A Literature Review for Wildfires Prediction

Kaylee Pham, Macoto Ward, Tyler Poplawski
Saulo Rubio, David Shin, Lior Zlotikman, Sergio Ramirez*
{kaylee.pham.704,david.ward.761,tyler.poplawski.456}@my.csun.edu
{saulo.rubio.491,david.shin.904,lior.zlotikman.183,sergio.ramirez.754}@my.csun.edu
California State University, Northridge
Northridge, California

Figure 1: A firefighter works during the Creek Fire in Madera County, California on Sept. 7. Photographer: Josh Edelson/AFP via Getty Images

ABSTRACT

According to the incident record on http://www.fire.ca.gov, between 1-1-20 and 10-1-20 wildfires have burned 3.75 million acres, killed 29 people, and destroyed 8,169 structures in California. Within the top six largest California wildfires, there were five incidents that happened in 2020 [1]. This problem needed an urgent attention, because it is getting worse due to climate change. As computer science students and enthusiasts in machine learning, our goal is to create a reliable wildfire prediction model to help our community in solving this major problem. The first step is looking into different approaches from existing works and learn from them. In this literature review, we have gone through several articles and conference papers. The purpose was to find examples on machine learning algorithms that apply to a certain type of data and resulted in an accuracy score above 95 percent.

KEYWORDS

Wildfires Prediction, Machine Learning, Neural Network (NN), Support Vector Machine (SVM), Big data, Meteorological data, Satellite data

1 INTRODUCTION

While much research has been conducted on how to predict wildfires with great accuracy, wildfires are still a major ongoing issue due to climate change. Thus, a better model is still needed to predict the creation and spread of wildfires. This paper examines different tools, techniques, and datasets that were used in alternative works. This paper also evaluates different machine learning models and their effectiveness, different datasets used with the most accurate models, and the analysis concluded from each of these techniques. Machine learning technology is evolving at a rapid rate, thus only the most recent wildfire prediction technologies are covered in this paper. The goal of this paper is to analyze the most accurate and effective prediction of wildfires. First, we will cover the different machine learning models that can be applied to wildfire prediction,

*All authors contributed equally to this research.

COMP 490 Senior Design Project, California State University, Northridge,

next we will discuss the datasets used in the most accurate models, and lastly we will evaluate the findings from the previous sections.

2 MACHINE LEARNING MODELS

Machine learning (ML) models play a big role in wildfire prediction, especially when dealing with big data that is extremely large and growing exponentially with time. ML models come as the main key component in analyzing this type of data [11]. In the fourteen papers that we read, a variety of ML models have been used in the past for similar prediction systems. These models were: Support Vector Machine (SVM), Relevance Vector Machine (RVM), K-Nearest Neighbor (KNN), K-means, Decision Tree (DT), Boosted Regression Tree (BRT), Random Forest (RF), and Monte Carlo Search Trees (MCST). As well as some Neural Network (NN) models such as Long Short-Term Memory (LSTM), Back-Propagation (RPROP), Particle Swarm Optimized Neural Fuzzy (PSO-NF), and Extreme Learning Machine (ELM) [2–15]. Standing out from all these models are a few types of NN models and the SVM model, which have proven to produce high accuracy results in their predictions.

2.1 Neural Networks

Artificial neural networks, or for short just neural networks (NN), are a type of machine learning model that consist of a large collection of conjoined nodes that feed data to one another across edges. Edges are the connections between multiple layers of nodes and the network typically includes an input layer, one or more hidden layers, and an output layer. Each of these edges is associated with a weight, which can be refined to produce different transformations and increased accuracy of results. The weights can be split into two categories: excitatory being positive and inhibitory being negative [5].

The output of a given node is determined by a propagation function, which commonly involves summing the inputs via linear combination with the weights as coefficients in the equation. An activation function then determines whether or not the node passes its data onto the next layer by comparing the input against a threshold value. Neural network models can be used for many things, and can be applied to predict the spread of a fire based on environmental and meteorological data using models such as Long Short-Term Memory (LSTM) and Back-Propagation (BPNN).

In a research paper by Liang et al., three different neural network models were tested against each other, those being a BPNN model, a Recurrent NN model, and the LSTM model mentioned earlier. The goal was to determine which of these methods is best to build a prediction model in order to help firefighters and emergency personnel assess the risk and spread of a fire before it grows too large. The results of this paper conclude that the LSTM model produced the best predictions of the three, with an accuracy of 90.9% [6]. Contrary to this discovery, a conference paper by Lall and Mathibela used a slight modification of BPNN called the Resilient Back-Propagation algorithm (RPROP), which resulted in the system's overall performance having an accuracy of 97%, as well as 87% precision and 88% recall [5].

Both papers proved to have created successful wildfire risk prediction systems, however they also faced similar drawbacks such as ensuring the model was not overfitting the data. To avoid this, Liang et al. performed a multi-collinearity test on the data to remove factors that were proven to skew the model's interpretation rather than benefit it. They also made note that an overall limitation of the system was due to the modeling data coming from a single area [6].

2.2 Support Vector Machine

A Support Vector Machine (SVM) is a supervised machine learning model commonly used for classification and regression problems. The main objective of an SVM is to find the optimal hyperplane to classify the members of two different classes. The points closest to the hyperplane are called support vectors, and they serve as training points that define the maximum margin. The SVM uses a kernel function to transform high-dimensional data.

Some of the benefits of using an SVM are that it requires only a couple of sets of data for training and offers one of the most accurate methods when compared to other models. In the paper: Wildfire Predictions: Determining Reliable Models using Fused Dataset, the authors found that SVM's had a higher percentage of accuracy when compared to Decision Tree and K's Nearest Neighbor models [7].

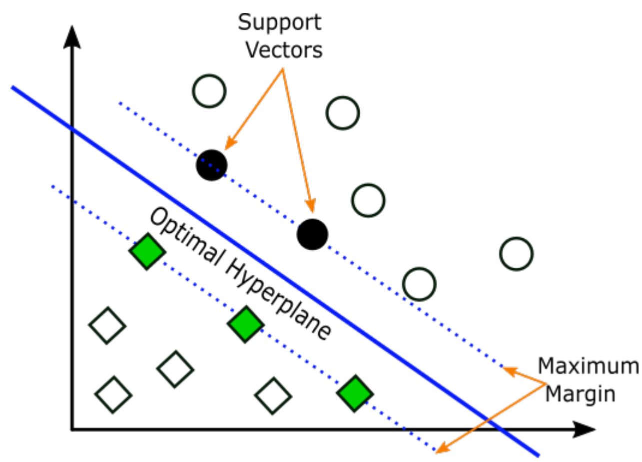

Figure 2: Support Vector Machine

3 DATASETS

Datasets in these studies generally use satellite data to deduce most of their information pertaining to spatial tracking of past fires and vegetative health for an area. A common source of satellite data was the Moderate Resolution Imaging Spectroradiometer (MODIS) which is an instrument on two satellites that are maintained by NASA. In addition, many studies used fire databases that already collected spatial information about past fires deduced from satellite data. Meteorological data such as wind, rainfall, surface temperature, and humidity amongst other things were collected either from satellite data or local meteorology stations.

3.1 Datasets used in SVM Models

One group used [7] a combination of US meteorological department (no longer under this name) data and the US forest fire database.

The meteorological data consists of temperature data as it pertains to weather, humidity, rain and snowfall levels among other data. The original database format could not be retrieved due to the inexistence of the original department. The US forest fire database included data such as geographical coordinates, area affected by fire, and severity, all in comma separated value (.csv) format. The two datasets were chosen due to their reliability.

Sayad et al. [11] used satellite data from MODIS to gather data on: crop health via normalized difference vegetation index (NDVI) data showing the photosynthetic activity of the foliage in an area, meteorological data to determine land surface temperature (LST), and thermal anomalies (given it is large enough) to deduce fire area.

3.2 Datasets used in NN Models

Another group [6] focused their study around fire prone forest in Alberta Canadian by using Canadian National Fire Database (CNFDB) and weather data from weather stations in the area, both datasets of which contained 30 years worth of data. The CNFDB contained information on the ignition and extinction dates and coordinates, as well as burn area and cause. Meteorological data included 11 elements pertaining to temperature, rain, snowfall, and wind speed.

Storer et al. [12] focused on a National Park in Portugal, using a combination of manually collected data in the form of a 9x9 grid of the park, meteorological data from a station inside the park including wind speed among other weather factors, and Fire Weather Index data from a Canadian database which yielded moisture levels, Initial Spread Index (correlated to fire velocity). The study had a notably small dataset from the first two sources of 3 years, which the author described as a hurdle when training a model.

Tien Bui et al. [13] built their model around data from a fire database, topographic maps, satellite data, local climate data, and data from locals authorities. The fire database detailed 430 fires in the region courtesy of the Department of Forest Protection. The other 10 factors they used include: NDVI, geographical factors from maps (slope, elevation), data from locals on land use (distance to road, and distance to residential area), and some meteorological factors (temperature, wind speed, rain levels).

4 EVALUATION

The assessment of our research papers demonstrated that there are a wide range of neural networks (NN), machine learning (ML) algorithms/models and datasets used for wildfire predictions. In this section, we outline the summary of results from our papers and our analysis thus far.

4.1 Results

The research papers outlined several approaches to improve the efficiency and/or accuracy of fire predictions. The approaches that yielded positive results were Support Vector Machine, Extreme Learning Machine, Random Forest, Long Short Term Memory, Resilient Backpropagation Algorithm, hybrid artificial intelligence, Particle Swarm Optimization, and the 2D Convolutional Neural Network.

The Support Vector Machine (SVM) was utilized in three papers out of the fourteen reviewed. When evaluated against the K-Nearest Neighbor, Decision Trees, and K-Mean Clustering models, the SVM had a higher accuracy in both binary and multi-class classifiers [7]. In another experiment, the SVM was compared to an ANN and produced a 97.48% accuracy compared to the ANN which produced an accuracy of 98.32% [11]. Yet in another experiment, the SVM was compared against the Regression Trees, Artificial Neural Network, and Random Forest, but the results concluded to be comparable to a Binary Logistic Regression model [10].

The Extreme Learning Machine was tested against a pre-existing system, the Fire Dynamic Simulator, and a Generalized Regression Neural Network. It was concluded that the ELM had the shortest calculation time, which led to less overhead [4].

The Random Forest model was tested against Regression Trees, Artificial Neural Network, Boosted Regression Trees and Support Vector Machine models and had the highest accuracy and required less predicative variables for high performance, and had more cartographic outputs [10].

The Long Short Term Memory was evaluated with the Backpropagation and the Recurrent Neural Network and had the highest accuracy results of 90.9% [6].

The Resilient Backpropagation Algorithm was able to categorize fire risk into four categories (low, medium, high, and extreme) with a 97% accuracy and 87% precision [5].

A hybrid artificial intelligence was concluded to capable of producing the best result in predicting fires [13].

The Particle Swarm Optimization produced exceptional results by reducing the Root Mean Squared Error by 260% when it was utilized to train an artificial neural network [12].

The 2D-Convolutional Neural Network was capable of producing an image with different color schemes to pinpoint point-of-interests and their predicted values of fire combustion [15].

These results gave us a general direction and approach that we should take as a group but as notable as the results are, there are pros and cons to each approach as well as constraints.

4.2 Analysis

After reviewing the results we noticed that there were many possibilities to derive a solution to wildfire predictions. Some approaches yielded more favorable results than others, but every approach had their own set of disadvantages and limitations. Some of the disadvantages are time to implement and calculation time. As for the limitations, we noticed that a common limitation is data or more specifically the updating of data. Since we will be predicting wildfire predictions, climate conditions will be essential to calculate the prediction. However, because the weather is extremely volatile it is difficult to retrieve live updates. Although, there were experiments that retrieved data from weather stations, we will have to research if there are weather stations that will provide the data for us and more importantly the locations of the weather stations. In short, our research has shown that there are a lot of factors that will go into this project, but it is possible to improve pre-existing models and systems.

5 CONCLUSION

With the abundant amount of research papers that were read and discussed, it is valid to say that in the realm of fire prediction, accurate results have been achieved. This is to say that there are several valid ways to drive prediction models to their highest potential. Throughout the various papers, we consistently came across the usage of geological data, topographic data, burn index data, and data mining applications. When this type of data is used in unison with machine learning models, the results proved to be more precise. For every unique paper that was read, the datasets were astonishingly similar. Moreover, the datasets were accurate and useful enough to use in a predictive model. The application of Neural Networks and Support Vector Machines were found to be the most widely used, and are some of best learning methods for this domain. With the increasingly positive trend towards generating correct predictions, most papers noted that there was a goal in mind; a goal where a model's usage will not only be limited to its targeted geographical location, but also be flexible enough to have usage in a variety of locations. As such, there was a notable issue that what worked for one area may not work for another. While there does not exist a specific set of tools that will completely work for any location, the data to make predictions exists and machine learning models await implementation. Therefore, while it is important to keep in mind that the field of study is still evolving and is not as efficient as it can be, when machine learning methods and highly usable data are paired together, they allow for highly accurate and useful wildfire prediction results.

REFERENCES

[1] 2020 (accessed October 1, 2020). *2020 Incident Archive.* https://www.fire.ca.gov/incidents/2020/

[2] Sriram Ganapathi Subramanian and Mark Crowley. 2018. Combining MCTS and A3C for Prediction of Spatially Spreading Processes in Forest Wildfire Settings. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10832 LNAI, d (2018), 285–291. https://doi.org/10.1007/978-3-319-89656-4_28

[3] HPE/SGI ICE XA System (NCAR Community Computing). 2017. Computational and Information Systems Laboratory. Cheyenne.Computational Information Systems Laboratory. (2017). https://doi.org/10.5065/D6RX99HX

[4] Chun Jiang and Qinqin Fan. 2019. State Parameter Prediction of Fire Based on Extreme Learning Machine. *2019 IEEE Symposium Series on Computational Intelligence, SSCI 2019* (2019), 1108–1114. https://doi.org/10.1109/SSCI44817.2019.9003049

[5] Shruti Lall and Bonolo Mathibela. 2017. The application of artificial neural networks for wildfire risk prediction. *International Conference on Robotics and Automation for Humanitarian Applications, RAHA 2016 - Conference Proceedings* (2017). https://doi.org/10.1109/RAHA.2016.7931880

[6] Hao Liang, Meng Zhang, and Hailan Wang. 2019. A Neural Network Model for Wildfire Scale Prediction Using Meteorological Factors. *IEEE Access* 7 (2019), 176746–176755. https://doi.org/10.1109/ACCESS.2019.2957837

[7] By Hariharan Naganathan, Sudarshan P Seshasayee, Jonghoon Kim, Wai K Chong, and Jui-sheng Chou. 2016. Wildfire Predictions : Determining Reliable. 16, 4 (2016).

[8] Rylan Perumal and Terence L.Van Zyl. 2020. Comparison of Recurrent Neural Network Architectures for Wildfire Spread Modelling. *2020 International SAUPEC/RobMech/PRASA Conference, SAUPEC/RobMech/PRASA 2020* 2020-Janua (2020). https://doi.org/10.1109/SAUPEC/RobMech/PRASA48453.2020.9078028

[9] David Radke, Anna Hessler, and Dan Ellsworth. 2019. Firecast: Leveraging deep learning to predict wildfire spread. *IJCAI International Joint Conference on Artificial Intelligence* 2019-Augus (2019), 4575–4581. https://doi.org/10.24963/ijcai.2019/636

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

[11] Younes Oulad Sayad, Hajar Mousannif, and Hassan Al Moatassime. 2019. Predictive modeling of wildfires: A new dataset and machine learning approach. *Fire Safety Journal* 104, September 2018 (mar 2019), 130–146. https://doi.org/10.1016/j.firesaf.2019.01.006

[12] Jeremy Storer and Robert Green. 2016. PSO trained Neural Networks for predicting forest fire size: A comparison of implementation and performance. *Proceedings of the International Joint Conference on Neural Networks* 2016-Octob (2016), 676–683. https://doi.org/10.1109/IJCNN.2016.7727265

[13] Dieu Tien Bui, Quang Thanh Bui, Quoc Phi Nguyen, Biswajeet Pradhan, Haleh Nampak, and Phan Trong Trinh. 2017. A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. *Agricultural and Forest Meteorology* 233 (feb 2017), 32–44. https://doi.org/10.1016/j.agrformet.2016.11.002

[14] Dieu Tien Bui, Hung Van Le, and Nhat Duc Hoang. 2018. GIS-based spatial prediction of tropical forest fire danger using a new hybrid machine learning method. *Ecological Informatics* 48 (nov 2018), 104–116. https://doi.org/10.1016/j.ecoinf.2018.08.008

[15] Guoli Zhang, Ming Wang, and Kai Liu. 2019. Forest Fire Susceptibility Modeling Using a Convolutional Neural Network for Yunnan Province of China. *International Journal of Disaster Risk Science* 10, 3 (2019), 386–403. https://doi.org/10.1007/s13753-019-00233-1