FOREST FIRE PREDICTION

Sai Pavana Niharika Kali
dept. of Computer Science(Masters)
University of Central Missouri
SXK06030@ucmo.edu

Pallavi Harishchandra Meher
dept. of Computer Science(Masters)
University of Central Missouri
SXN99850@ucmo.edu

Nithisha Majety
dept. of Computer Science(Masters)
University of Central Missouri
NXM09780@ucmo.edu

ABSTRACT— **Forest fires are one of the major natural disasters, which is creating economic and ecological damage and endangering human lives. It not only damages the forest wealth but also affects the human lives. It causes ecological imbalance. Future effects of forest fire are anticipated to be lessened via forest fire prediction. Future effects of forest fire are anticipated to be lessened via forest fire prediction. Detecting the fire at the early stage can prevent the forest fire to occur. If the forest fire is detected in early stage and quick action is taken, then the damage caused by forest fire can be minimized. We have taken the latest dataset for then details when fire was happened and not happened in previous atmospheric conditions. The suggested system procedures employed meteorological variables including temperature, rain, wind, and humidity to anticipate when a forest fire will occur. There are some other features which have been calculated based on the above primary features which have been taken. This paper proposes the SVM (Support Vector Machine), SVM with SMOTE (Synthetic Minority Oversampling technique) which helps to overcome the overfitting problem and Random forest regression algorithms to predict the forest fires based on the latest dataset having all the required features. By performing all the above techniques on the data we find the accuracy of each algorithm and consider the highest accuracy as the best technique for predicting the forest fire. This can be used to predict the data before the forest fire to happen or we can use the pre-defined values or pre-defined datasets to analyze the previous data's of the forest fire that is happened. In this we can predict the burnt area, or the day, or the month that the fire may occur based on the current parameters and the data can be visualized.**

Keywords: SVM, SMOTE, Random forest regression, forest fire, prediction

I. INTRODUCTION

One major environmental concern is that the prevalence of forestfires (also known as wildfires), that have an effect on forest preservation, cause economic and ecological harm and cause human suffering . annually a lot of forest hectares (ha) are destroyed all round the world. quick detection could be a key component for a successful firefighting. The scale of the losses might be nearly incomprehensible when a wildfire flames out of control. The cause of the forest fires are many: natural causes i.e. temperature, wind direction, moisture level etc. and human related like the burning by grazers and gatherers of forest produce, shifting cultivation, fires to ward off wild animals . The cost of disaster may be millions of trees, in addition to lose of structures, animals (wild and farm), and human life. Forest fire lead to global warming, soil erosion, ozone layer depletion and the loss of livelihood of forest products. Wildfires may only be prevented from destroying forests if they are foreseen in advance.

A. ***OBJECTIVE OF THE PROJECT**:* To predict or detect the occurrence of forest fire using the support vector machine and by visualizing the data in python, Early stage detection of the forest fire to prevent the occurrence of the forest fire.
B. ***SIGNIFICANCE**:* The importance of the forest fire prediction is, by calculating the probability of fire using the meteorological conditions will help us to find whenever there are chances of fire accident it will notify us by the visual representation so that we can alert at that time and prevent the occurrence of fire in forest.
C. ***FEATURES**:* The latest features of the forest fire prediction we have taken are Month, day, Wind, Area, Rain, Temperature, FFMC(Fine fuel Moisture Code), DMC(Duff Moisture Code), DC(Drought code), ISI(Initial spread Index), RH(Relative humidity)
D. ***DETAIL DESIGN OF FEATURES**:* From the figure design we understand that FFMC(Fine fuel Moisture Code) is based on the factors i.e, Rain, Relative Humidity, Temperature, and Wind. DMC(Duff Moisture Code) is based on the factors like Rain, Relative humidity, Temperature. DC(Drought code) is based on Rain and Temperature factors. Based on all these factors we predict the fire occurrence.

Github link: https://github.com/pallavi234/ML_Project

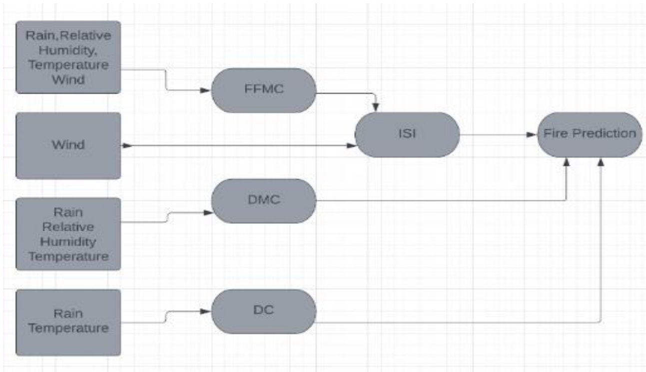

Figure 1

II. MOTIVATION

The cause of the forest fires are many: natural causes i.e. temperature, wind direction, moisture level etc. and human related like the burning by grazers and gatherers of forest produce, shifting cultivation, fires to ward off wild animals . The cost of disaster may be millions of trees, in addition to lose of structures, animals (wild and farm), and human life. Forest fire lead to global warming, soil erosion, ozone layer depletion and the loss of livelihood of forest products. The only way of protecting forest from wildfires is their early prediction . satellite system based on monitoring form satellites: Usually for monitoring the crown fire satellite based system is more useful. When a satellite connection is unavailable or problematic (for example some points are invisible to satellite). There are two ways to construct a wireless network for fire detection in a forest. One is a satellite-based system that uses satellite monitoring: Typically, a satellite-based system is more effective for keeping an eye on the crown fire. An option to using a satellite link to monitor a forest fire is to utilize an observation place instead (for example, some points are unseen to satellite, notably floors in confined forests, and there could be several such blind spots are identified in a forest; in India, most flames occur at the surface). Traditionally, direct human observation was used to guide this process. Here, we are keeping an eye on the forest fire using a camera. Camera data's are sufficient only for the detection of forest fire, but for the prediction we need some more information. We are placing sensor nodes in the woodland region to achieve this goal. Sensor nodes are capable of measuring the parameters like temperature and humidity. For the purpose of forecasting natural forest fires, these characteristics are crucial.

III. MAIN CONTRIBUTIONS & OBJECTIVES

The objective of this study is to simulate the likelihood that a fire would occur in relation to the day, the month, specific meteorological variables here for the dataset The question we are attempting to answer is how confident we are that a forest fire will occur, not whether it will occur or not.
The following are the steps where we need to implement accordingly
1. Import library and dataset
2. Data pre-processing
3. Implementing the support vector machine algorithm, SVM using SMOTE, Linear Regression and Random forest Regression
4. Prediction using the algorithms and then finding the accuracies.
5. Visualization of data.

IV. RELATED WORK

The pre-processing or purification of data is a crucial stage in the job of a machine learning engineer, and the great majority of them invested a lot of time and effort before creating a model from scratch. Data pre-processing methods include removing unwanted or noisy data, treating missing values, and detecting outliers.

To get more accurate findings, some background noise should be eliminated before feature extraction. The image is transformed from RGB to gray scale and then smoothed with a Gaussian filter. To gauge the amount of green color existing in the image, the image is first transformed to the HSV color space.

A crucial part of controlling forest fires is forest fire prediction. This is a serious environmental issue that threatens a landscape of endangered species and results in ecological catastrophe. Natural resources that decrease resources like water that contribute to climate change and water pollution as well as resources that disturb the ecosystem's stability and increase the risk of other natural hazards. A crucial component for managing such occurrences is fire detection. It is anticipated that forest fire prediction would lessen the effects of forest firing in the future. There are numerous fire detection algorithms available with various fire detection approaches. The region affected by the fire is forecasted using satellite imagery in the current work processes.

The proposed system procedures employed weather variables like temperature, rain, wind, and humidity to forecast the likelihood of a forest fire. With the help of the RandomizedSearchCV algorithm, we performed Random Forest Regression and Hyper - parameter Tuning using a variety of sub-samples of the dataset. The predictive accuracy and over-fitting were improved by averaging the results from several decision trees that were fitted on the dataset. The events of forest fires can be modeled using the evaluation of the models and all the chosen meteorological parameters. In this essay, we compare and contrast several models for predicting forest fires, including Decision Tree, Random Forest and SVM(Support Vector Machine), and Artificial Neural Networks (ANN) algorithms.

The study of RandomizedSearchCV coefficient calculation with Hyperparameter tuning yields the best outcomes of Root mean square error (RMSR) is equal to 0.07, Mean absolute error (MAE) is 0.03 and Mean square error (MSE) is 0.004.

V. PROPOSED FRAMEWORK

In order to predict the forest fire from the above given factors between, we proposed the below algorithms for the prediction

1. ***SVM:***

The foundation of an SVM classifier (Support Vector Machine) method is the creation of a hyperplane that divides data into several groups. An SVM classifier's primary goal is to construct hyperplanes using a variety of methods. This hyperplane building procedure varies and is the main task of an SVM classifier. The main focus while drawing the hyperplane is on maximizing the distance from hyperplane to the nearest data point of either class. These nearest data points are known as Support Vectors. In machine learning, **support vector machines** are used for classification and regression analysis. However, they are mostly used in classification problems. SVM needs labeled training data, much as other supervised learning machines. Materials are designated in groups for categorization. SVM training materials are divided into distinct groups and categorized individually at various locations in space.

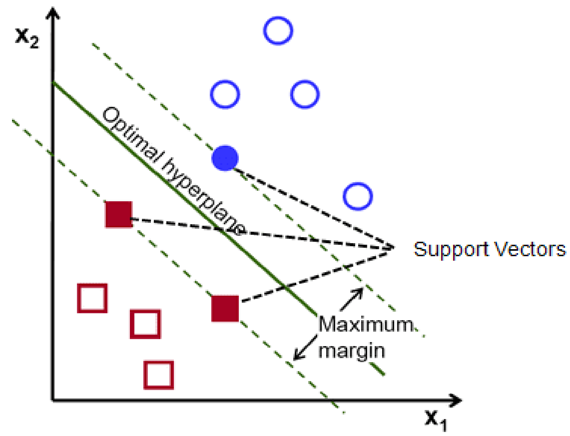

Figure 2

The Hyperplane dividing the data points and a tangent from the data point nearest to the hyperplane forms the support vectors

SVMs may do unsupervised learning after digesting a large number of training instances. The algorithms will work to increase the perimeter around the hyperplane and maintain equality between the two sides in order to obtain the optimal data separation.

2. ***SMOTE:***

When observed frequencies of a categorical data vary significantly among its conceivable values, the data is said to be imbalanced. In general, there are many of observations of one kind and few of another.

SMOTE is a remedy for data that is unbalanced. Undersampling is the simplest way to balance out class disparity. When you under sample, you exclude some of the data points from the class that are present too frequently.

Undersampling has the drawback of causing you to waste a lot of important data. Oversampling is another quick fix for unbalanced data. Under sampling's opposite is oversampling. Duplicating the information that is least prevalent in your data collection called oversampling. After that, you include those duplicates in your data collection. The drawback of oversampling is that it results in a lot of duplicate data points. SMOTE is indeed an algorithm that adds artificial data points to the actual data points to accomplish data augmentation. SMOTE can be thought of as a specific data augmentation technique or as an enhanced version of oversampling. With SMOTE, you avoid producing duplicate data points and instead produce synthetic data points that really are

marginally distinct from the original data points. SMOTE is a superior oversampling substitute.
Following is how the SMOTE algorithm operates:

- A random sample is chosen out from minority class.
- Find the k nearest neighbors for such observations in this sample.
- The vector between both the existing data point and one of the neighbors will then be determined using that neighbor.
- The vector is multiplied by a chance number ranging from 0 and 1.
- You combine this with the existing data point to get the synthetic data point.

3. *LINEAR REGRESSION:*

An algorithm for machine learning based on supervised learning is linear regression. It carries out a regression job. Based on independent variables, regression models a goal prediction value. Finding the connection between variables and predicting is its main purpose. The type of link that different regression models take into account between the dependent and independent variables, as well as the quantity of independent variables utilized, are what make them different. In order to make predictions, linear regression models offer a straightforward mathematical formula. A well-known statistical method, linear regression is simple to use in software and computers. It is utilized by businesses to consistently and predictably transform the raw data to business intelligence & useful insights. Several behavioral, environmental, and social sciences, as well as biology, employ linear regression to do early data analysis and forecast future trends. The use of linear regression in several data science techniques, including machine learning and artificial intelligence, helps to resolve complicated issues. A basic linear regression method's fundamental goal is to draw a trend line between the two data variables x and y. The horizontal axis is used to plot x, the independent variable. Explanatory or predictive variables are other names for independent variables. On the vertical axis, the dependent variable, y, is shown. Y values can alternatively be referred to as anticipated variables or response variables.

Here are the few types of Linear Regression:

- SIMPLE LINEAR REGRESSION:
  To represent the connection between two variables, such as these, simple linear regression can be used:

- MULTIPLE LINEAR REGRESSION:
  Modeling numerous factors and their effects on an outcome using multiple linear regression

- LOGISTIC REGRESSION:
  To calculate the likelihood that an event will occur, data scientists utilize logistic regression. The forecast is a number ranging from zero and 1, wherein 0 denotes an event as unlikely to occur and 1 denotes the highest probability that it will. Logarithmic functions are used in logistic equations to determine the regression line.

4. *RANDOM FOREST REGRESSION:*

Like its name suggests, a random forest is an ensemble of several different decision trees. Each each tree inside the random forest spew out a class prediction, and the class with the highest votes becomes our model's forecast.

The crucial factor is the poor correlation between models. Uncorrelated models can provide ensemble forecasts which are more precise than any of the individual predictions, much like assets with low correlations (such as stocks and bonds) combine to make your portfolio more than the finished product. As long as they don't consistently all error in the same direction, the trees shield one another from their individual mistakes, which accounts for this lovely effect.

The random forest algorithm's steps are as follows:
Step 1: From a data collection with k records, n random records are selected at random and used in the Random Forest algorithm.

Step 2: For each sample, a unique decision tree is built.

Step 3: An output will be produced by each decision tree.

Step 4: For classification and regression, the final result is evaluated using a majority vote or an average.

VI. DATA DESCRIPTION

Forest fires may harm the environment and put people's lives and property in jeopardy. It's critical to managing them to know when they happen and what triggers them. The information we'll use in this project is related to a scientific study that used modeling techniques to forecast the likelihood of forest fires in Portugal. For this project, modeling and visualization will be equally important. In order to fully comprehend the data and uncover any potential links, exploratory analysis will also be performed on it.

Following is the description of the variables that are present in the dataset and the range of values taken

- **X**: X-axis is the spatial coordinate area inside the park map: 1 to 9

- **Y**: Y-axis is the spatial coordinate area inside the park map: 2 to 9
- **month**: Months from the year: from 'jan' to 'dec'
- **day**: Days of the week: from 'mon' to 'sun'
- **FFMC**: Fine Fuel Moisture Code index: 18.7 to 96.20
- **DMC**: Duff Moisture Code index system: 1.1 to 291.3
- **DC**: Drought Code index : 7.9 to 860.6
- **ISI**: Initial Spread Index : 0.0 to 56.10
- **temp**: Temperature (Degree Celcius) : 2.2 to 33.30
- **RH**: Relative humidity(In percentage): 15.0 to 100
- **wind**: Wind speed in km/h: 0.40 to 9.40
- **rain**: Outside rain in mm/m2 : 0.0 to 6.4
- **area**: The burned area of the forest (in hector area): 0.00 to 1090.84
- **Size Category**: Considered as small when area is 0 and large when the area is greater than 0

The dataset was split into three categories, with training datasets and testing datasets being separated by 70% and 30%, respectively for SVM, Linear Regression and Random Forest Classifier, to prevent overfitting, SMOTE technique is used and then the accuracy is calculated for SVM. Data pre-processing methods include removing unwanted or noisy data, treating missing values, and detecting outliers.

VII. RESULTS & COMPARISION/ANALYSIS

Results:

After performing all the algorithms to the data set we got different accuracy percentages for different algorithms and performed visualization for data when the forest fire is happened and representing the area of fire that was happened at that time.

SVM: The below image is the result for Accuracy percentages which is obtained by performing SVM to the training and testing data set.

```
Train Result
Accuracy Score: 55.68%
Test Result
Accuracy Score: 57.69%
```

Figure 3

SVM_SMOTE: The below image is the result for Accuracy percentages which is obtained by performing SMOTE techniques to the SVM which is used to overcome the fitting problem

```
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
SVC with ACTUAL
============================================================
Confusion Matrix
============================================================
[[72  2]
 [ 0 82]]

Classification Report
============================================================
              precision    recall  f1-score   support

           0       1.00      0.97      0.99        74
           1       0.98      1.00      0.99        82

    accuracy                           0.99       156
   macro avg       0.99      0.99      0.99       156
weighted avg       0.99      0.99      0.99       156


AUC-ROC
============================================================
1.0
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
SVC with SMOTE
============================================================
Confusion Matrix
============================================================
[[73  1]
 [ 0 82]]

Classification Report
============================================================
              precision    recall  f1-score   support

           0       1.00      0.99      0.99        74
           1       0.99      1.00      0.99        82

    accuracy                           0.99       156
   macro avg       0.99      0.99      0.99       156
weighted avg       0.99      0.99      0.99       156


AUC-ROC
============================================================
1.0
```

Figure 4

Linear regression: The below image is the result for accuracy which is obtained by performing linear regression to the dataset.

```
R2 score for Linear Regression 0.6361427176005193
```

Figure 5

```
Adjusted R2 score for Linear regression: 0.548816969824644
```

Figure 6

We have two different accuracies in Linear regression which is R2 and Adjusted R2.Accuracy of model increases whenever we increase the features. But the adjusted R2 will remain unaffected inspite of features.

Random Forest Regression: The below image is the result for accuracy which is obtained by performing Random forest regression algorithm to the data set.

```
Random Forest Regressor
R2 Score value: 1.0000
MAE value: 0.0001
```

Figure 7

Visualization:

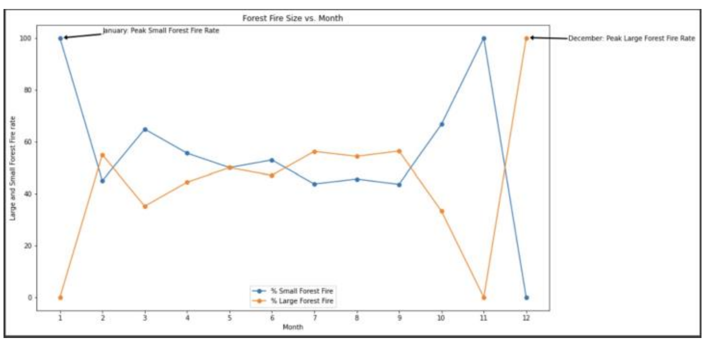

Figure 8

The above line graph represents the relation between Forest Fire size vs Month in above figure x-axis represents the numbers of each month and y-axis represents Forest fire rate, the blue line represents the small forest fire and orange line represents the large forest fire.

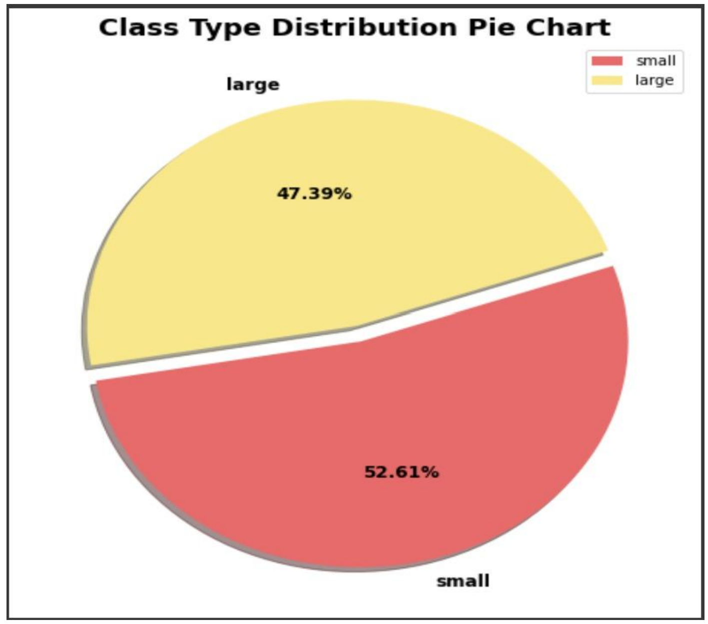

Figure 8

The above pie chart represents the area percentage where the fire is small and large in percentages.

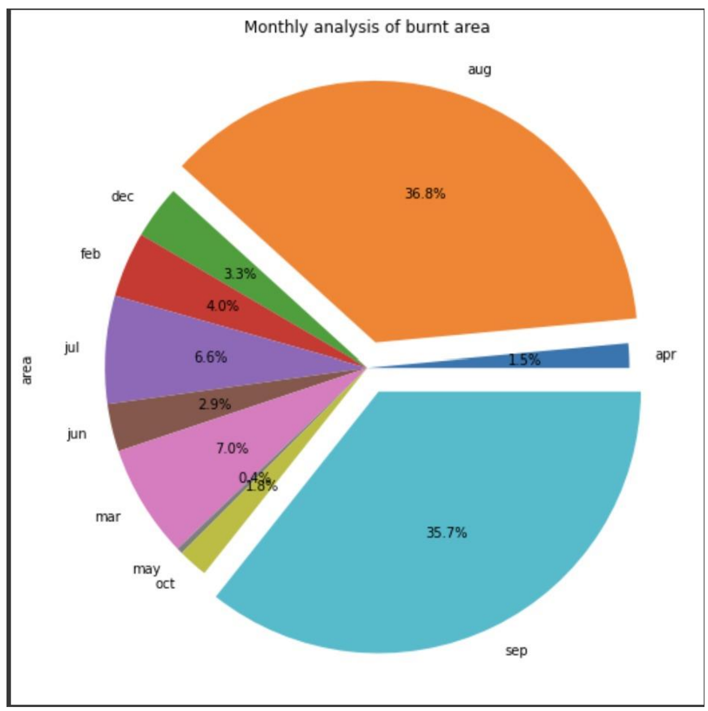

Figure 9

The above pie chart represents the Monthly analysis of burnt area where it represents the percentage of area has burnt at that particular time of month which is easier to identify when the most fire accidents were occurred and it will be easier to be cautious at that time.

Analysis:

We used SVMs (Support Vector Machines) to determine the likelihood of a fire occurring using the provided data. The results are shown above, and we can see that they are biased in favor of the majority as a result of the dataset's imbalance. We have employed some sort of oversampling techniques to balance the dataset because the model accurately classifies all of the majority classes but misclassifies the minority classes using SMOTE techniques. We also added the Linear regression, Random forest Regression algorithms by comparing all the accuracies we can say that SVM with SMOTE techniques is more accurate compared to other. We also visualized the data by representing the relation between the forest fire area and months.

REFERENCES

[1] P. Rakshit et al., "Prediction of Forest Fire Using Machine Learning Algorithms: The Search for the Better Algorithm," 2021 6th International Conference on Innovative Technology in Intelligent System and Industrial Applications (CITISIA), 2021, pp. 1-6, doi: 10.1109/CITISIA53721.2021.9719887..

[2] T. Preeti, S. Kanakaraddi, A. Beelagi, S. Malagi and A. Sudi, "Forest Fire Prediction Using Machine Learning Techniques," 2021 International Conference on Intelligent Technologies (CONIT), 2021, pp. 1-6, doi: 10.1109/CONIT51480.2021.9498448.

[3] S. Natekar, S. Patil, A. Nair and S. Roychowdhury, "Forest Fire Prediction using LSTM," 2021 2nd International Conference for Emerging Technology (INCET), 2021, pp. 1-5, doi: 10.1109/INCET51464.2021.9456113.

[4] G. E. Sakr, I. H. Elhajj, G. Mitri and U. C. Wejinya, "Artificial intelligence for forest fire prediction," 2010 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, 2010, pp. 1311-1316, doi: 10.1109/AIM.2010.5695809.

[5] G. Demin, L. Haifeng, J. Anna and W. Guoxin, "A forest fire prediction system based on rechargeable wireless sensor networks," 2014 4th IEEE International Conference on Network Infrastructure and Digital Content, 2014, pp. 405-408, doi: 10.1109/ICNIDC.2014.7000334.

[6] T. Bhatt and A. Kaur, "Automated Forest Fire Prediction Systems: A Comprehensive Review," 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2021, pp. 1-5, doi: 10.1109/ICRITO51393.2021.9596528.

[7] N. Omar, A. Al-zebari and A. Sengur, "Deep Learning Approach to Predict Forest Fires Using Meteorological Measurements," 2021 2nd International Informatics and Software Engineering Conference (IISEC), 2021, pp. 1-4, doi: 10.1109/IISEC54230.2021.9672446.

[8] W. Liu and P. Li, "Forest fire prediction model based on Improved BP neural network," ICETIS 2022; 7th International Conference on Electronic Technology and Information Science, 2022, pp. 1-5.

[9] P. Thirumal and L. Shylu Dafni Agnus, "Forest Fire Detection and Prediction – Survey," 2022 International Conference on Inventive Computation Technologies (ICICT), 2022, pp. 1295-1302, doi: 10.1109/ICICT54344.2022.9850518.

[10] M. Anshori, F. Mar'i, M. W. Alauddin and W. F. Mahmudy, "Prediction of Forest Fire using Neural Network based on Extreme Learning Machines (ELM)," 2019 International Conference on Sustainable Information Engineering and Technology (SIET), 2019, pp. 301-305, doi: 10.1109/SIET48054.2019.8986106.

[11] J. Zhang, W. Li, Z. Yin, S. Liu and X. Guo, "Forest fire detection system based on wireless sensor network," 2009 4th IEEE Conference on Industrial Electronics and Applications, 2009, pp. 520-523, doi: 10.1109/ICIEA.2009.5138260.

[12] N. Hamadeh, A. Hilal, B. Daya and P. Chauvet, "Studying the factors affecting the risk of forest fire occurrence and applying neural networks for prediction," 2015 SAI Intelligent Systems Conference (IntelliSys), 2015, pp. 522-526, doi: 10.1109/IntelliSys.2015.7361189.

[13] Z. Qu, H. Hu and L. Yu, "Study of a Prediction Model for Forest Fire-Initial Burnt Area on Meteorological Factors," 2009 International Workshop on Intelligent Systems and Applications, 2009, pp. 1-4, doi: 10.1109/IWISA.2009.5073121.

[14] X. Gao, X. Fei and H. Xie, "Forest fire risk zone evaluation based on high spatial resolution RS image in Liangyungang Huaguo Mountain Scenic Spot,"

Proceedings 2011 IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services, 2011, pp. 593-596, doi: 10.1109/ICSDM.2011.5969116.

[15] K. Satoh, Song Weiguo and K. T. Yang, "A study of forest fire danger prediction system in Japan," Proceedings. 15th International Workshop on Database and Expert Systems Applications, 2004., 2004, pp. 598-602, doi: 10.1109/DEXA.2004.1333540.

[16] R. S. Aakash, M. Nishanth, R. Rajageethan, R. Rao and R. Ezhilarasie, "Data Mining Approach to Predict Forest Fire Using Fog Computing," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), 2018, pp. 1582-1587, doi: 10.1109/ICCONS.2018.8663160.

[17] R. Ghali, M. A. Akhloufi, M. Jmal, W. S. Mseddi and R. Attia, "Forest Fires Segmentation using Deep Convolutional Neural Networks," 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2021, pp. 2109-2114, doi: 10.1109/SMC52423.2021.9658905.

[18] B. K. Singh, N. Kumar and P. Tiwari, "Extreme Learning Machine Approach for Prediction of Forest Fires using Topographical and Metrological Data of Vietnam," 2019 Women Institute of Technology Conference on Electrical and Computer Engineering (WITCON ECE), 2019, pp. 104-112, doi: 10.1109/WITCONECE48374.2019.9092926.

[19] S. Natekar, S. Patil, A. Nair and S. Roychowdhury, "Forest Fire Prediction using LSTM," 2021 2nd International Conference for Emerging Technology (INCET), 2021, pp. 1-5, doi: 10.1109/INCET51464.2021.9456113.

[20] B. Li, J. Zhong, G. Shi and J. Fang, "Forest Fire Spread Prediction Method based on BP Neural Network," 2022 9th International Conference on Dependable Systems and Their Applications (DSA), 2022, pp. 954-959, doi: 10.1109/DSA56465.2022.00134.