RESEARCH ARTICLE

INTERNATIONAL RESEARCH JOURNAL OF
MULTIDISCIPLINARY TECHNOVATION

Asian Research Association

DOI: 10.54392/irjmt2413

# Machine Learning based Forest Fire Prediction: A Comparative Approach

**Rohini Patil [a,*], Janhvi Pawar [a], Kamal Shah [b], Disha Shetty [a], Aparna Ajith [a], Sakshi Jadhav [a]**

[a] Department of Computer Engineering, Terna Engineering College, Navi Mumbai-400706, India

[b] Department of Information Technology, Thakur College of Engineering and Technology, Mumbai-400101, India

*Corresponding Author Email: rohiniapatil01@gmail.com

DOI: https://doi.org/10.54392/irjmt2413

Received: 15-09-2023; Revised: 19-11-2023; Accepted: 24-12-2023; Published: 17-01-2024

Check for updates

**Abstract:** Wildfires are among the world's most pressing issues, and they are getting more prevalent as global warming and other environmental conditions deteriorate. These wildfires might be caused by humans or by natural causes. Wildfires are one of the factors contributing to the extinction of rare flora and wildlife that serve to maintain our planet's ecological balance. In this paper, a comparative analysis of various machine learning classifier models for predicting forest fires was undertaken using two separate datasets. The suggested system's processing is dependent on a few characteristics such as temperature, humidity, oxygen, and wind. Several machine learning classification techniques, including logistic regression, support vector classifier, decision tree, k neighbors and random forest, were used in this study. For further optimization of the model, K-fold cross validation method and hyperparameter tuning were implemented. The system reveals Support Vector Machine as the best strategy for the forest fire dataset, with 96.88% accuracy. Random Forest method was found to be the best for the Cortez and Morais dataset, with 90.24% accuracy.

**Keywords:** Machine Learning, Classification Techniques, Prediction, Wildfires.

## 1. Introduction

Globally, forests constitute about 31% of the total land area. However forest fire affects about 36% of the forest land [1], representing as one of the most common natural catastrophes in past few years. Recently in third week of February 2023, a total of 1156 forest fires were reported in India [2].

Due to these forest fires, many hectares of forest are being destroyed. National economies are closely linked and strongly affected by wildfires. Ecological and environmental damage affects the humans and livestock directly as well as indirectly [3].

Wildfires can be caused by a number of factors including changing weather conditions, rising temperatures, campfires, and many more. Many causes of wildfires are due to an increase in the planet's average temperature and human ignorance. In areas with high wildfire risk, management services can improve fire prevention by setting up watchtowers and using additional tracking equipment [4].

The main obstacle to regional forecasting is its uncertainty. This feature is undesirable but unavoidable for fire fighters' response. This uncertainty seems to remain without its complete elimination. However, efforts can be done to minimize its severity. An automated prediction system and planning the necessary measures beforehand can be an effective solution to this problem. Thus, predicting the forest fires has become a crucial issue nowadays [5]

This research can serve as a guide for choosing a highly accurate fire prediction model and offer a scientific foundation for forest fire prevention [6-13]. In the recent years, many researchers have used the concept of machine learning (ML) and deep learning based model for predicting the forest fire [14-20]. The current model plans to predict the forest fire well in advance in a less complex way. Furthermore, this model can help us to be panic-free from the last-minute chaos. Several machine learning approaches like Logistic Regression (LR), Random Forest (RF), Decision Tree (DT), Support Vector Classifier (SVC), K Nearest Neighbors Classifier (KNN) were applied on two separate datasets. Hyperparameter tuning and K-fold cross validation methods were used for optimizing the model. and avoid overfitting and improve the accuracy.

The rest of the paper is organized as follows. Materials and methods are described in section 2. Result and discussions are described in section 3. Overall conclusion is provided in section 4.

## 2. Materials and Methods

### 2.1. Description of the Dataset Used

The system consisted of two datasets with various attributes. Data from these datasets were further used for training and testing the machine learning models. Dataset 1 [21] consisted of attributes like oxygen, temperature (°F), humidity and fire occurrence. The Dataset 1 consisted of 201 instances and 4 features. Dataset 2 [22] was the Cortez and Morais dataset which consisted of attributes such as X, Y, month, FFMC, DMC, DC, ISI, Temp, RH, wind, rain, area and fire occurrence. The Dataset 2 consisted of 517 instances and 13 features. Both the datasets were collected from Kaggle.

### 2.2 Proposed Methodology Was Divided into Two Phases

*Phase 1: K-Fold Cross Validation.* In the initial phase (Figure. 1), the dataset 1 and 2 were filtered using data pre-processing. After the cleaning of the datasets for the training and testing phases, data were split into 70% to 30% respectively. The datasets were trained with ML classifiers such as LR, DT, RF, KNN and SVC to increase the model's accuracy. K-fold cross-validation technique was used to get more precise results. In K-fold cross-validation procedure, K=10 was chosen.

*Phase 2: Optimization.* In order to achieve the optimal values for the proposed model, hyperparameter tuning was done on the datasets. In this technique, the correct combination of hyperparameters was identified to enhance the performance of the model. For the suggested design, the simple grid search method was chosen. Here in the grid search, a grid of the hyperparameters was formed, and then iterated in combination until left with the best model along with their hyperparameter values.

### 2.3 Performance Measure

**Accuracy:** The accuracy of the dataset is given by:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

**Precision:** The precision of the dataset is given by:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

**Recall:** The recall of the dataset is given by:

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

**F1 Score.** The F1 Score of any dataset is given by:

$$F1 - score = \frac{2*(TP)}{2*(TP) + FP + FN} \quad (4)$$

## 3. Result and Discussions

### 3.1 Training Accuracy, Testing Accuracy and K-Fold Accuracy

The Proposed approach worked on two different datasets.

**Dataset 1: Forest Fire Dataset** Figure. 2 represents the comparison between the training, testing and k-fold accuracies for dataset 1. Highest training accuracy was showed by RF and DT classifiers. RF model showed highest testing and k-fold accuracy of 95.12% and 96% respectively.

**Dataset 2: Cortez-Morais Dataset.** Figure. 3 shows the comparison between the training, testing and k-fold accuracies for dataset 2. Among all classifiers, RF provided maximum training, testing and k-fold accuracy of 99.76%, 89.32% and 89.07% respectively.

## Flowchart

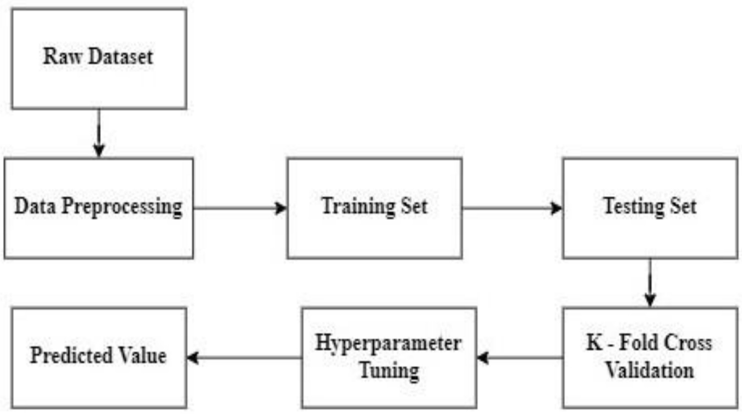


**Figure1.** Proposed System

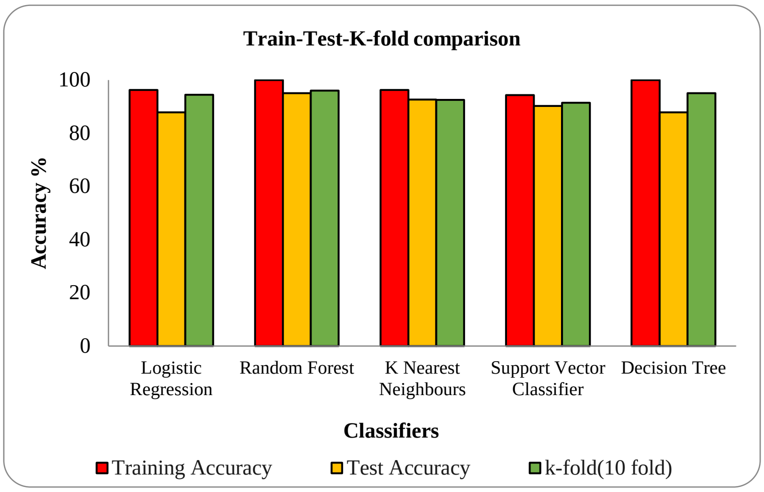


**Figure 2.** Training, Testing and K-fold comparison graph for Dataset 1

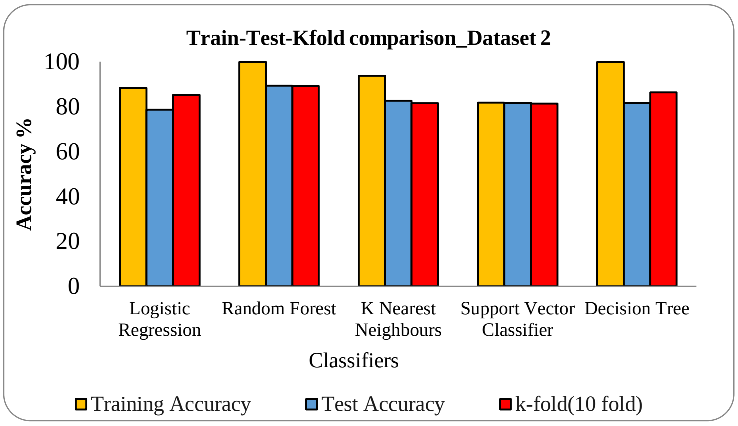


**Figure 3.** Training, Testing and K-fold comparison graph for Dataset 2

## 3.2 K-Fold Accuracy and Hyperparameter Tuning Accuracy

Figure. 4 represents the comparison between the k-fold and hyperparameter tuning accuracies for dataset 1. RF provided highest k-fold accuracy of 96% while lowest accuracy was seen with SVC of 91.50%. After applying hyperparameter tuning with best estimator as, C=50, gamma=scale, kernel= rbf, SVC produced an accuracy of 96.88%, suggesting it to be the best model whereas DT showed lowest accuracy of 94.38%. Accuracy details of dataset 1 are mentioned in Table 1

Below fig. 5 shows accuracy comparison on dataset-2. The detailed analysis is mentioned in Table 2. Observation showed that RF gives highest k-fold accuracy of 89.07% while lowest accuracy observed with SVC of 81.27%. In hyperparameter tuning accuracy, RF showed 90.48% using the best estimator as, max_features-sqrt, 'n_estimators'=100 whereas KNN showed 87.56% making it the poor model. There was an improvement in accuracy after applying the hyperparameter tuning.

## 3.3 Performance Measure

Receiver Operating Characteristic (ROC-curve) for dataset1 and dataset 2 is shown in Table 3. All classifiers showed good results on both datasets. Highest ROC-AUC score of 0.988 for dataset 1 was shown by LR and RF of 0.9382 for dataset 2. Corresponding ROC curve analysis is shown in fig 6 & 7 respectively.

Other performance measures i.e., precision, recall and f1-score values are mentioned in Table 4 and Table 5 for the dataset 1 and dataset 2 respectively.

RF model worked better on both datasets. For dataset1, RF provided greatest precision, recall and F1 score value of 96%, 95% and 95% respectively while for dataset2 precision, recall and F1 score were 90%, 89% and 89% respectively.

## 3.4 Comparative Analysis

Table 6 showed comparative analysis with other researcher's best models. The proposed work SVC model was validated on the same dataset. The result showed that proposed work SVC model provided better accuracy of 96.88% in comparison with others.

Figure. 8 represents the graphical representations of models.

Similarly Table 7 showed comparative analysis validated on Cortez-Morais Dataset (dataset 2). The proposed approach applied on RF model showed best result in comparison with Sharma and co-workers [25].

Figure. 9 represents the visualization of models used for validation purpose on dataset 2.

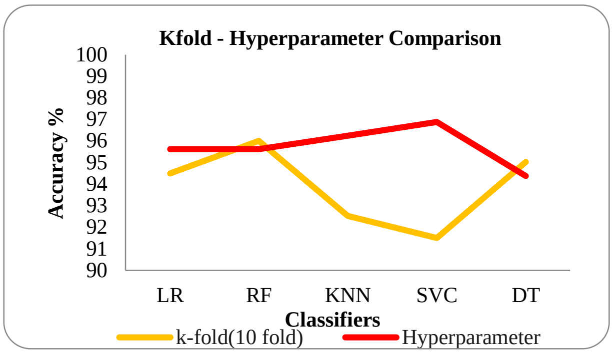


**Figure 4.** K-fold and Hyperparameter Accuracy-Dataset 1

**Table 1.** Accuracy Analysis

| **Dataset 1** | **Accuracy** | | |
|---|---|---|---|
| **Classifiers** | **Testing** | **k-fold** | **Hyperparameter tuning** |
| Logistic Regression | 87.80 | 94.50 | 95.63 |
| Random Forest | 95.12 | 96.00 | 95.63 |
| K Nearest Neighbours | 92.68 | 92.52 | 96.25 |
| Support Vector Classifier | 90.24 | 91.50 | 96.88 |
| Decision Tree | 87.80 | 95.02 | 94.38 |

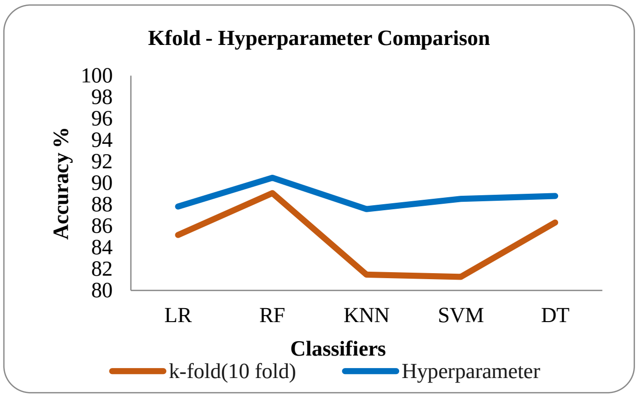


**Figure 5.** K-fold and hyper parameter Accuracy-Dataset 2

**Table 2.** Accuracy Comparison

| **Dataset 2** | **Accuracy** | | |
|---|---|---|---|
| **Classifiers** | **Testing** | **k-fold** | **Hyperparameter tuning** |
| Logistic Regression | 78.64 | 85.17 | 87.80 |
| Random Forest | 89.32 | 89.07 | 90.48 |
| K Nearest Neighbours | 82.52 | 81.49 | 87.56 |
| Support Vector Classifier | 81.55 | 81.27 | 88.54 |
| Decision Tree | 81.55 | 86.34 | 88.78 |

**Table 3.** ROC-AUC Score for Dataset 1 and Dataset 2

| Classifiers | ROC-AUC Score | |
|---|---|---|
| | Dataset1 | Dataset2 |
| RF | 0.9808 | 0.9382 |
| LR | 0.9880 | 0.8897 |
| KNN | 0.9282 | 0.8986 |
| DT | 0.9138 | 0.8930 |
| SVC | 0.9856 | 0.8955 |

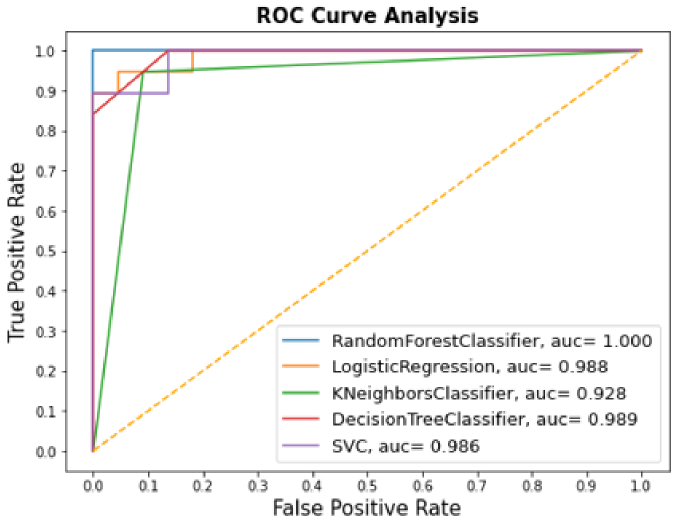


**Figure 6.** ROC Curve Analysis for Dataset 1

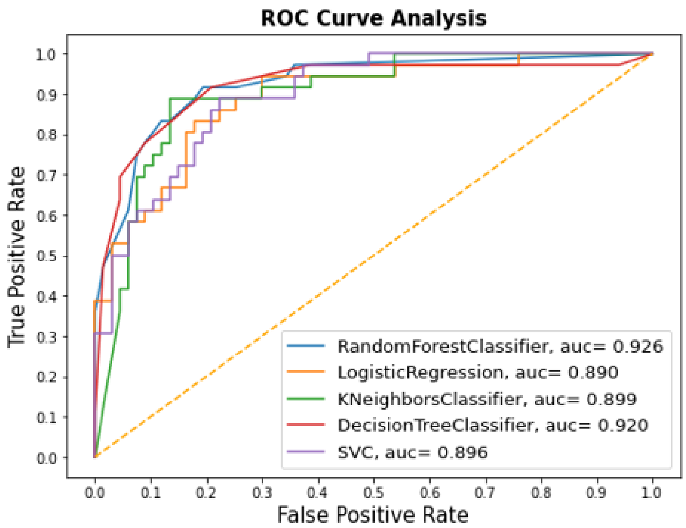


**Figure 7.** ROC Curve Analysis for Dataset 2

**Table 4.** Performance Measure for Dataset 1

| Dataset 1 | | | |
|---|---|---|---|
| Classifier | Precision | Recall | F1 score |
| LR | 89 | 88 | 88 |
| RF | 96 | 95 | 95 |
| KNN | 93 | 93 | 93 |
| SVC | 91 | 90 | 90 |
| DT | 85 | 85 | 85 |

**Table 5.** Performance Measure for Dataset 2

| Dataset 2 | | | |
|---|---|---|---|
| Classifier | Precision | Recall | F1 score |
| LR | 80 | 81 | 80 |
| RF | 90 | 89 | 89 |
| KNN | 84 | 84 | 84 |
| SVC | 79 | 80 | 79 |
| DT | 85 | 84 | 85 |

**Table 6.** Comparative Analysis for Dataset 1

| Sr. No. | Ref. | Best Model | Accuracy (%) |
|---|---|---|---|
| 1 | [8] | Extreme Machine Learning | 85.42 |
| 2 | [23] | LR | 86.9 |
| 3 | [24] | Random Forest Regression | 91.2 |
| 4 | Proposed Work | SVC | **96.88** |

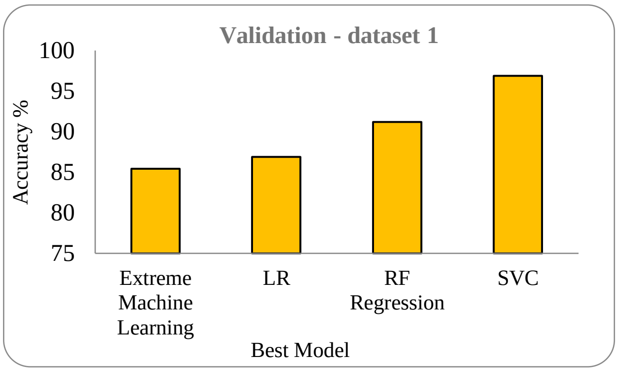


**Figure 8.** Validation with other researchers – Dataset 1

**Table 7.** Comparative Analysis for Dataset2

| Sr. No | Ref | Best Model | Accuracy (%) |
|---|---|---|---|
| 1 | [25] | Boosted Decision Tree | 72 |
| 2 | Proposed Work | RF | **90.48** |

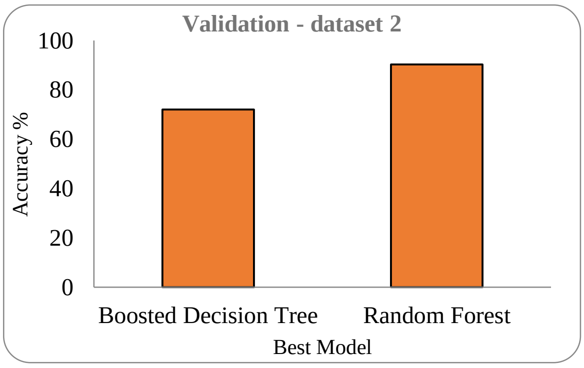


**Figure 9.** Validation with other researchers – Dataset 2

## 4. Conclusion

Forest fires or wildfires are the major threats created by mankind or natural disasters. This leads to measure losses to our nature as well as the humans. Predicting these forest fires may help the environment and humans to protect themselves from the uncontrollable fires turning each and every bit into ashes. In this paper, comparative analysis and prediction tasks helped to get better results. Comparative analysis helped to get better idea of the subject and provided better results for the models. The prediction task using the ML classification techniques like LR, DT, KNN, SVC and RF helped to outperform other models implemented before in other research work. Overall, the current model plans to predict the forest fire well in advance in a less complex way. Furthermore, this model may help us to be panic-free from the last-minute chaos. Further work on increasing the accuracy and speed of our model is required. Live predictions or on-site predictions using satellite images can also be performed. Ensemble approach may be used for prediction of forest fire.

## References

[1] Global Forest Resources Assessment 2020, Food and Agriculture Organization (2020) Available at https://www.fao.org/forest-resources-assessment/2020/en/

[2] S. Yashwant, (2023) Forest fires threaten India's climate and biodiversity. Available at https://www.deccanherald.com/opinion/forest-fires-threaten-india-s-climate-and-biodiversity-1194415.html

[3] G. Demin, L. Haifeng, J. Anna, W. Guoxin, (2014) A forest fire prediction system based on rechargeable wireless sensor networks. 4th International Conference on Network Infrastructure and Digital Content, IEEE, Beijing, China. https://doi.org/10.1109/ICNIDC.2014.7000334

[4] S. Natekar, S. Patil, A. Nair, S. Roychowdhury, (2021) Forest Fire Prediction using LSTM. 2nd International Conference for Emerging Technology (INCET), IEEE, Belagavi, India. https://doi.org/10.1109/INCET51464.2021.9456113

[5] T. Bhatt, A. Kaur, (2021) Automated Forest Fire Prediction Systems: A Comprehensive Review. 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), IEEE, Noida, India. https://doi.org/10.1109/ICRITO51393.2021.9596528

[6] G.E. Sakr, I.H. Elhajj, G. Mitri, U.C. Wejinya, (2010) Artificial intelligence for forest fire prediction. IEEE/ASME International Conference on Advanced Intelligent Mecha tronics, Montreal, QC, Canada. https://doi.org/10.1109/AIM.2010.5695809

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

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

[9] M. Anshori, F. Mar'i, M.W. Alauddin, W.F. Mahmudy, (2019) Prediction of Forest Fire using Neural Network based on Extreme Learning Machines (ELM). International Conference on Sustainable Information Engineering and Technology (SIET), Lombok, Indonesia. https://doi.org/10.1109/SIET48054.2019.8986106

[10] G.F. Shidik, K. Mustofa, (2014) Predicting Size of Forest Fire Using Hybrid Model. Information and Communication Technology. ICT-EurAsia. Lecture Notes in Computer Science, 8407, (2014). https://doi.org/10.1007/978-3-642-55032-4_31

[11] A. Gaikwad, N. Bhuta, T. Jadhav, P. Jangale, S. Shinde, (2022) A Review on Forest Fire Prediction Techniques. 6th International Conference on Computing, Communication, Control and Automation (ICCUBEA), IEEE, Pune, India. https://doi.org/10.1109/ICCUBEA54992.2022.10010889

[12] M. Singh, C. Sharma, T. Agarwal, M.S. Pal, (2022) Forest Fire Prediction for NASA Satellite Dataset Using Machine Learning. 10th International Conference on Reliability, Info com Technologies and Optimization (Trends and Future Directions) (ICRITO), IEEE, Noida, India. https://doi.org/10.1109/ICRITO56286.2022.9964975

[13] F. Abid, A Survey of Machine Learning Algorithms Based Forest Fires Prediction and Detection Systems. Fire Technology, 57,

(2021). https://doi.org/10.1007/s10694-020-01056-z

[14] K.V. Murali Mohan, A.R. Satish, K. Mallikharjuna Rao, R.K. Yarava, G.C. Babu, (2021) Leveraging Machine Learning to Predict Wild Fires. 2nd International Conference on Smart Electronics and Communication (ICOSEC), IEEE, India. https://doi.org/10.1109/ICOSEC51865.2021.9591952

[15] D. Rosadi, W. Andriyani, D. Arisanty, D. Agustina, (2020) Prediction of Forest Fire Occur rence in Peatlands using Machine Learning Approaches, 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), IEEE, Yogyakarta, Indonesia. https://doi.org/10.1109/ISRITI51436.2020.9315359

[16] P. Rakshit, S. Sarkar, S. Khan, P. Saha, S. Bhattacharyya, N. Dey, S.M.N. Islam, S. Pal, (2021) Prediction of Forest Fire Using Machine Learning Algorithms: The Search for the Better Algorithm, 6th International Conference on Innovative Technology in Intelligent System and Industrial Applications (CITISIA), IEEE, Sydney, Australia. https://doi.org/10.1109/CITISIA53721.2021.9719887

[17] Y.O. Sayad, H. Mousannif, H. Al Moatassime, Predictive modeling of wildfires: A new dataset and machine learning approach, Fire Safety Journal, 104, (2019) 130–146. https://doi.org/10.1016/j.firesaf.2019.01.006

[18] Z. Liu, K. Zhang, C. Wang, S. Huang, Research on the identification method for the forest fire based on deep learning. Optik International Journal for Light and Electron Optics, 223 (2020) 165491. https://doi.org/10.1016/j.ijleo.2020.165491

[19] S. Saha, B. Bera, P. Shit, S. Bhattacharjee, N. Sengupta, Prediction of forest fire susceptibility applying machine and deep learning algorithms for conservation priorities of forest resources. Remote Sensing Applications: Society and Environment, 29 (2023) 100917. https://doi.org/10.1016/j.rsase.2022.100917

[20] M. Amiri, H.R. Pourghasemi, Predicting areas affected by forest fire based on a machine learning algorithm. Computers in Earth and Environmental Sciences, (2022) 351-362. https://doi.org/10.1016/B978-0-323-89861-4.00004-X

[21] P. Panditi, (2021) Forest fire prediction, Kaggle. Available at: https://www.kaggle.com/datasets/pradeeppanditi/forest-fire-prediction

[22] P. Cortez, A. Morais, (2008). Forest Fires. UCI Machine Learning Repository. https://doi.org/10.24432/C5D88D

[23] L. Si, L. Shu, M. Wang, F. Zhao, F. Chen, W. Li, W. Li, Study on forest fire danger prediction in plateau mountainous forest area. Natural Hazards Research, 2 (2022) 25-32. https://doi.org/10.1016/j.nhres.2022.01.002

[24] T. Preeti, S. Kanakaraddi, A. Beelagi, S. Malagi A. Sudi, (2021) Forest Fire Prediction Using Machine Learning Techniques. International Conference on Intelligent Technologies (CONIT), IEEE. https://doi.org/10.1109/CONIT51480.2021.9498448

[25] R. Sharma, S. Rani, I. Memon, A smart approach for fire prediction under uncertain conditions using machine learning. Multimedia Tools and Applications, 79 (2020) 28155–28168. https://doi.org/10.1007/s11042-020-09347-x

## Authors Contribution Statement

**Rohini Patil:** Conceptualization, Methodology,Writing Draft, Implementation, Review and Editing; **Janhvi Pawar:** Conceptualization and Implementation, Methodology, Writing Draft and Editing; **Kamal Shah:** Formal analysis, Review and Editing; **Disha Shetty:** Writing Draft, Methodology and performed the simulation work and generated the results, Review and Editing; **Aparna Ajith**: Methodology, Implementation, paper formatting, Review and Editing; **Sakshi Jadhav**: Writing Draft and reviewing the literature. All the authors read and approved the final version of the manuscript.

## Has this article screened for similarity?

Yes

## Conflict of Interest

The Authors have no conflicts of interest on this article to declare.

## About the License