Journal of Industrial Information Technology and Application *ISSN(Online): 2586-0852*
Vol.9 No. 4

Research Study on Forest Fire Prediction System Using KNN

Dr.N.M Sangeetha [1)], Sathya Seelan S.A [2)], and Sanjay S [2,)]*

[1)] Assistant Professor Department of Computer Science (UG&PG) Dwaraka Doss Goverdhan Doss Vaishanav College, Chennai, TN, India

[2)] B.Sc Computer Science 2nd Year Department of Computer Science (UG&PG) Dwaraka Doss Goverdhan Doss Vaishanav College, Chennai, TN, India

Abstract. Forest fires are one of the most disastrous natural calamities causing significant harm to human life and property. Early detection of wildfires is important to prevent their spread and minimize their damage. Wildfire research is an important environmental and scientific issue, especially in wildfire-prone regions such as South Carolina (SC). Machine Learning techniques offer an opportunity to improve wildfire prevention despite the limited research on forest fires. This application of machine learning techniques presents an opportunity to enhance forest fire prevention and control efforts. In this paper, we present a comparative study of four popular ML methods tree, random forest, k-nearest neighbors (KNN), and support vector machine (SVM)- for forest fire detection.

Keywords; Forest Fires, Machine Learning

Cite this paper as: Dr.N.M Sangeetha, Sathya Seelan S.A, and Sanjay S (2025) "Research Study on Forest Fire Prediction System Using KNN", Journal of Industrial Information Technology and Application, Vol. 9. No. 4, pp. 1188 - 1194

[*] Corresponding author: Sanjaysridhar156@gmail.com
Received: Aug. 18. 2025 Accepted: Nov. 05. 2025 Published: Dec. 31. 2025

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright©2017. Journal of Industrial Information Technology and Application (JIITA)

1. Introduction

Forest fires pose significant threats to ecosystems, human health, and economies, with their frequency and severity increasing due to climate change and human activity. Accurately predicting these fires minimizes damage and enhances emergency response efforts. Recent advancements in machine learning (ML) offer promising solutions for forest fire prediction by leveraging large datasets that include environmental factors such as weather conditions, vegetation types, and historical fire incidents. This paper explores using ML algorithms to develop predictive models for forest fires, aiming to identify key risk factors and improve decision-making in fire prevention and management [1].

This research seeks to enhance the ability to forecast and mitigate the impact of forest fires in vulnerable regions by integrating data-driven approaches [9].

2. Methodology

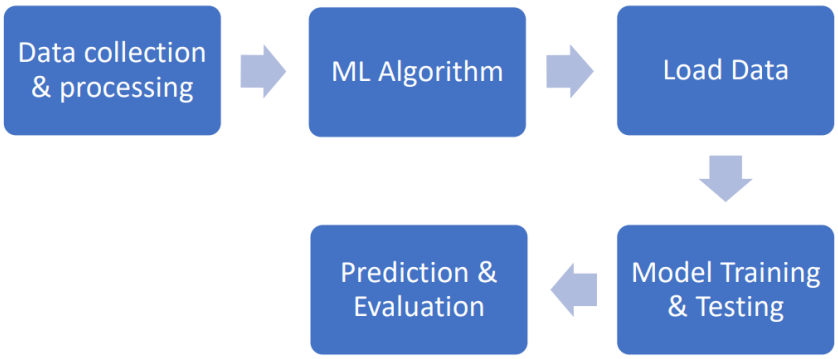

Figure 1. Machine Learning (ML) Workflow

2.1 Data Collection

Data collection involves gathering key information such as weather data (temperature, humidity, wind speed), historical fire data (locations and times of past fires), satellite imagery, and terrain information (elevation and slope). This data forms the foundation for the model to identify patterns that influence fire occurrence and behavior [4].

2.2 Data Processing

The collected data is first cleaned by handling missing or inconsistent values in data processing. Feature engineering is then performed to create new, meaningful variables, such as fire risk indices. The data is then normalized and scaled to ensure all features are in a comparable range, which helps improve model accuracy. Finally, the data is encoded to convert categorical variables into numerical format [13].

2.3 Machine Learning Algorithm

The ML algorithm is crucial in learning patterns from the processed data to predict fire risks. Algorithms like Random Forest, Support Vector Machines (SVM), and Neural Networks analyze features such as weather conditions, vegetation type, and historical fire data to identify high-risk areas [5]. The selected algorithm helps classify or forecast fire occurrences based on realtime inputs.

2.4 Model Training

Model training in a forest fire prediction system involves feeding historical data to the machine learning algorithm to learn patterns related to fire risks. The model adjusts its parameters through techniques like cross-validation to improve accuracy and avoid overfitting. This process enables the model to generalize and make reliable predictions on new data [12][5].

2.5 Model Testing

Model testing in a forest fire prediction system involves evaluating the trained model on a separate test dataset that it has not seen before. This helps assess the model's accuracy and performance in predicting fire risks on new, unseen data. The results are measured using metrics like precision, recall, and F1-score to determine the model's effectiveness

2.6 Prediction and Evaluation

Prediction and evaluation in a forest fire prediction system involve using the trained model to predict fire risks based on real-time data. The model's predictions are then compared with actual outcomes to assess its accuracy and reliability. Evaluation metrics like accuracy, precision, recall, and F1- score help determine the model's performance and areas for improvement [6][8].

3. Workflow For a Forest Fire Prediction Model

A simple example of how a machine learning model for predicting forest fires might be implemented using Python. The program makes use of the RandomForestClassifier to predict whether a fire will occur based on environmental features [3][6].

STEP 1: Import the necessary libraries

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report
```

- pandas: Used to handle data manipulation tasks, such as loading datasets and manipulating DataFrames.
- train_test_split: This helps to split the dataset into two parts: one for training the model and the other for testing it.
- RandomForestClassifier: This is used for classification tasks. It builds multiple decision trees and combines their results.

STEP 2: Load Data

```
data = pd.read_csv('forest_fire_data.csv')
```

- data: This line loads the dataset from a CSV file (forest_fire_data.csv). The dataset contains information about various environmental features (e.g., temperature, humidity) and the target variable (whether a forest fire occurred, labeled as 1 for fire and 0 for no fire).

STEP 3: Data Preprocessing

```
X = data.drop('target', axis=1)  # Features
y = data['target']  # Target variable
```

- X: These are the independent variables (the factors that influence the fire occurrence).

- Y: This is the target variable representing the dependent variable, which is the outcome (fire or no fire). It is assigned the values from the 'target' column in the dataset.

STEP 4: Split data into training and testing sets

- X_train and y_train are the features and target variables for training the model.
- X_test and y_test are the features and target variables for testing the model after it has been trained.

STEP 5: Train random forest model

```
model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train, y_train)
```

- RandomForestClassifier: Random Forest is an ensemble learning algorithm that uses multiple decision trees to make predictions, improving accuracy and reducing overfitting
- model.fit(X_train, y_train): This step trains the model using the training data. The model learns the relationship between the input features (X_train) and the output target (y_train).

STEP 6: Make Predictions

```
y_pred = model.predict(X_test)
```

- model.predict(X_test): After training, the model is used to make predictions on the test set (X_test). The predicted values are stored in y_pred. These predictions represent whether the model predicts a fire (1) or no fire (0) based on the input features.

STEP 7: Evaluate Model

```
print("Accuracy:", accuracy_score(y_test, y_pred))
print("Classification Report:\n", classification_report(y_test, y_pred))
```

- accuracy_score(y_test, y_pred): This calculates the accuracy of the model by comparing the predicted values (y_pred) with the actual values from the test

set (y_test). Accuracy is the percentage of correct predictions made by the model.

- classification_report(y_test, y_pred): This generates a detailed classification report, which includes precisions and recall.

Using the above algorithm, we can predict whether a forest fire will occur by collecting and processing the data.

4. Challenges and Limitations

- Gathering comprehensive datasets, especially in remote forested areas, can be difficult. Incomplete or missing data can reduce the accuracy of predictions [11].
- Models trained on data from one region may not perform well in other areas due to different climates, vegetation, or geographical conditions. This imbalanced nature of the data can result in the model failing to predict fires that occur, which is critical for public safety [2].
- Many machine learning models, such as deep learning and ensemble methods, lack interpretability, which makes it difficult for forest rangers, and authorities to trust and act on the predictions[10].

5. Conclusion

In conclusion, the use of machine learning for forest fire prediction offers a promising approach to enhancing early warning systems and improving fire management. ML models can provide valuable insights into fire risk by analyzing various environmental factors such as weather conditions, vegetation, and historical fire data. However, challenges like data quality, regional specificity, and model interpretability must be addressed to improve the reliability of predictions. Additionally, integrating real-time data and overcoming class imbalances in datasets will further enhance model performance. Despite these obstacles, ML-based prediction systems have the potential to significantly reduce the impact of wildfires. Future advancements in data collection, model adaptation, and computational power will strengthen these systems. Ultimately, ML holds the key to more proactive and efficient forest fire management.

References

[1] Madden, M., & Rothermel, R. C. Fire behavior and its relation to fire spread. International Journal of Wildland Fire.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., Vol. 2. Oxford: Clarendon, 1892, pp.68–73. (2003).

[2] Chen, Y., Li, D., & Li, J. Forest fire prediction and risk analysis using machine learning techniques: A review. Journal of Environmental Management.K. Elissa, “Title of paper if known,” unpublished. (2018).

[3] Rashid, S. A., & Sultana, N. A survey on machine learning techniques for wildfire prediction. Journal of King Saud UniversityComputer and Information Sciences. (2019).

[4] Jain, R., & Tiwari, A. Predicting forest fire using machine learning algorithms. International Journal of Computer Applications. (2020).

[5] Liu, Q., & Liang, X. A machine learning approach to predict forest fires in a remote sensing context. Remote Sensing of Environment. (2017).

[6] Zhao, Y., & Peng, X. Predicting wildfire risk using machine learning methods: A case study from California. Environmental Modelling & Software. (2021).

[7] Kassambara, A. Machine learning essentials: A practical guide for beginners. Springer. (2017).

[8] De Moura, F. A., & Filho, G. A. Using machine learning techniques to predict forest fire occurrences in the Amazon rainforest. (2019).

[9] Carvalho, F. P., & Sanches, A. P. A data-driven approach for wildfire forecasting: Case study in Portugal. Computers, Environment and Urban Systems. (2020).

[10] Alonso, F., & Trivedi, M. Fire risk prediction and its application in forest fire management. Fire Safety Journal. (2019).

[11] Xie, J., & Zhi, Y. A review of forest fire risk prediction models and methods using machine learning. Journal of Environmental Engineering. (2020).

[12] Oliveira, F., & Nogueira, M. Fire prediction and simulation using AI techniques: Applications in forest management. Journal of Artificial Intelligence Reserch. (2021).

[13] United Nations Environment Programme (UNEP). Global Forest Resources Assessment: Forest Fires and Deforestation Data. United Nations Environment Programme. (2020)