Wildfire Prediction

Theegela Pavan Kumar, Pulavarthy Harshik, Aryan Sridharala, Shanmukh Chowdary Morampudi
Computer Science Engineering
Bennett University
Greater Noida, India
E19CSE373@bennett.edu.in, E19CSE372@bennett.edu.in, E19CSE162@bennett.edu.in, E19CSE374@bennett.edu.in

***Abstract*—A wildfire is an uncontrolled fire that burns in the vegetation of the wildlife, usually in rural locations. Wildfires have been burning for hundreds of millions of years in forests, grasslands, savannas, and other habitats. Forest fires have become more common in recent decades as a result of deforestation and climate change. We propose a novel, cost effective model to predict the wildfires in the forests and prevent or minimize the damages that are caused by wildfires with the help of previous data about wildfires in different locations and climatic conditions**

***Index Terms*—Machine Learning, Decision Tree, ADA Boost, XGBoost, KNN**

I. Introduction

A wildfire is a natural disaster fire that burns in the vegetation of the wildlife, usually in rural locations. Wildfires have been known to burn in forests, grasslands, savannas, and other environments. For hundreds of millions of years, this has been the case. This is not the case. It is one that is exclusive to a specific continent or environment.one of the United Nations' key priorities is to defend and protect the environment from natural and man-made disasters. Wildfires have had an impact on climate change. According to numerous reports by the United Nations. When the temperatures rise higher, then they will have a significant impact on Changes in the climate. Temperatures around the world are rising, increasing the risk of wildfires.

As you can see how many problems are caused by wildfires many people, animals, species in some situation are loosing their habitat, loosing home is such a heinous thing that can happen to anyone despite many measures taken by governments and by WHO organization many wildfires are still occurring we all know about 2019-2020 bushfire season in Australia also known as black summer it effected both environmentally and ecologically, The Air Quality Index of Sydney has raised to 2552 which is more than a dozen times have been raised over the dangerous level of 200. According to NASA, approximately 25000 koalas were dead, or around 50 percent of the overall population of the overall species on the island. A 2020 research projected that the flames displaced or killed at least 3 billion terrestrial vertebrates alone. In India, we experienced numerous tragic wildfires as well. Large forest fires erupted in a number of locations across the Bandipore National Park in India's Karnataka state in February 2019. In just five days, an estimated 10,920 acres were burned. The fire also moved to Tamil Nadu's Mudumalai forest region, destroying roughly 40 acres. The fire was started by a "act of sabotage," according to Karnataka's top forest official. The Uttarakhand forest fires were one of India's worst wildfires in 2016. The fires, which were mostly started in pine forests on the sub-Himalayan region's slopes, created clouds of smoke. The Indian Air Force Mi-17 helicopters and the National Disaster Response Force used "Bambi buckets" to put out the fires. According to the forest department, 3,500 hectares (8,600 acres) of woodland were burned. By the 2nd of May, about 1,600 fires had been identified and were under control. The rains on May 3 aided in reducing the severity of the fires. When we heard about these wildfires we got chills down the spine and we felt responsible to do something from our side to help to minimise these wildfires.

Wildfires will be often occurring due to human activity or a natural phenomenon such as lightning, poor historic management and so on. These disasters can happen anywhere at any time Extremely dry conditions, such as drought, and high winds enhance the risk of wildfires. Wildfires can affect transportation, power, and gas supplies, as well as the water supply. Depending on the size, speed, and proximity of the fire, as well as whether or not the population has been given advance warning to evacuate, wildfires or forest fires can have a substantial impact on mortality and morbidity. They also have a negative impact on air quality since wildfires generate a large quantity of CO2 and greenhouse gases, which raises global temperatures. The particles from the fires go a long way and pollute the air. These particles also settle on snow, causing more sunlight to be absorbed. A climate feedback loop is a phenomena that causes climate conditions to deteriorate over time. Approximately 6.2 million people are estimated to be affected by the World Health Organization's reports. These particles also settle on snow, causing more sunlight to be absorbed. A climate feedback loop is a phenomena that causes climate conditions to deteriorate over time. Approximately 6.2 million people are estimated to be affected by the World Health Organization's reports. The wildlife environment's occupants have all been displaced. If the fires are close enough to human housing, people frequently lose their homes as well Vegetation is now obsolete If this location is near a farm or other people's food,

vegetation is no longer necessary. . Repairing the damages and re-building homes and vegetation areas costs millions of dollars. The earth along the wildfire's path has been completely depleted. The soils in the forest are made up of decomposing nutrients and debris that contain a variety of natural components that contribute to the creation of the earth. When a wildfire burns through this soil, it becomes too hot, and the nutrients are lost forever. Trees and plants will also be lost. Plants and trees contribute to the world's oxygen production. We have less clean air to breathe as there are fewer trees and plants. The animals that did survive no longer have anything to eat due to the lack of plants and trees. Animals will be the ones that suffer the most as a result of this calamity. Birds, squirrels, bunnies, and other nature species will no longer be a part of this world, which is a sad but genuine fact. The extinction of certain animals and other forest species can be triggered by the loss of animals. Workers in the agricultural field have lost income and jobs as a result of the wildfire, which damaged their field produce and animals. When people lose their jobs, the economy suffers, making it more difficult to recover. Wildfires also release large levels of mercury into the air, which can cause speech, hearing, and movement impairments, as well as muscle weakness and vision difficulties in people of all ages. The mixture of tiny solid and liquid particles suspended in air called particle pollution is one of the most dangerous form of air pollution. Particles emitted by the Wildfire Smoke will be even smaller than one fourth of the diameter of a person's hair, they can get into lungs and stay there for a long time.

II. Related Work

j Alexander. predicted the wildfire with accuracy of 60 percent by using Random forest method. He stored the data SQLite database with four labels that are causes for wildfires i.e., natural causes, arson, negligence and infrastructure. [1]

Current Wildfire determining technologies, For example Wagner's (1987) famous WildFire Weather Prediction technology from Canada, works on data provided weather prediction systems on the ground (nearly 750 in this case). The data is taken as an input and then analysed using accurate-form mathematical equations to produce the Fire Weather Index (FWI), which represents the intensity of Fire in numerical for for theWildFire Weather Prediction technology of Canada. But our research also aims to build cost-effective prediction systems.

Fire weather has an important role in deciding whether or not a fire will start, how quickly it will spread, and where it will spread.There have been very few publications that use machine learning to predict fire weather and hazard. [2]SOMs were originally utilised to investigate the synoptic climatology of severe fire weather in the southwestern United States (Crimmins 2006). (US). Crimmins discovered three significant patterns involving southwesterly flow and substantial geopotential height gradients, which were linked to more than 80 percent of the extreme fire weather days as measured by a fire weather index. [3]SOMs were utilised by Nauslar et al. (2019) to calculate the timing of the North American Monsoon, which has a significant impact on the length of the active fire season in the southwestern United States. [4]SOMs were also utilised by Lagerquist et al. (2017) to anticipate severe fire conditions in northern Alberta, Canada.

Some state and federal fire agencies utilise the Wildland Fire Decision Support System (WFDSS), a web-based geospatial fire management portal, to manage and document major fires [5][O'Connor et al., 2016]. WFDSS records daily fire advancement using fire perimeter mapping and a modelling framework that incorporates real-time short-term weather forecasts, but it does not try to anticipate future development. WFDSS is meant to assist flexible and informed decision making by fire management regarding the current status of a fire by eliminating forecasts and just presenting the user with existing facts.

Finney, a Research Scientist of the US Forest Service, is the main creator of Farsite, the most widely implemented fire modelling tool, which uses a two-dimensional (2D) deterministic fire development model to forecast fire perimeters [6] [7][Finney, 1994; Finney, 1998]. Landscape and vegetation features, as well as meteorological data, are utilised to produce anticipated future fire perimeters for a certain time period using physics and mathematical-based models. A significant number of location-based variables are required for the programme, including a manually produced landscape file1, weather, winds, fuels, fuel moisture, fire spread rate modifications, and other optional inputs.

Other machine-learning-based methods to forest fire prediction exist, but they do not exclusively focus on fire prediction. [8]Wijayanto et al. investigated the link between fire and human activities using an Adaptive Neuro Fuzzy Interface System to determine if a fire alarm trigger is an FP (false positive) (Wijayanto et al. 2017). Given meteorological station data, [9]Cortez and Morais utilised logistic regression, decision trees, and support vector machines to estimate the probable fire area [**?**](Cortez and Morais 2007). FireNet detects and annotates fire boundaries in drone video feeds in real time [10](Doshi et al. 2019). Also [11](Radke, Hessler, and Ellsworth 2019) developed FireCast to anticipate how a fire would spread in the near future.

III. Methodology

A. Data Pre-processing

As the dataset contains 5 different types of datasets, the preprocessing is done for all the five datasets separately.

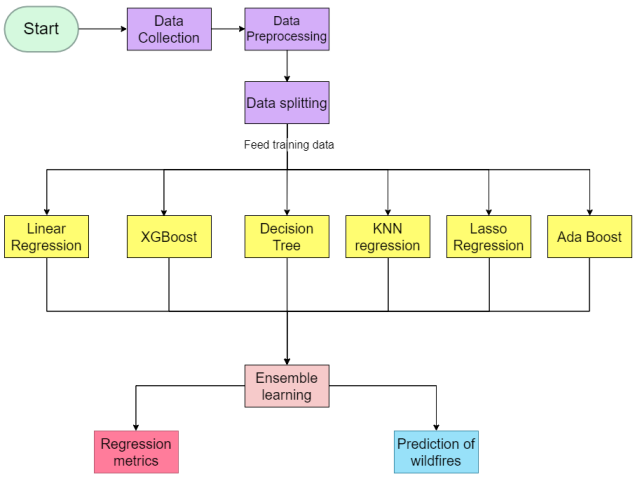

Fig. 1. block diagram

1) Historical Wildfires dataset: Changed all the dates into datetime format. And then checked for null values. After running the command, we observed standard deviation confidence and variance confidence have null values, so we replaced these null values with zeroes. Then removed the duplicates.

2) Historical Weather: Changed all the dates into datetime format. There are no null values in this dataset. There are some parameters like precipitation, humidity, temperature etc., they are placed in rows, then rearranged such that the parameter values become columns contains values 'area', 'min', 'max', 'mean' and '2ndmoment' while keeping distinct values for 'Date','Region' and 'area'. This is done using 'pivottable'. For better understanding we renamed the column names, for example mean of temperature is changed into temperaturemean. And then sorted the columns in such a way that all the parameter values of same kind comes next by next. Now the data had NULL values because of the rearrangements, and if we cross check them they have no reading in the original database. So, replaced them with zeroes. And then merge wildfire's data with weather data.

3) Historical Weather Forecasts: Changed all the dates into datetime format. Then repeated the process that is done in historical weather dataset preprocessing, rearranging data using pivottable, again renamed them as mentioned above in an understandable way. And then sorted the columns in such a way that all the parameter values of same kind comes next by next. Now the data had NULL values, and if we cross check them they have no reading in the original database, so replace them with zeroes. Now merge this data with the already merged data.

For historical vegetation index datset, changed all the dates into datetime format and there are no null values. For landclass dataset, there are no null values and no preprocessing required.

B. Algorithms used and Proposed methodology

For predicting the estimated fire area, we built 6 different models. They are linear regression, lasso regression, decision tree, KNN regression, XGBoost and, AdaBoost. Target variable is the estimated fire area. The dataset is divided into two parts training set and testing set. The test set size and training set size have a lot of impact on the accuracy of the model. The accuracy changes when the training dataset size changes. So, it is important to divide the data efficiently. For this dataset, the data division is done in 75:25 ratio. 75 percent of dataset is considered for training and remaining 25 percent is given for testing dataset. This is concluded after verifying by replacing different values. And random state is taken 42.

1) Linear regression: Using scaler transform, normalization of data is done. Then feed the training data to the linear regressor. And examined the importance of features using regression coefficients.

2) Lasso Regression: Lasso regression is done by taking alpha value 0.0001. By examining the results compared to linear regression, they are almost equal.

3) KNN Regression: K-NN algorithm assumes the similarity between the test data and available data and put the new data into the category that is most like the available categories. For KNN regression the nneighbors impacts the accuracy of the model. So first, test the nneighbors using different values. The accuracy is better when nneighbors value is taken as 4.

4) XGBoost Regression: XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses gradient boosting framework. Model is trained using XBRegressor, taken all parameters with default values. Tried taking different parameters with different values. And the model performed good with parameters with default values only.

5) Decision Tree: Decision tree regressor is used to build the model. The parameters taken are maxdepth and randomstate. For maxdepth we tried all the values from 6 to 24. After examining the outputs, maxdepth=19 gave the better results and randomstate is 42. And examined the importance of features using featureimportance.

6) AdaBoost: AdaBoost is a decision-tree-based ensemble Machine Learning algorithm that uses gradient boosting framework. AdaBoost is such an amazing algorithm because it learns from previous mistake and tries to get best accuracy.

By examining all the regression models and their outputs, we can see that AdaBoost gave the best results. The AdaBoost accuracy is highest among all, followed by Decision Tree.

Hence AdaBoost is the proposed algorithm for predicting fires in forest.

IV. Result and Analysis

A. Hyper parameter Analysis

The model is trained using 6 training methods i.e., Linear regression, Decision tree, K-NN, XGBoost, AdaBoost, Lasso regression. 25 percent of data is separated a test data with random state value 42. In all the above 6 methods AdaBoost gave better accuracy of 95.1 percent. AT first we trained the model using linear regression. Results given by the model both RMSE and accuracy are not as we expected. So, we then normalized the data and applied linear regression again, then we saw the improved results with an accuracy of 49.2 percent and low root mean squared error than before. Feature importance graph of linear regression is plotted and shown in fig 2

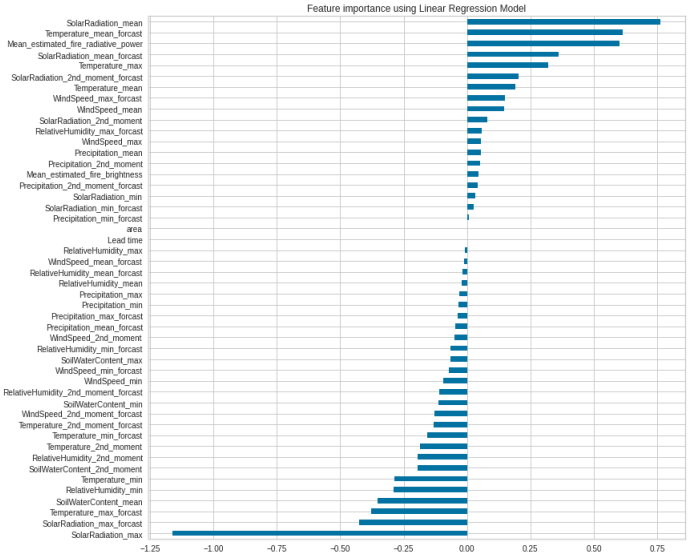

Fig. 2. Linear feature importance diagram

Even though the accuracy got improved it's not up to mark, so then we used the K-NN algorithm to train the model. K-NN algorithm assumes the similarity between the test data and available data and put the new data into the category that is most like the available categories. It is tested using different k values. Graph between k value and MSLE (Mean square logarithmic error) is plotted, shown in fig.3. The least MSLE value is observed at value of 4. The accuracy predicted by K-NN algorithm with k= 4 is 60.02 percent which is better than linear regression.

Next, we used Lasso regression to train the model. But Lasso regression didn't increase the accuracy it even got high MSLE value compared to previous models. This model got an accuracy of 49.44 percent.

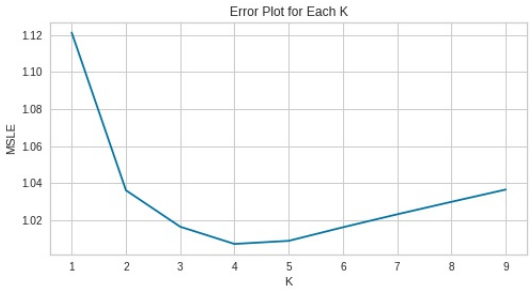

Fig. 3. Graph between k value and MSLE

Then we trained the model using XGBoost method. XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses gradient boosting framework. XGBoost is such an amazing algorithm because it learns from previous mistake and tries to get best accuracy. As expected, we got the highest accuracy and least MSLE value better than all the other algorithms used till now. XGBoost Feature importance graph is plotted and shown in fig.4.

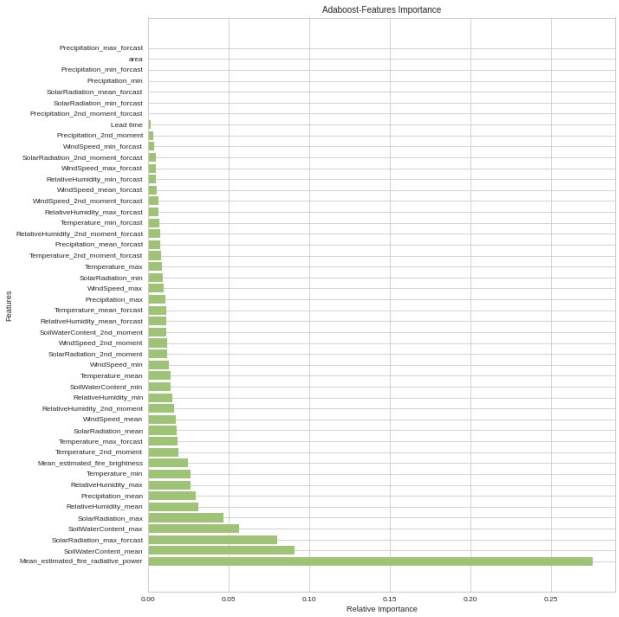

Fig. 4. XGBoost feature Importance diagram

Next, Decision tree is used to train the model. In decision tree model is trained by splitting the source set into subsets based on an attribute value test. Model is trained using different values of max depth. The graph plotted between max depth and MSLE is plotted shown in fig.6, it is observed that at max depth value of 19 MSLE value is low. So the model is trained at max depth value of 19, and it gave the accuracy of 84.04 percent which is the highest value seen. After training the model feature importance is plotted in graph shown in fig.5.

FInally, AdaBoost is used to train the model. As expected AdaBoost performed very well both in terms of accuracy and MSLE value. It gave the accuracy of 95.1 percent which is the highest value seen and 0.12 MSLE value which is the least error till now.

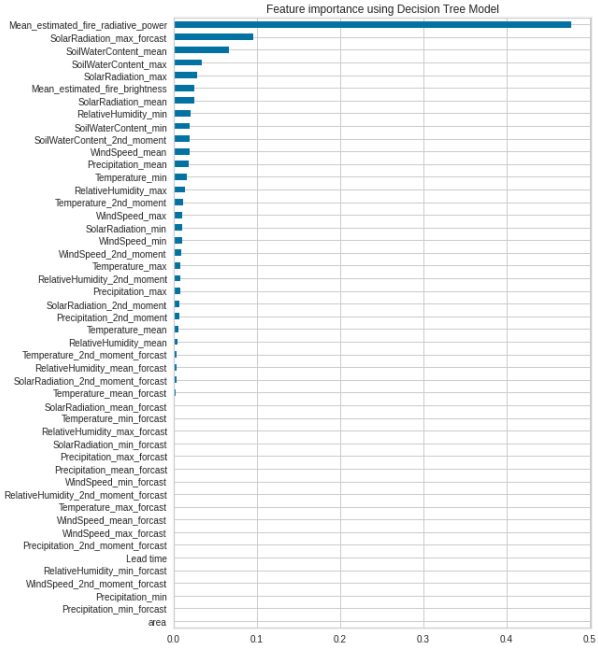

Fig. 5. Decision Tree feature importance diagram

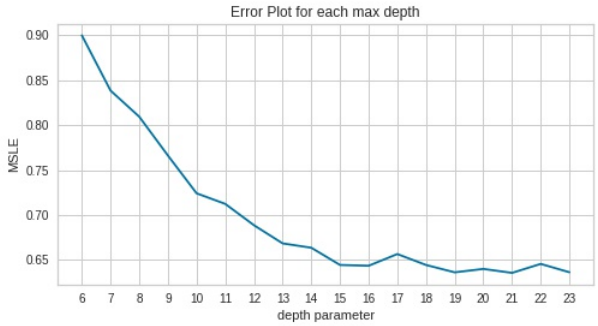

Fig. 6. Graph between max depth and MSLE

B. Comparison

The model is trained using 6 training methods i.e., Linear regression, Decision tree, K-NN, XGBoost, AdaBoost Lasso regression. . Among all the above 6 methods ADABoost gave the highest accuracy of 95.1 percent with least MSLE value. And then Decision Tree gave the highest accuracy of 84.04 percent with least MSLE value after AdaBoost. In table 1 metrics for all the models is mentioned. Among all the models AdaBoost has the least MSLE value followed by Decision tree, XGBoost, KNN, Lasso regression and Linear regression. In terms of error Linear regression has poor performance when compared to other models. And in terms of accuracy order also follows the same pattern. AdaBoost performed exceptionally well followed by Decision tree. Lasso regression and Linear regression shown the almost same reults in terms of both accuracy and error value. A Graph is plotted between accuracy of different models shown in the fig 7

Models	MSLE	Root MSLE	R2 Score	Accuracy
Linear Regression	1.283386784	1.132866622	0.494151378	49.4151
Decision Tree	0.404990375	0.636388541	0.840372501	84.0373
KNN	1.014297682	1.007123469	0.600213208	60.0213
Lasso Regression	1.282874821	1.132640641	0.494353168	49.4353
XG Boost	0.550297643	0.741820492	0.783099446	78.3099
ADA Boost	0.121948631	0.349211441	0.951933784	95.1934

TABLE I
CAPTION

Previously there are some researches done on the wildfire prediction. Many researches have failed to get a descent accuracy. Many of them fell behind 70 percent accuracy. J Alexander. [1] predicted the wildfire with accuracy of 58 percent by using Random forest method. He stored the data SQLite database. Then he has done some improvements and labelled the data as four different parts that are causes for wildfires i.e., natural causes, arson, negligence and infrastructure. Then he applied Random forest and observed the accuracy of 68 percent. Compared to the previous works our results are far ahead of them. Not only prediction of wildfire but also estimated wildfire area also gave very good accuracy.

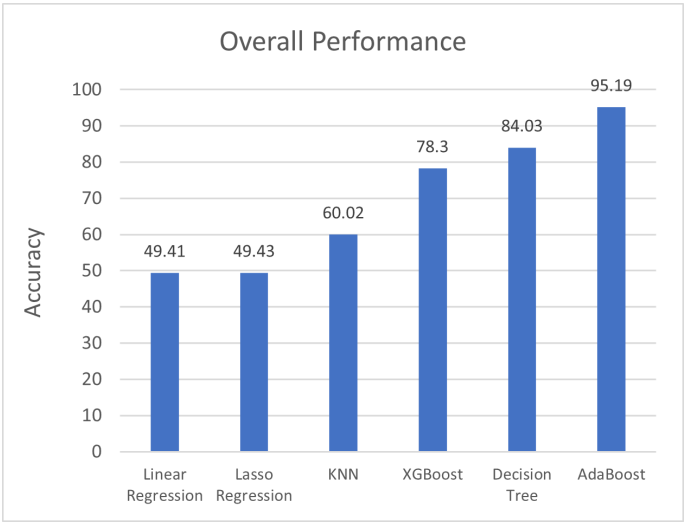

Fig. 7. Overall Performance of the models

REFERENCES

[1] J. Alexander, "Wildire prediction using improved random forest method."

[2] C. M. 2006., "Synoptic climatology of extreme fire-weather conditions across the southwest united states. int. j. climatol. 26(8): 1001–1016."

[3] B. T. J. . K. M. L. Nauslar N J, Hatchett B J and M. J. F, "Impact of the north american monsoon on wildfire activity in the southwest united states. int. j. climatol. 39(3): 1539–1554."

[4] F. M. . W. X. Lagerquist R. and M. G. 2017., “Automated prediction of extreme fire weather from synoptic patterns in northern alberta, canada. can. j. for. res. 47(9): 1175–1183.”

[5] M. P. T. Christopher D. O’Connor and F. R. y Silva., “Getting ahead of the wildfire problem: quantifying and mapping management challenges and opportunities. geosciences, 6(3):35, 2016.”

[6] M. A. Finney., “Farsite: a fire area simulator for fire managers. the biswell symposium: fire issues and solutions in urban interface and wildland ecosystems, pages 55–56, feb 1994.”

[7] I. C. G. Mark Finney and C. W. McHugh., “Modeling containment of large wildfires using generalized linear mixed-model analysis. forest science, 55(3):249–255.”

[8] S. . O. K. . N. D. Wijayanto, A. K and Y. . Herdiyeni, “Classification model for forest fire hotspot occurrences prediction using anfis algorithm. iop conference series: Earth and environmental science 54(1): 012059.”

[9] P. Cortez and A. . Morais, “A data mining approach to predict forest fires using meteorological data.”

[10] J. Doshi, “Firenet: Real-time segmentation of fire perimeter from aerial video.”

[11] H. . A. Radke, D. and E. . D. 2019., “Firecast: Leveraging deep learning to predict wildfire spread. in proc. of ijcai, 4575–4581. international joint conferences on artificial intelligence organization. doi:10.24963/ijcai.2019/636. url .”