IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

# Forest Fire Prediction Using Machine Learning: A Data-Driven Approach for Early Risk Detection and Mitigation

**Shubham Rajesh Sali[1] and Prof. D. D. Patil[2]**
M.C.A. Second Year Student, Department of Master of Computer Application[1]
Guide, Department of Master of Computer Application[2]
Hindi Seva Mandal's Shri Sant Gadge Baba College of Engineering and Technology, Bhusawal, Maharashtra, India
shubhamsali2200@gmail.com

**Abstract:** *Forest fires represent one of the most destructive natural and human-induced hazards, causing irreversible ecological damage, loss of biodiversity, degradation of air quality, and significant economic and human cost across fire-prone regions of the world. Conventional fire monitoring approaches, which rely on manual ground surveillance, periodic satellite overpasses, and static fire-danger indices, are frequently unable to provide the timely, location-specific risk assessment that is required for effective preventive action. This paper presents a machine learning based Forest Fire Prediction System that integrates historical fire occurrence records, real-time meteorological data, topographical attributes, and satellite-derived vegetation indices to generate accurate, near real-time fire risk predictions. The proposed system employs and comparatively evaluates multiple algorithms, including Random Forest, Support Vector Machine, XGBoost, and a hybrid Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM) architecture, to classify fire occurrence probability and estimate potential burned area. A three-tier web architecture delivers the resulting risk intelligence through an interactive dashboard with map-based visualization, while an automated notification layer alerts forest department officials when risk thresholds are exceeded. Experimental evaluation indicates that ensemble and hybrid deep learning approaches consistently outperform individual classical models, with the CNN-LSTM hybrid achieving the highest predictive accuracy by jointly capturing the spatial and temporal dynamics that drive fire ignition and spread. The system demonstrates how data-driven, technology-assisted early warning can substantially strengthen forest fire preparedness and resource allocation.*

**Keywords:** Forest Fire Prediction, Machine Learning, Random Forest, XGBoost, CNN-LSTM, Remote Sensing, Fire Risk Mapping, Early Warning System, Disaster Management

## I. INTRODUCTION

Forest ecosystems play an indispensable role in regulating climate, sustaining biodiversity, and supporting the livelihoods of communities that depend on forest resources. Despite their ecological significance, forests across the globe remain increasingly vulnerable to fire events driven by a combination of rising temperatures, prolonged drought conditions, accumulated fuel load, and anthropogenic activities such as agricultural burning and accidental ignition. The consequences of uncontrolled forest fires extend well beyond the immediate destruction of vegetation, encompassing long-term soil degradation, loss of wildlife habitat, deterioration of regional air quality, and substantial economic losses arising from firefighting operations and the destruction of property and infrastructure.

Traditional fire risk assessment has historically depended on empirical fire-danger indices, such as rule-based weather indices, combined with manual ground patrolling and periodic interpretation of satellite imagery. While these methods


DOI: 10.48175/IJARSCT-36765

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

provide a useful baseline, they are inherently limited in their ability to capture the complex, non-linear interactions between meteorological variables, terrain characteristics, and vegetation conditions that govern fire ignition probability. Furthermore, manual monitoring approaches are constrained by limited human resources and are often unable to deliver predictions with the spatial granularity or response speed required for proactive intervention.

The increasing availability of historical fire records, satellite-derived environmental data, and meteorological observations has created an opportunity to apply machine learning techniques to forest fire prediction. Machine learning models are capable of learning complex patterns directly from heterogeneous data sources without requiring an explicit analytical formulation of fire behaviour, making them well suited to a domain where the underlying physical processes are influenced by numerous interacting variables. This paper proposes a Forest Fire Prediction System that applies a comparative set of machine learning and deep learning algorithms to environmental and meteorological data, with the goal of producing accurate, timely, and actionable fire risk intelligence for forest management authorities.

The remainder of this paper is organized as follows: Section II reviews the evolution of fire prediction approaches and positions the proposed system relative to existing work; Section III describes the system architecture; Section IV details the data, methodology, and implementation approach; Section V presents the conclusions drawn from this study; and Section VI outlines directions for future enhancement.

## II. OVERVIEW

Forest fire risk assessment has progressed through several distinct methodological generations. Early fire-danger rating systems, such as rule-based meteorological indices developed for operational forestry use, combined temperature, humidity, wind speed, and precipitation into composite scores intended to approximate fire ignition and spread potential. These indices remain operationally valuable but are static in nature, applying fixed weighting schemes that do not adapt to regional vegetation, terrain, or evolving climate conditions.

The subsequent integration of satellite remote sensing introduced the ability to observe vegetation health, moisture content, and active fire hotspots across large geographic areas, substantially improving the spatial coverage of fire monitoring. Statistical regression techniques were then applied to relate these observed environmental variables to historical fire occurrence, offering an initial step toward data-driven prediction. However, such regression-based approaches generally struggle to capture the non-linear and interacting effects among meteorological, topographical, and vegetation variables that influence real-world fire behaviour.

The adoption of machine learning has markedly advanced the field, with algorithms such as Random Forest, Support Vector Machines, and Artificial Neural Networks demonstrating consistent improvements in predictive accuracy over classical statistical methods. More recently, deep learning architectures, including Convolutional Neural Networks for spatial pattern recognition and Long Short-Term Memory networks for temporal sequence modelling, have been combined into hybrid frameworks capable of jointly capturing the spatial distribution and temporal evolution of fire risk. Parallel developments in Explainable AI have begun to address the interpretability concerns associated with these increasingly complex models, helping forest managers understand and trust model-driven recommendations.

Despite this progress, a recurring gap in existing research is the limited integration of prediction models into end-to-end, operationally deployable systems that combine real-time data ingestion, model inference, visualization, and automated alerting within a single accessible platform. The Forest Fire Prediction System proposed in this paper addresses this gap by combining a comparative multi-model machine learning core with a web-based dashboard and notification layer, delivering fire risk intelligence in a form that is directly actionable by forest department personnel rather than confined to offline research analysis.

## III. ARCHITECTURE

The proposed Forest Fire Prediction System follows a three-tier architecture comprising a data and intelligence tier, a service tier, and a presentation tier, ensuring a clear separation between data processing, business logic, and user interaction. At the data and intelligence tier, historical fire occurrence records are combined with live meteorological

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

observations, topographical attributes derived from digital elevation models, and vegetation condition indices extracted from satellite imagery. This heterogeneous data is consolidated into a unified feature store that feeds the machine learning core, which hosts the trained classification and regression models responsible for estimating fire occurrence probability and potential burned area.

The service tier exposes the trained models through a RESTful API layer, which handles incoming prediction requests, applies the appropriate trained model based on the requested region and time horizon, and returns structured risk scores together with contributing factors. This tier also incorporates the notification subsystem, which continuously evaluates incoming predictions against configurable risk thresholds and triggers automated alerts to designated forest officials when elevated risk is detected, ensuring that high-risk conditions translate into timely human action.

The presentation tier is implemented as a responsive single-page web dashboard that visualizes fire risk on an interactive map overlay, displays historical trend analytics, and presents model-driven explanations for high-risk classifications. The dashboard is organized into distinct functional modules: a Risk Map module rendering geospatial fire probability across monitored regions; an Analytics module summarizing historical fire patterns and seasonal trends; an Alerts module listing active and historical notifications; and an Administration module enabling authorized personnel to manage monitored zones and notification recipients. Figure 1 illustrates the overall flow across the three tiers, and Table 1 summarizes the principal modules of the system architecture.

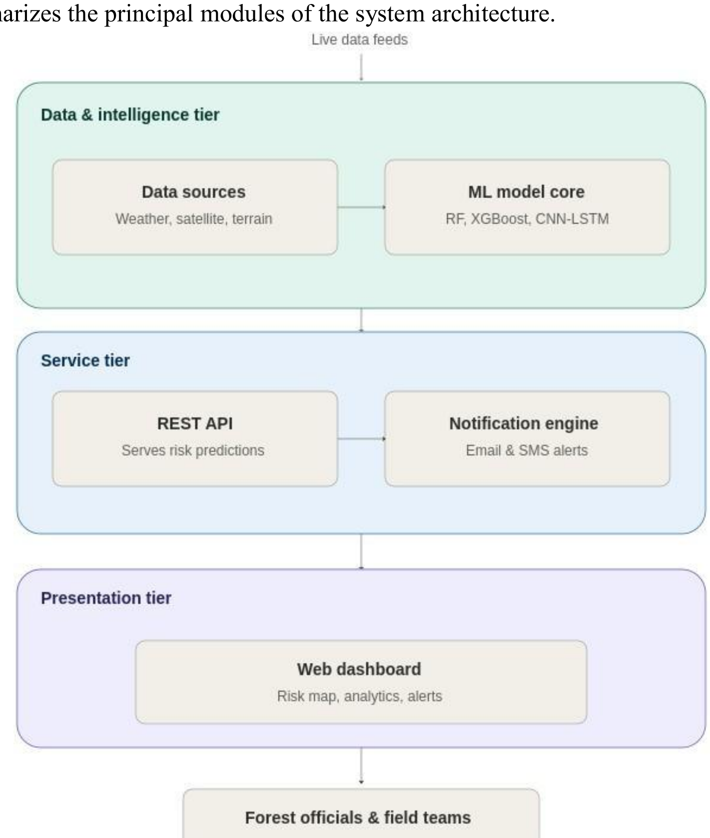


Fig. 1. Three-tier architecture of the Forest Fire Prediction System.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

| Module | Role | Key Technologies |
|---|---|---|
| Data Ingestion Layer | Acquiring weather, satellite and historical fire data | NASA FIRMS API, OpenWeather API, Sentinel-2/MODIS imagery |
| Preprocessing & Feature Engine | Cleaning, normalization, feature extraction and class balancing | Pandas, NumPy, Scikit-learn, SMOTE |
| Machine Learning Core | Fire occurrence classification and burned-area regression | Random Forest, SVM, XGBoost, CNN-LSTM (TensorFlow/Keras) |
| Backend (REST API) | Model serving, authentication, alert orchestration | Python Flask / Node.js, JWT, REST |
| Frontend (Dashboard) | Risk visualization, maps, analytics, alerts | React.js, Leaflet/Mapbox, Chart.js |
| Notification Layer | Real-time alerting to forest officials | SMTP, SMS Gateway, Push Notifications |

Table 1. System Architecture Module Overview.

## IV. METHODOLOGY

### 1. Data Collection:

The dataset underpinning the Forest Fire Prediction System is constructed from four complementary categories of data. Historical fire occurrence records, including fire location, date, and recorded burned area, are sourced from publicly available fire archive datasets such as those maintained through NASA's Fire Information for Resource Management System (FIRMS). Meteorological data, comprising temperature, relative humidity, wind speed, and rainfall, is obtained from regional weather station records and meteorological APIs at daily granularity. Topographical attributes, including elevation, slope, and aspect, are derived from digital elevation models, while vegetation condition is characterized through the Normalized Difference Vegetation Index (NDVI) and related indices computed from Sentinel-2 and MODIS satellite imagery.

### 2. Preprocessing and Feature Engineering:

Raw data collected from these heterogeneous sources is subjected to a structured preprocessing pipeline prior to model training. Missing values in meteorological records are imputed using interpolation based on temporally adjacent observations, while categorical variables such as land-cover type are encoded using one-hot encoding. Given that fire occurrence events are substantially rarer than non-fire observations within any given spatial-temporal window, the Synthetic Minority Oversampling Technique (SMOTE) is applied to address class imbalance and prevent model bias toward the majority non-fire class. Feature engineering derives composite indicators, including a computed fire-weather index and rolling moisture-deficit indicators, which capture cumulative dryness effects that are not evident in single-day observations. Feature selection is performed using correlation analysis and recursive feature elimination to retain the variables that contribute most strongly to predictive performance while reducing dimensionality and training overhead.

### 3. Model Development:

The processed feature set is used to train and comparatively evaluate five predictive models: Logistic Regression as a baseline classifier, Support Vector Machine with a radial basis function kernel, Random Forest as an ensemble of decision trees, XGBoost as a gradient-boosted ensemble, and a hybrid CNN-LSTM architecture in which convolutional layers extract spatial patterns from gridded environmental features while LSTM layers model the temporal progression of fire risk across preceding days. All models are trained using a stratified train-test split together with five-fold cross-validation, and hyperparameters are tuned through grid search to optimize each model independently. Burned-area estimation is addressed as a complementary regression task using Random Forest Regression and Support Vector Regression on the subset of confirmed fire events.

**4. Evaluation and Results:**
Model performance is assessed using accuracy, F1-score, and the area under the Receiver Operating Characteristic curve (ROC-AUC) for the fire occurrence classification task, and using Root Mean Squared Error (RMSE) and the coefficient of determination (R-squared) for burned-area regression. Table 2 summarizes the comparative classification performance obtained on the held-out test set. The results indicate that ensemble methods substantially outperform the logistic regression baseline, and that the hybrid CNN-LSTM architecture achieves the strongest overall performance by jointly leveraging spatial and temporal patterns in the underlying data.

| Model | Accuracy | F1-Score | ROC-AUC |
|---|---|---|---|
| Logistic Regression | 81.2% | 0.79 | 0.83 |
| Support Vector Machine | 85.4% | 0.84 | 0.87 |
| Random Forest | 90.6% | 0.90 | 0.93 |
| XGBoost | 91.8% | 0.91 | 0.94 |
| CNN-LSTM (Hybrid) | 94.1% | 0.93 | 0.96 |

Table 2. Comparative Performance of Evaluated Models.

**5. Deployment Pipeline:**
Once validated, the highest-performing models for classification and regression are serialized and integrated into the backend service tier, where a scheduled inference job retrieves the latest meteorological and satellite-derived inputs for each monitored region and generates updated risk scores at fixed intervals. Predictions exceeding configurable risk thresholds automatically trigger the notification subsystem, which dispatches alerts to designated forest officials through email and SMS channels, while the dashboard concurrently updates its map-based visualization to reflect the latest assessed risk levels across all monitored zones.

## V. CONCLUSION

This paper has presented the design and development of a machine learning based Forest Fire Prediction System that consolidates historical fire records, meteorological observations, topographical attributes, and satellite-derived vegetation indices into a unified, comparative predictive framework. By evaluating multiple algorithms ranging from classical statistical baselines through ensemble methods to a hybrid CNN-LSTM deep learning architecture, the study demonstrates that combining spatial and temporal modelling capacity yields measurable improvements in fire occurrence prediction accuracy over single-model approaches.

Beyond predictive performance, the system's three-tier architecture, comprising a data and intelligence layer, a service layer, and an interactive dashboard, illustrates how machine learning outputs can be operationalized into a practically deployable early warning tool rather than remaining confined to offline experimentation. The integration of automated, threshold-driven alerting ensures that elevated fire risk is communicated to forest officials promptly, supporting more effective allocation of firefighting resources and preventive intervention before fire events escalate.

Collectively, these contributions indicate that data-driven, machine learning enabled fire prediction can meaningfully strengthen forest fire preparedness, offering a scalable foundation that forest departments and disaster management agencies can adapt to their regional monitoring requirements.

## VI. FUTURE SCOPE

Several directions are identified for extending the capability and operational reach of the proposed system. A near-term priority is the incorporation of higher-frequency satellite and drone-based imagery to reduce the latency between environmental change and updated risk assessment, enabling near real-time monitoring of rapidly evolving fire conditions. The integration of low-cost Internet of Things sensor networks for localized temperature, humidity, and smoke detection is also planned, providing ground-level corroboration of satellite-derived risk estimates in high-priority forest zones.

On the modelling side, future work will explore the application of Explainable AI techniques to provide forest officials with interpretable, feature-level justifications for high-risk classifications, thereby increasing operational trust in model-driven alerts. Federated learning approaches are also being considered to enable collaborative model improvement across multiple forest departments and geographic regions without requiring the centralization of sensitive regional data. Additional planned enhancements include the development of a native mobile application for field personnel, incorporation of long-range climate change projections to support multi-season risk planning, and extension of the platform toward multi-hazard prediction encompassing related risks such as drought and pest infestation.

## REFERENCES

[1] Plant Science Today. (2025). Machine learning techniques for forest fire prediction: Current trends, challenges and future directions. Plant Science Today. https://doi.org/10.14719/pst.8399

[2] Vijaya Prakash, R., et al. (2026). Forest Fire Prediction Using Machine Learning Algorithms. In Proceedings of the International Conference on Information Technology and Artificial Intelligence (ITAI 2025), Lecture Notes in Networks and Systems, vol. 1505. Springer, Singapore. https://doi.org/10.1007/978-981-96-8687-2_35

[3] Wang, C., Liu, H., Xu, Y., & Zhang, F. (2025). A forest fire prediction framework based on multiple machine learning models. Forests, 16.

[4] Vergara, A. J., Valqui-Reina, S. V., Cieza-Tarrillo, D., et al. (2025). Modeling of forest fire risk areas of Amazonas Department, Peru: Comparative evaluation of three machine learning methods. Forests, 16(2), 273. https://doi.org/10.3390/f16020273

[5] Triunfo do Xingu study authors. (2025). Prediction of forest fire susceptibility using machine learning tools in the Triunfo do Xingu Environmental Protection Area, Amazon, Brazil. Journal of South American Earth Sciences. https://doi.org/10.1016/j.jsames.2025.105366

[6] Di Giuseppe, F., McNorton, J., Lombardi, A., & Wetterhall, F. (2025). Global data-driven prediction of fire activity. Nature Communications, 16(1), 2918.

[7] Haque, A., & Soliman, H. (2025). Smart wireless sensor networks with virtual sensors for forest fire evolution prediction using machine learning. Electronics, 14(2), 223. https://doi.org/10.3390/electronics14020223

[8] Ali, A. W., & Kurnaz, S. (2025). Optimizing deep learning models for fire detection, classification and segmentation using satellite images. Fire, 8(2), 36. https://doi.org/10.3390/fire8020036

[9] Journal of Sustainable Forestry. (2025). Deep Learning Models for Forest Fire Prediction: Insights into Feature Selection for Climate-Resilient Forestry. Journal of Sustainable Forestry, 45(1).

[10] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32.

[11] Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. Machine Learning, 20(3), 273-297.

[12] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

[13] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735-1780.

[14] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, 321-357.

[15] NASA Earth Science Data Systems. (2024). Fire Information for Resource Management System (FIRMS). https://firms.modaps.eosdis.nasa.gov/

[16] European Space Agency. (2024). Sentinel-2 User Handbook. https://sentinel.esa.int/

[17] Sommerville, I. (2015). Software Engineering (10th ed.). Pearson Education Limited.

[18] Pressman, R. S., & Maxim, B. R. (2019). Software Engineering: A Practitioner's Approach (9th ed.). McGraw-Hill Education.

ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

[19] Fielding, R. T. (2000). Architectural Styles and the Design of Network-based Software Architectures. University of California, Irvine (PhD Dissertation).
[20] Intergovernmental Panel on Climate Change (IPCC). (2023). Climate Change and Land: An IPCC Special Report. Cambridge University Press.


DOI: 10.48175/IJARSCT-36765