- [17] NASA. Nasa. URL: <https://www.nasa.gov/> (visited on 01/15/2024).
- [18] NASA. Nasa — lance — firms us/canada. URL: <https://firms.modaps.eosdis.nasa.gov/usfs/map/> (visited on 01/15/2024).
- [19] National Park Service, U.S. Department of the Interior. NPS.gov homepage (u.s. national park service). URL: <https://www.nps.gov/> (visited on 01/15/2024).
- [20] Jim Nilsson and Tomas Akenine-Möller. Understanding ssim. *arxiv*, June 2020. DOI: <https://doi.org/10.48550/arXiv.2006.13846>.
- [21] NWCG. Nwcg — nwcg is an operational group designed to coordinate programs of the participating wildfire management agencies. URL: <https://www.nwcg.gov/> (visited on 01/15/2024).
- [22] U.S. Department of Agriculture Forest Service. Landfire map viewer. URL: <https://landfire.gov/viewer/> (visited on 2023).
- [23] U.S. Department of Agriculture Forest Service. Landfire program: home. URL: <https://landfire.gov/> (visited on 01/15/2024).
- [24] USDA Farm Production and Conservation Business Center. National agriculture imagery program - naip hub site. URL: <https://naip-usdaonline.hub.arcgis.com/> (visited on 01/15/2024).

- [25] Eftychios Protopapadakis, Ioannis Rallis, Anastasios Doulamis, Nikolaos Doulamis, and Athanasios Voulodimos. Unsupervised 3d motion summarization using stacked auto-encoders. *Applied Sciences*, 10(22):8226, November 2020. ISSN: 2076-3417. DOI: 10.3390/app10228226. URL: <https://www.mdpi.com/2076-3417/10/22/8226>.
- [26] Joe H Scott. *Introduction to Wildfire Behavior Modeling*. National Interagency Fuels, Fire, & Vegetation Technology Transfer, 2012. URL: [http://pyrologix.com/wp-content/uploads/2014/04/Scott\\_2012.pdf](http://pyrologix.com/wp-content/uploads/2014/04/Scott_2012.pdf) (visited on 01/15/2024).
- [27] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun WOO. Convolutional lstm network: a machine learning approach for precipitation nowcasting. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL: [https://proceedings.neurips.cc/paper\\_files/paper/2015/file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2015/file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf) (visited on 01/15/2024).
- [28] João Silva, João Marques, Inês Gonçalves, Rui Brito, Senhorinha Teixeira, José Teixeira, and Filipe Alvelos. A systematic review and bibliometric analysis of wildland fire behavior modeling. *Fluids*, 7(12), 2022. ISSN: 2311-5521. DOI: 10.3390/fluids7120374. URL: <https://www.mdpi.com/2311-5521/7/12/374> (visited on 01/15/2024).
- [29] TensorFlow. Tensor flow. DOI: 10.5281/zenodo.4724125. URL: <https://www.tensorflow.org/> (visited on 01/15/2024).
- [30] US Dept of Commerce, National Oceanic and Atmospheric Administration, National Weather Service. National weather service. URL: <https://www.weather.gov/> (visited on 01/15/2024).
- [31] Zhou Wang and A.C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, 2002. DOI: 10.1109/97.995823.
- [32] Xingbang Wu, Fangchen Hu, Peng Zou, Xingyu Lu, and Nan Chi. The performance improvement of visible light communication systems under strong nonlinearities based on gaussian mixture model. *Microwave and Optical Technology Letters*, 62(2):547–554, October 2019. DOI: 10.1002/mop.32080.