

# Enhancing Forest Fire Prognosticating Using Machine Learning Algorithm – A Survey

K. Mathivanan<sup>1</sup>, G. Bavadharani<sup>2</sup>, R. Nithya<sup>3</sup>

<sup>1</sup>*K. Mathivanan, Assistant Professor, Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women, Tamil Nadu, India*

<sup>2</sup>*G. Bavadharani, Student, Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women, Tamil Nadu, India*

<sup>3</sup>*Dr.R. Nithya, Associate Professor, Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women, Tamil Nadu, India*

**Abstract**—Forest fires pose significant threats to ecosystems, human health, and economies, making accurate prediction crucial for effective management and mitigation. Traditional fire prediction models often rely on meteorological data, historical fire records, and expert systems, which face limitations in terms of precision and adaptability. Recent advances in machine learning (ML) offer promising solutions by enabling the analysis of large, complex datasets and the recognition of intricate patterns that predict fire occurrence and behavior. This survey paper explores the application of various ML techniques, including supervised (e.g., decision trees, support vector machines), unsupervised (e.g., clustering, PCA), and hybrid models (e.g., ensemble methods, deep reinforcement learning) for enhancing fire prognostication. We examine their effectiveness in fire risk mapping, behavior modeling, and real-time prediction, highlighting both the strengths and challenges. Despite notable progress, issues related to data quality, model interpretability, and scalability persist, suggesting that further research is needed to integrate ML with traditional fire management practices and improve model transparency. This paper also identifies future directions, including the incorporation of real-time data, hybrid ML systems, and cross-disciplinary approaches to create more adaptive and reliable fire prediction tools.

**Index Terms**—Machine Learning, Forest Fire Prediction, Forest Fire Fighting, Formation Control, Random Forest, Support Vector Machine.

## I. INTRODUCTION

Forest fires are one of the most destructive natural disasters, causing significant damage to ecosystems, biodiversity, air quality, and human livelihoods. They

are increasingly frequent and intense due to a combination of factors, including climate change, human activities, and land-use changes. Over the years, the frequency of forest fires has escalated, resulting in severe ecological, social, and economic consequences. The effect of these fires is the same-reaching: they not only destroy large areas of forest but also disrupt wildlife habitats, reduce biodiversity, and degrade soil quality, which in This shift can lead to lasting harmful impacts on the environment. In addition, forest fires often contribute to air pollution, releasing large quantities of carbon dioxide, particulate matter, and other harmful gases into the atmosphere. This leads to poor air quality, affecting both human health and the broader environment. Human communities living in or near forests are at considerable risk from forest fires. These fires can destroy homes, businesses, and infrastructure, leading to economic instability and loss of livelihoods. Moreover, the costs associated with firefighting efforts, recovery, and rehabilitation are substantial. The damage to forest ecosystems also affects the livelihoods of people dependent on forests for resources such as timber, fuelwood, and non-timber forest products. Additionally, forest fires can disrupt social structures, particularly in areas where communities rely on forests for cultural practices and traditional knowledge. In regions like North-East India (NEI), forest fires are particularly frequent and pose a significant challenge to sustainable land management. The unique climatic and geographical conditions in NEI, combined with human activities such as slash-and-burn agriculture, make the region highly vulnerable to fire outbreaks. With forest cover

constituting 19.27% of the land area, the risk of fires spreading rapidly is high, especially during dry seasons. This necessitates the development of effective fire management strategies, which take into account the diverse environmental, social, and economic factors that influence fire occurrence and behavior. Understanding the underlying causes and risk factors of forest fires is critical for developing mitigation measures that can help reduce their frequency and impact. Traditional fire prediction models have often relied on statistical techniques and expert judgment, which, although useful, have limitations when it comes to handling large and complex datasets.

These methods typically focus on historical fire data, weather conditions, and land-use factors to estimate the likelihood of a fire starting or spreading. However, they may not fully capture the complexities of fire behavior, such as how different variables interact with each other to influence fire risk. In recent years, machine learning algorithms have shown promise in improving fire prediction accuracy by handling large volumes of data and identifying patterns that might not be instantly visible through conventional methods. Two popular machine learning algorithms used in fire prediction are Random Forest and Support Vector Machine (SVM). Random Forest is an ensemble learning technique that constructs several decision trees simultaneously.

Each tree is constructed using a random subset of the input features, and the final prediction is made by aggregating the outputs of all the individual trees. This approach helps reduce the risk of overfitting, making it particularly effective for handling large, complex datasets with multiple variables. Random Forest can capture nonlinear relationships between features and the target variable, making it useful for predicting fire-prone areas and estimating fire risk based on various environmental factors such as temperature, humidity, wind speed, and vegetation density. Support Vector Machine (SVM) is another powerful supervised learning technique that excels in classification tasks, particularly binary classification problems like predicting the occurrence of a fire. SVM works by finding an optimal hyperplane that separates data into different classes—such as fire versus no-fire. This hyperplane is chosen to maximize the margin between the two classes, ensuring that the classification is as accurate as possible. SVM is particularly effective

when the data is high-dimensional, meaning that there are many variables or features that influence the fire risk. It can handle complex, non-linear relationships between these variables, making it a suitable choice for predicting fire behavior under different conditions. In fire prediction, combining Random Forest and SVM can provide a more comprehensive and accurate framework. While Random Forest excels in handling diverse datasets with multiple variables, SVM is highly effective at distinguishing between classes when the data has clear margins

## II. LITERATURE SURVEY

<sup>[1]</sup> Zechuan Wu The paper introduces a model to predict how forest fires spread, called the Forest Fire Spread Behavior Prediction (FFSBP) model. It has two parts: one predicts the direction and speed of the fire spread (FFSPP model) using two modeling techniques, and the other predicts the burned area (FFSRP model) using machine learning. The models were tested on real fire cases in China and Portugal, showing that they perform better than some existing models. The FFSPP model is especially good at predicting fire spread, while the FFSRP model works well for small to medium fires. However, the study points out that the models still need more data and real-world testing to improve their accuracy.

<sup>[2]</sup> Nazimur Rahman Talukdar Forest fires in Northeast India are increasing, especially during the pre-monsoon season, and are strongly influenced by human activities like farming and shifting cultivation. The areas most at risk are districts with tree cover between 60% and 80%, especially in lower-elevation forests with moderate slopes. The Lawngtlai, Dhalai, and Ri-Bhoi districts are the most affected, and similar risks exist in protected areas. A stronger management plan, using geospatial technology and local community involvement, is needed to predict and prevent fires, as most are caused by human actions.

<sup>[3]</sup> Kajol R Singh Forest fires are a major environmental threat, and predicting them accurately is key to prevention. Traditional methods like SVM (Support Vector Machine) have limitations, especially with large datasets, due to overfitting. This study develops a parallel SVM model using Apache Spark to improve prediction efficiency and reduce computational time. The model uses weather data, including the Forecast Weather Index (FWI), and

performs well on data from India and Portugal. It also includes an alert system to notify authorities in real time, aiming to improve forest fire prevention and response in the future.

<sup>[4]</sup> Ananya Shahdeo Wildfires are a growing environmental threat, and this paper proposes a solution to predict the likelihood of wildfires using factors like weather, terrain, vegetation, and fire weather indices. The prediction is done using machine learning, which helps identify high-risk areas for wildfires, enabling early interventions to reduce damage. Additionally, the solution uses satellite images to detect actual wildfires, testing the accuracy of predictions with real fire data. The model uses data from NASA's MODIS system, which includes meteorological, topographical, and vegetation data. The Random Forest algorithm is used for prediction because of its high accuracy, ability to handle large datasets, and its capacity to estimate missing data. In the future, the process of mapping the prediction data to satellite images can be automated.

<sup>[5]</sup> Quansheng Hai The research outlined Focuses on predicting and understanding forest fires in Central-South China using advanced technology. It combines Geographic Information Systems (GIS) with machine learning, specifically the LightGBM model, to analyze factors like weather, terrain, vegetation, and human activities that contribute to fire risk. The model accurately forecasts the likelihood of fires, with high accuracy in identifying fire-prone areas and predicting seasonal and regional variations in fire risk. The study shows that forest fire risks are higher in certain months and locations, helping to guide better resource allocation and fire prevention efforts. Overall, it emphasizes the importance of using advanced data analysis and technology for effective forest fire management and prevention.

<sup>[6]</sup> Yashnil Mohanty Forest fires in the United States pose serious risks, making early and accurate predictions essential for effective management. This study used the "Algerian Forest Fires Dataset" to develop a machine learning model for predicting forest fires. The model, built using Random Forest Classifier (RFC), was trained on five key meteorological factors: month, temperature, humidity, wind, and rain. Compared to other models like Decision Tree, Logistic Regression, and Artificial Neural Networks, the RFC showed slightly better performance, with an accuracy of 86.5% and an F1 score of 88.9%. This

approach provides a probability of fire occurrence, which can help forest management agencies prioritize high-risk areas and take timely preventive actions.

<sup>[7]</sup> Arunkumar V Explores various machine learning techniques for predicting forest fires in India, using factors like temperature, humidity, wind speed, and location (latitude and longitude). Several classification algorithms, including decision trees, logistic regression, support vector machines (SVM), and Random Forest with Bagging, were compared. The goal is to build a predictive model that can alert people about potential forest fires based on their location (zip code). The model was tested using data from various regions, and Python was used for model training and evaluation. Additionally, the study discusses key techniques like data pre-processing, feature selection, and hyper parameter tuning, aimed at improving the accuracy and effectiveness of the forest fire prediction system. The final model is intended to help forest management agencies by predicting fire risks and providing real-time alerts to local authorities and tourists.

<sup>[8]</sup> Jakkamsetti Bharath The research focuses on predicting forest fires using machine learning techniques like Random Forest, Decision Tree, and Logistic Regression. The model uses weather data, including temperature, humidity, wind speed, and Fire Weather Index (FWI) components to predict fire likelihood. The dataset, sourced from Algerian Forest Fires, contains data on fire incidents from June to September 2012. After cleaning and processing the data, various classification models are trained and tested to determine the best performer. A web application will allow users to input their location and receive real-time fire alerts based on the model's predictions. This system can help forest authorities manage resources and respond proactively to fire risks. The project also includes exploratory data analysis (EDA) to identify key factors influencing fire occurrences, ensuring a more accurate prediction model.

<sup>[9]</sup> Hoang Thi Hang Introduces a new method for predicting forest fire risk in Nainital district using a combination of machine learning models, with a Deep Neural Network (DNN) as the main model. The approach combines several strong models like AdaBoost, XGBoost, and Random Forest, improving prediction accuracy through an ensemble method. It also uses Explainable AI (XAI) techniques, such as

SHAP and LIME, to make the models more understandable and to identify key factors like annual rainfall and evapotranspiration that influence fire risk. While the results are promising, future studies could improve accuracy by adding more variables and considering spatial patterns in the data.

<sup>[10]</sup> Mriganka Shekhar Sarkar Examines forest fire risk in North-East India (NEI) using five machine learning models and an ensemble method. It identifies key

factors driving fire susceptibility, such as solar radiation, vegetation, human population density, and traditional farming practices. The Random Forest model showed the best performance, with an AUC of 0.87, helping to map areas at high risk of forest fires. The findings provide valuable insights for targeted fire management and can be applied globally using satellite data, GIS, and machine learning techniques to improve forest fire prevention and response strategies.

### III. PERFORMANCE METRICS

Study	Model/Method	Dataset	Accuracy	Analysis
Fire Spread Prediction	Artificial Neural Networks (ANNs)	Historical Fire Data, Environmental Data (weather, vegetation)	Accuracy, Precision & Recall, F1-Score	Adjust the model with old data, fine-tune the settings, and test it using actual fire events
Forest Fires in Northeast India	Parallel SVM (Support Vector Machine)	Local data from Northeast India	Efficient with large datasets, reduces overfitting	Uses Apache Spark for parallel processing, alerts for authorities.
Wildfire Prediction (Global Focus)	Random forest	NASA MODIS (satellite data)	High accuracy in wildfire prediction, data handling	Uses weather, terrain, vegetation data to predict fires.
Forest Fires in Central-South China	LightGBM (Machine Learning)	GIS, weather, terrain, vegetation	High accuracy, identifies fire-prone areas, seasonal patterns	Accurate for forecasting regional fire risks, improves resource allocation.
Forest Fires in the United States	Random Forest Classifier (RFC)	Algerian Forest Fires Dataset	Accuracy = 86.5%, F1 score = 88.9%	Better than Decision Tree, Logistic Regression, and Neural Networks.
Forest Fires in India	Decision Trees, Logistic Regression, SVM, Random Forest with Bagging	Data from various regions of India	Model trained using various ML algorithms and data pre-processing	Designed for real-time alerts for authorities and tourists.
Forest Fire Prediction with Weather Data	Random Forest, Decision Tree, Logistic Regression	Algerian Forest Fires Dataset	Accuracy after cleaning and processing; classification models	Explores weather data, FWI, and classification models.
Forest Fire Risk in Nainital District	Deep Neural Network (DNN), AdaBoost, XGBoost, RF(ensemble)	Local weather and environmental data	Increases accuracy, key factors using SHAP and LIME	Uses Explainable AI to make models interpretable.

Forest Fire Risk in North-East India	Random Forest, Ensemble Method (5 models)	Local data, GIS, machine learning	AUC = 0.87, best performance with Random Forest	Uses key factors like solar radiation, vegetation, and human population.
Forest Fire Prediction Using Machine Learning	Random Forest, Decision Tree, Logistic Regression	Algerian Forest Fires Dataset	Accuracy = 86.5%, F1 Score = 88.9%	The model is trained using weather data from June to September 2012.

#### IV. OVERVIEW

The Fire Spread Prediction study uses ANNs with historical data and environmental factors, measuring performance with accuracy and F1-score. The Northeast India study employs Parallel SVM with local data and Apache Spark for fast processing and alerts. Wildfire Prediction combines Random Forest with NASA MODIS data for accurate predictions. Forest Fires in the U.S. achieves 86.5% accuracy and 88.9% F1-score using Random Forest. Other studies, including those in India, China, and Nainital, use models like LightGBM, DNN, and XGBoost for improved fire risk prediction and resource planning.

#### V. CONCLUSION

Machine learning models like Random Forest, SVM, and Light GBM are effective at predicting forest fires by looking at factors such as weather, terrain, vegetation, and human activities. These models can predict fire risks, track fire spread, and identify areas at high risk. By using real-time data from satellites and weather systems, the models can make accurate predictions to help manage and prevent fires. However, challenges such as limited data, overfitting, and scalability need to be addressed. Future work will focus on improving accuracy, using more local and real-time data, and handling large-scale predictions. The use of Explainable AI (XAI) has been helpful in making the models more transparent for decision-makers. Going forward, improving these models with more data and advanced techniques, like combining multiple models (ensemble methods), will make them more reliable and effective for global fire management.

#### VI. REFERENCE

- [1] Zechuan Wu, Bin Wang, Mingze Li, Yuping Tian, Ying Quan, Jianyang Liu, Simulation of forest fire spread based on Artificial Intelligence, Ecological Indicators, 2022, Ecological Indicators 136 (2022) 108653.
- [2] Nazimur Rahman Talukdar, Firoz Ahmad, Laxmi Goparaju, Parthankar Choudhury, Rakesh Arya, Abdul Qayum, Javed Rizvi, Forest fire estimation and risk prediction using multispectral satellite Images, Natural Hazards Research, 2024, Natural Hazards Research 4 (2024) 304–319.
- [3] Kajol R Singh, K.P. Neethu, K Madhurekaa, A Harita, Pushpa Mohan, Parallel SVM model for forest fire prediction, Soft Computing Letters, 2021, Soft Computing Letters 3 (2021) 100014
- [4] Ananya Shahde, Aditi Shahdeo, Prakruthi S Reddy, Chaitra K, Wildfire Prediction and Detection using Random Forest and Different Color Models, International Research Journal of Engineering and Technology (IRJET), Volume: 07 | June 2020
- [5] Quansheng Hai, Xiufeng Han, Battsengel Vandansambuu, Yuhai Bao, Byambakhuu Gantumur, Sainbuyan Bayarsaikhan, Narantsetseg Chantsal and Hailian Sun, Predicting the Occurrence of Forest Fire in the Central-South Region of China, MDPI (Multidisciplinary Digital Publishing Institute) Forests 2024, 15, 844. <https://doi.org/10.3390/f15050844>.
- [6] Yashnil Mohanty, A Machine Learning Approach to Predict the Occurrence of Forest Fires with Meteorological Parameters, high school edition journal of student research, Volume 13 |(2024).
- [7] Arunkumar, Kanimozhi, Forest Fire Prediction Using Machine Learning Techniques, International Journal of Scientific Research in Engineering and Management (IJSREM),

- [8] Volume 07 | (2023), DOI: 10.55041/IJSREM19604.
- [9] Jakkamsetti Bharath, K.Swapna, Forest Fire Prediction Using Supervised Machine Learning Models, IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES, Volume 13, 2024.
- [10] Hoang Thi Hang, Javed Mallick, Saeed Alqadhi, Ahmed Ali Bindajam, Hazem Ghassan Abdo, Exploring forest fire susceptibility and management strategies in Western Himalaya: Integrating ensemble machine learning and explainable AI for accurate prediction and comprehensive analysis, Sciencedirect, Volume 35, August 2024, 103655.
- [11] Mriganka Shekhar Sarkar, Bishal Kumar Majhi, Bhawna Pathak, Tridipa Biswas, Soumik Mahapatra, Devendra Kumar, Indra D. Bhatt, Jagadish C. Kuniyal, Sunil Nautiyal, Ensembling machine learning models to identify forest fire-susceptible zones in Northeast India, Sciencedirect, Volume 81, July 2024, 102598.
- [12] N. Ntinopoulos, M. Spiliotopoulos, L. Vasiliades, and N. Mylopoulos, "Contribution to the study of forest fires in semi-arid regions with the use of Canadian fire weather index application in Greece," *Climate*, vol. 10, no. 10, p. 143, Sep. 2022.
- [13] X. Deng, Z. Zhang, F. Zhao, Z. Zhu, and Q. Wang, "Evaluation of the regional climate model for the forest area of Yunnan in China," *Frontiers Forests Global Change*, vol. 5, Jan. 2023, Art. no. 1073554.
- [14] Y. Pang, Y. Li, Z. Feng, Z. Feng, Z. Zhao, S. Chen, and H. Zhang, "Forest fire occurrence prediction in China based on machine learning methods," *Remote Sens.*, vol. 14, no. 21, p. 5546, Nov. 2022.
- [15] C. Gao, H. Lin, and H. Hu, "Forest-fire-risk prediction based on random forest and backpropagation neural network of Heihe area in Heilongjiang province, China," *Forests*, vol. 14, no. 2, p. 170, Jan. 2023.
- [16] C. Shi and F. Zhang, "A forest fire susceptibility modeling approach based on integration machine learning algorithm," *Forests*, vol. 14, no. 7, p. 1506, Jul. 2023.
- [17] Yanzhi Li, Guohui Li, Kaifeng Wang, Zumin Wang and Yanqiu Chen, "Forest Fire Risk Prediction Based on Stacking Ensemble Learning for Yunnan Province of China" *MDPI fire*, 2023.
- [18] A. R. Ervilha, J. M. C. Pereira, and J. C. F. Pereira, "On the parametric uncertainty quantification of the Rothermel's rate of spread model," *Appl. Math. Model.*, vol. 41, pp. 37–53, Jan. 2017.
- [19] Y. Zhang and L. Tian, "Examining and reforming the Rothermel surface fire spread model under no-wind and zero-slope conditions for the Karst ecosystems," *Forests*, vol. 14, no. 6, p. 1088, May 2023.
- [20] M. Fečkan and J. Pačuta, "Existence of solution of a forest fire spread model," *Appl. Math. Lett.*, vol. 83, pp. 227–231, Sep. 2018.
- [21] U. Kc, J. Aryal, S. Garg, and J. Hilton, "Global sensitivity analysis for uncertainty quantification in fire spread models," *Environ. Model. Softw.*, vol. 143, Sep. 2021, Art. no. 105110