iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** V **Month of publication:** May 2025

DOI: https://doi.org/10.22214/ijraset.2025.71034

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Enhancing Forest Fire Prognosticating Using Machine Learning Algorithms

K. Mathivanan[1], G. Bavadharani[2], R. Nithya[3]
[1]Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women, Tamil Nadu, India
[2]PG Scholar, Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women, Tamil Nadu, India
[3] Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women, Tamil Nadu, India

Abstract: Introduces a machine learning-based approach that uses certain geographic and environmental characteristics to categorize different types of forest cover. The Random Forest Classifier technique, which effectively manages classification problems by building many decision trees and combining their output for increased accuracy and robustness, is used to design the model. Important input variables including height, aspect, slope, distances to roads and hydrology, hill shade values, and particular soil and wilderness kinds are all taken into account by the algorithm. To generate predictions, users can manually enter numbers or upload datasets using an intuitive interface created using Streamlit. Apart from producing forecasts, the system offers assessment metrics like accuracy, precision, recall, and F1 score, which are bolstered by visual aids like metric bar charts and confusion matrix plots. Through a dependable and understandable machine learning model, this structured prediction approach provides insights for applications involving the analysis and classification of forest data.
Keywords: Random Forest Classifier, Forest Cover Type Prediction, Machine Learning, Feature-Based Classification, Environmental Data Analysis

I. INTRODUCTION

In order to analyze structured data and identify patterns, it is important to comprehend how different types of forest cover are classified according to topographical and environmental characteristics. It is feasible to distinguish different types of forest land by looking at quantifiable elements including height, slope, aspect, soil composition, and spatial distances. Predictive models that can accurately and efficiently classify cover kinds can be created by using machine learning approaches to such data. In order to create a model that can accurately classify data based on structured input attributes, this research focuses on integrating statistical techniques with algorithmic learning. In the context of forest-type classification, the method facilitates informed analysis and improves the capacity to identify data patterns.

A. Random Forest Classifier

A popular ensemble machine learning approach called the Random Forest Classifier aggregates the predictions of several decision trees to carry out categorization. To introduce variation among the models, each decision tree is trained on a randomly selected portion of the training data and a randomly selected set of features. Majority voting across all trees determines the output during the prediction phase, which helps to increase accuracy and stability. This method is resistant to noise and overfitting, especially when dealing with complicated variable interactions or large dimensionality datasets. Additionally, the algorithm offers insights into feature relevance, making it possible to pinpoint the variables that have the most effects on the classification result. It is a popular option for many classification tasks involving structured datasets because of its strong generalization across a wide range of data.

B. Forest Cover Type Prediction

The method of classifying land areas according to particular forest characteristics using input variables obtained from topographical and environmental data is known as forest cover type prediction. These features could include distance-based variables like proximity to roads and hydrology, as well as topographical parameters like elevation, slope, and aspect. The classifying method is further aided by category indications like wilderness area and soil type. The objective is to classify each sample or site into a predetermined forest cover type class by examining these variables. In addition to being useful for mapping and land assessment, this classification helps to understand the makeup of landscapes. The methodology is based on statistical learning, which finds patterns in labeled datasets and uses them to predict comparable unseen samples.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

C. Machine Learning

Through the use of computation, machine learning allows computers to automatically learn from data and perform better on a given task. Machine learning models find patterns and correlations in datasets to create predictive reasoning, in contrast to rule-based programming, which requires explicit coding of instructions. A fundamental area of machine learning is supervised learning, which uses labeled data with established input-output relationships to train models. A frequent supervised job is classification, in which each input instance is given a discrete category by the model. Machine learning's versatility and scalability, which enable its application in a variety of fields requiring both structured and unstructured data, are its key advantages. In structured data analysis, algorithms like decision trees, support vector machines, and ensemble techniques like random forests are essential for creating efficient classifiers.

D. Feature-Based Classification

In feature-based classification, a data sample's category or class is ascertained by applying a predetermined set of quantifiable factors, or features. Features can be discrete values, like region indications or encoded soil kinds, or continuous values, such altitudes and numerical distances. Because these characteristics have a direct impact on the model's capacity to discriminate across classes, choosing pertinent features is an essential stage in the classification process. To further improve model performance, feature engineering can also be used to convert unstructured data into more insightful representations.

II. LITERATURE REVIEW

The Forest Fire Spread Behavior Prediction (FFSBP) model is introduced to predict the spread of forest fires. It comprises two parts: the FFSPP model, which uses two modeling techniques to predict the direction and speed of fire spread, and the FFSRP model, which applies machine learning to estimate the burned area. These models were evaluated using real fire incidents in China and Portugal, demonstrating superior performance compared to some existing approaches. The FFSPP model is very effective at predicting the spread of fires, whereas the FFSRP model is especially accurate for fires of small to medium size. However, the study notes that further data and real-world validation are needed to enhance their accuracy [1]

In Northeast India, forest fires are becoming more frequent, especially before the monsoon season, largely due to human activities such as agriculture 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 districts of Lawngtlai, Dhalai, and Ri-Bhoi are among the hardest hit, with comparable risks also present in protected regions. 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 [2].

Forest fires pose a significant environmental risk, and accurate prediction is essential for effective 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. change the sentence same meaning. It also includes an alert system to notify authorities in real time, aiming to improve forest fire prevention and response in the future [3].

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. Machine learning is used for prediction, allowing the identification of high-risk wildfire areas and supporting early intervention to minimize damage. Moreover, the approach incorporates satellite imagery to detect real wildfires, validating prediction accuracy with actual 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 [4].

The Model focuses on predicting and understanding forest fires in Central-South China using advanced technology. The approach integrates Geographic Information Systems (GIS) with machine learning, specifically the LightGBM model, to examine factors such as weather, topography, vegetation, and human activities that influence 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 [5].

Forest fires in the United States pose serious risks, making early and accurate predictions essential for effective management. This research utilized the "Algerian Forest Fires Dataset" to build a machine learning model for forest fire prediction. The model, built using Random Forest Classifier (RFC), was trained on Five essential meteorological factors include month, temperature, humidity, wind, and rainfall. 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. [6].

This 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 [7].

The Model focuses on predicting forest fires using machine learning techniques like Random Forest, Decision Tree, and Logistic Regression. The model utilizes weather data, such as temperature, humidity, wind speed, and components of the Fire Weather Index (FWI), to forecast the likelihood of a fire. The dataset, obtained from Algerian Forest Fires, includes information on fire incidents that occurred between June and 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 [8].

The Model 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 [9].

This Method 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. [10]

III. EXISTING SYSTEM

Predicting forest fires is a crucial element in successfully managing and controlling them. This is a serious environmental issue that leads to ecological devastation in the form of a hovered geography of natural coffers that disturbs ecosystem stability, raises the risk of other natural hazards, and depletes coffers that are similar to water, which results in water pollution and global warming. An essential component of managing similar events is fire discovery. It is expected that forest fire vaticination will lessen the impact of future forest fires. The suggested system uses meteorological characteristics like temperature, precipitation, wind, and moisture to forecast the likelihood of a forest fire. Our idea is centered on creating an advanced system that combines machine literacy technology to detect forest fires early. Similar to temperature, moisture, and wind speed, the system uses environmental data to find minute patterns and irregularities that could indicate implicit fire effort. Our technology may provide timely warnings to firefighting agencies and forest operating authorities by assaying these environmental elements in real-time, allowing them to react to emerging fire events quickly and efficiently. We want to improve the delicacy and reliability of our discovery system by integrating sophisticated data processing techniques and predictive modeling algorithms, reducing false alarms and increasing the detection of actual fire incidents. Similarly, by relieving visionary intervention and mitigation stress, the implementation of our system implicitly aims to lessen the devastating effects of backfires on natural ecosystems and mortal communities. Our design seeks to help preserve forests and protect lives and livelihoods threatened by the growing threat of backfires in our changing environment by utilizing technology and data-driven methodologies.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

IV. PROPOSED SYSTEM

The suggested solution uses a machine learning-based methodology to categorize different forms of forest cover. It analyzes a predetermined collection of features, including topographical and environmental aspects like elevation, aspect, slope, and soil type, using the Random Forest Classifier algorithm. After processing the data which can be entered manually or through structured file input the algorithm predicts a classification label for each instance. To evaluate the model's performance, Performance metrics such as accuracy, precision, recall, and F1 score are assessed.. To give a clear picture of the classification findings, visual aids like bar graphs and confusion matrices are also included. With the help of a trained prediction model, this framework guarantees a methodical and effective way to provide accurate classifications.

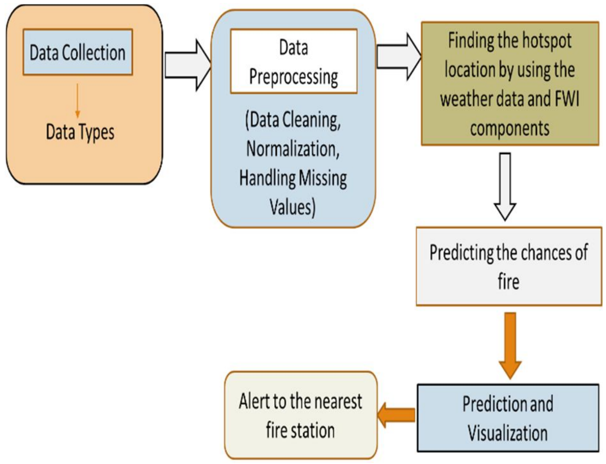

Figure : 1 System Architecture

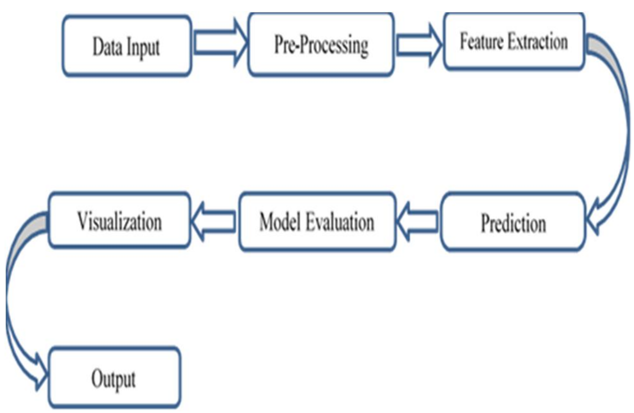

Figure : 2 Module Design

A. Data Input

This module is in charge of gathering user input data. Users have two options: manually enter the necessary characteristics in text boxes or upload a CSV file with the data. The system reads a file into a Pandas Data Frame upon upload and determines if it contains all the features needed for prediction. When manual input is used, users are asked to provide the values for every feature before the information is arranged in a Data Frame. After being collected, the data proceeds to the preparation phase.

B. Pre-Processing

The input data is prepared for prediction by the preprocessing module. To make sure the data is in the right format for the machine learning model, it uses a number of data transformation techniques. This involves using Standard Scaler to scale numerical features, One Hot Encoder to encode categorical features, and imputation approaches (mean for numerical features, most frequent for categorical features) to handle missing values. Preprocessing guarantees that the data is standardized, cleaned, and prepared for forecasting.

C. Prediction

The prediction module takes over after the input data has been pre-processed. Based on the pre-processed input, it generates predictions using the trained machine learning model that has been stored in memory. The predicted forest cover types generated by the model are added to the original input data. The predicted value is shown to the user for manual input, and the predictions are shown in a new column for file uploads. Prediction results can also be downloaded by the user in CSV format for additional analysis.

D. Model Evaluation

This module offers evaluation metrics to gauge the trained model's effectiveness. It computes important evaluation metrics as accuracy, precision, recall, and F1 score after loading model predictions (y_pred) and pre-saved test data (y_test) from disk. For ease of interpretation, these measures are presented as text. To see how well the model performs in identifying the data, a confusion matrix is also presented. A warning message appears if the required evaluation data cannot be located.

V. VISUALIZATION

The model's evaluation findings are shown graphically using the visualization module. To better comprehend the model's prediction accuracy across several classes, it contains a confusion matrix that is shown as a heat map. In order to display the model's performance based on important assessment criteria including accuracy, precision, recall, and F1 score, a bar graph is also created. This helps users to visualize how well the model performs and how effective it is.

VI. RESULT ANALYSIS

Assessing the model's performance and prediction accuracy using the given data is the main goal of the outcome analysis. Using a variety of evaluation indicators, the main goal is to determine how well the model classifies the kind of forest cover. Accuracy, precision, recall, and F1 score are among the metrics that together provide a comprehensive understanding of the model's effectiveness in distinguishing between classes. While precision and recall emphasize the model's capacity to accurately identify positive occurrences and its success in avoiding false negatives, accuracy gauges the overall correctness of the predictions made When Taking both false positives and false negatives into account, the F1 score, which represents the harmonic mean of precision and recall, acts as a balanced metric.

VII. CONCLUSION

To sum up, the forest cover type prediction system shows promise in efficiently categorizing different types of cover according to specific environmental characteristics. In addition to making predictions, the system uses a machine learning model and integrates assessment metrics including accuracy, precision, recall, and F1 score to assess the model's performance. Together with the confusion matrix and visualizations, the evaluation results provide insightful information about the model's advantages and shortcomings. The model can be improved to produce predictions with greater accuracy and dependability by comprehending these criteria. In order to keep the model efficient and flexible for upcoming use cases, this method emphasizes the significance of ongoing development and optimization in machine learning systems.

VIII. FUTURE WORK

There are numerous chances to improve the method for predicting the kind of forest cover in the future. Adding more characteristics is one such approach that can offer a more thorough comprehension of the variables affecting cover kinds. Predictions may be more accurate if the dataset is expanded to incorporate a wider range of environmental factors. Model performance could also be further enhanced by experimenting with more sophisticated machine learning methods and model optimization strategies. The creation of a more interactive user interface that facilitates prediction input and visualization could be another area of emphasis. Additionally, incorporating model deployment solutions would make it possible to smoothly include this system into broader ecosystems, increasing its usability for a variety of user types. Maintaining the model's applicability and efficacy throughout time will require constant assessment and improvement.

REFERENCES

[1] 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

[2] According to Nhu, V.-H., Janizadeh, S., Avand, M., Chen, W., Farzin, M., Omidvar, E., Shirzadi, A., Shahabi, H., Clague, J.J., Jaafari, A., et al. A Comparison of Computational Ensemble Data Mining Models for GIS-Based Gully Erosion Susceptibility Mapping. Appl. Sci. 2020, 10, 2039. [Reference]

[3] Arunkumar, Kanimozhi, Forest Fire Prediction Using Machine Learning Techniques, International Journal of Scientific Research in Engineering and Management (IJSREM).

[4] Bui, D.T.; Khosravi, K.; Tiefenbacher, J.; Nguyen, H.; Kazakis, N. Using innovative hybrid machine-learning techniques to improve water quality index prediction. Environ. Sci. Total. 2020, 721, 137612. [Reference] [Med]

[5] Deepenbacher, J.P.; Pradhan, B.; Tien Bui, D.; Arabameri, A.; Chen, W.; Blaschke, T. Water 2020, 12, 16. Gully head-cut distribution modeling with machine learning techniques—A case study in NW Iran.

[6] 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, Science direct, Volume 35, August 2024, 103655.

[7] In Da Lat City, Lam Dong Province, Vietnam, Thanh, D.Q.; Nguyen, D.H.; Prakash, I.; Jaafari, A.; Nguyen, V.-T.; Van Phong, T.; Pham, B.T. employed a frequency ratio approach based on Gis for mapping landslide susceptibility. Earth Sci. Vietnam J. 2020, 42, 55–66.

[8] Jakkamsetti Bharath, K. Swapna, Forest Fire Prediction Using Supervised Machine Learning Models, IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES, Volume 13, 2024.

[9] J. Shi and J. Zhang, ‘‘Load forecasting based on multi-model by stacking ensemble learning,’’ Proc. CSEE, vol. 39, no. 14, pp. 4032–4042, 2019.

[10] 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

[11] Li, S.; Hong, H.; Wang, X.; Bian, H.; Zhang, S.; Pradhan, B.; Li, Y.; Xue, W.; Shahabi, H.; Li, S. Chen, W. utilizing data-driven techniques such as random forest, alternating decision tree, and naïve bayes tree to model flood susceptibility. Environmental Science, 701, 134979, 2020.

[12] M. Seddouki, M. Benayad, Z. Aamir, M. Tahiri, M. Maanan, and H. Rhinane, ‘‘Using machine learning coupled with remote sensing for forest fire susceptibility mapping. Case study Tetouan province, Northern Morocco,’’ Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci., vol. XLVIII-4/W6, pp. 333–342, Feb. 2023.

[13] 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.

[14] Performance evaluation of machine learning algorithms for forest fire modeling and prediction, Symmetry 12 (6) (2020) 1022, B.T. Pham, A. Jaafari, M. Avand, N. Al-Ansari, T. Dinh Du, H.P.H. Yen, T.V. Phong, D.H. Nguyen, H.V. Le, D. Mafi-Gholami, I. Prakash.

[15] Phong, T.V.; Ly, H.-B.; Le, T.-T.; Trinh, P.T.; Dao, D.V.; Jaafari, A.; Bayat, M.; Mafi-Gholami, D.; Qi, C.; Moayedi, H.; et al. A deep learning neural network model with spatial explicitness for landslide susceptibility prediction. 2020, Catena, 188, 104451.

[16] 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.

[17] R. Alkhatib, W. Sahwan, A. Alkhatieb, and B. Schütt, ‘‘A brief review of machine learning algorithms in forest fires science,’’ Appl. Sci., vol. 13, no. 14, p. 8275, Jul. 2023.

[18] Shahabi, H.; Geertsema, M.; Khosravi, K.; Amini, A.; Bahrami, S.; Shirzadi, A.; Ghaderi, K.; Omidvar, E.; Al-Ansari, N.; Clague, J.J. Using machine learning and sentinel-1 remote sensing data for flood detection and susceptibility mapping: Bagging ensemble hybrid intelligence based on k-nearest neighbor classification. 2020; Remote Sens. 12, 266.

[19] Shallow Landslide Susceptibility Mapping by Random Forest Base Classifier and its Ensembles in a Semi-Arid Region of Iran Nhu, V.-H.; Shirzadi, A.; Shahabi, H.; Chen, W.; Clague, J.J.; Geertsema, M.; Jaafari, A.; Avand, M.; Miraki, S.; Asl, D.T. 2020, 11, 421; Forests. [Reference]

[20] Tiefenbacher, J.P.; Pradhan, B.; Jahani, A.; Goshtasb, H.; Kornejady, A.; Shahabi, H.; Rahmati, O.; Panahi, M.; Ghiasi, S.S.; Deo, R.C.; et al. Neural fuzzy hybrid ensembles for modeling and forecasting dust sources. 2020, Atmos. Environ. 224, 117320. [Reference]

[21] Volume 07 | (2023), DOI: 10.55041/IJSREM19604.

[22] 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).

[23] Z. Wang, K. Wang, Y. Li, and G. Li, ‘‘Research on forest fire prediction in Yunnan province based on LightGBM and SHAP,’’ Fire Science Technol., vol. 42, no. 11, pp. 1567–1571, 2023.

[24] 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.

[25] Z. Zhou, Integrated Learning: Fundamentals and Algorithms. Beijing, China: Publishing House of Electronics Industry, 2020.