IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2025

DOI: 10.17148/IJIREEICE.2025.131113

Forest Fire Severity Prediction using Random Forest and Neural Network Stacking with SMOTE

Mani Shankar B[1], Jai Vignesh K[2], Arjun Ashkar N C[3], Dr G. Paavai Anand[4]

Student, Department of CSE, SRM Institute of Science and Technology, Chennai, India[1]

Student, Department of CSE, SRM Institute of Science and Technology, Chennai, India [2]

Student, Department of CSE, SRM Institute of Science and Technology, Chennai, India [3]

Assistant Professor (Sr. G), Department of CSE, SRM Institute of Science and Technology, Chennai, India [4]

Abstract: Forest fire prediction is essential for sustainable forest management, as wildfires cause significant ecological, environmental, and economic damage. This paper introduces a hybrid stacking model that improves the precision and resilience of forest fire risk prediction by merging Random Forest and Neural Network classifiers with Logistic Regression acting as a meta-learner. To solve the problem of class imbalance, new data samples were introduced and the SMOTE was used to add more high-risk fire events to the dataset, which was originally sourced from Kaggle. Additionally, feature engineering was used to create new variables that captured intricate connections between environmental and meteorological aspects. The suggested stacking framework achieves an accuracy of 87.2% on tests and an Area Under the Curve of 0.925 by combining the interpretability of RF with the nonlinear learning strength of NN. Strong dependability in differentiating between low-risk and high-risk fire incidents is demonstrated by these data. Overall, this approach provides an effective, data-driven foundation for intelligent wildfire monitoring and proactive forest management.

Keywords: Forest Fire Prediction, Random Forest, Neural Network, SMOTE, Machine Learning, Stacking, Feature Engineering.

I. INTRODUCTION

A) Impact of Forest Fires

Forest fires represent one of the most severe natural hazards, resulting in profound ecological, environmental, and economic consequences. They contribute to large-scale deforestation, degradation of ecosystems, and loss of biodiversity, while also releasing significant amounts of greenhouse gases that exacerbate global climate change. In addition to environmental damage, forest fires cause extensive property destruction, threaten human lives, and impose substantial costs on local economies and governments. Consequently, accurate and timely prediction of forest fire occurrences has become an essential aspect of sustainable forest management and disaster mitigation.

B) Role of Machine Learning in Fire Prediction

Machine learning (ML) methods have become effective instruments for forecasting the risk of forest fires in recent years. Data-driven decision-making in fire prevention and management is made possible by these models' ability to efficiently learn intricate, nonlinear correlations between meteorological, topographical, and environmental factors. “Artificial Neural Networks (ANN)” [3] and “Random Forest (RF)” [2] are two algorithms that have demonstrated great promise in predicting the behavior of forest fires and pinpointing high-risk areas.

C) Challenge of Class Imbalance in Fire Data

The problem of class imbalance, in which there are much more low-risk (non-fire) cases than high-risk (fire) cases, is one of the main obstacles encountered in such predictive modeling projects. This disparity frequently results in biased model training, which makes the classifier perform poorly on the minority class—which is essential for precise fire detection—while favoring the majority class. The current study used a two-stage data balancing technique to overcome this problem. To increase data diversity and enhance the minority class, more high-risk fire samples from validated environmental sources were included to the dataset in the first stage. A more balanced class distribution was ensured in the second stage by creating synthetic high-risk samples by interpolation using the “Synthetic Minority Over-sampling Technique (SMOTE)” [1]. This method improved the dataset's representativeness and made it possible for the model to successfully identify the traits that distinguished fire cases apart from non-fire ones.

DOI: 10.17148/IJIREEICE.2025.131113

D) Proposed Hybrid Stacking Model

To improve overall predictive performance, this research introduces a hybrid stacking model that combines the strengths of "Random Forest (RF)" [2], "Neural Network (NN)" [3], and "Logistic Regression (LR)" [4]. The RF model captures nonlinear dependencies and effectively handles noisy or high-dimensional features, while the NN model provides deep learning capabilities to learn complex feature interactions.

The outputs from these base learners are then integrated using Logistic Regression as a meta-learner, which refines the final predictions by learning optimal decision boundaries from the combined model outputs. This ensemble framework enhances generalization, robustness, and predictive reliability, outperforming individual models through complementary learning behavior.

II. RELATED WORK

A) Goal of the Research

The primary aim of this research is to develop an effective and understandable framework for forest fire risk prediction in order to address the problem of class imbalance while maintaining high accuracy and stability. The proposed model demonstrates its potential for application in early warning systems and forest fire management programs by displaying improved evaluation metrics.

B) Early Approaches and Statistical Models

Predicting forest fires with machine learning and data-driven methods has been the subject of extensive research. Among the first attempts to predict fire behavior were empirical models and traditional statistical approaches, but they frequently failed to capture the intricate and nonlinear interactions between environmental, meteorological, and geographic factors impacting fire ignition and spread. Machine learning (ML) algorithms have been used more and more in recent research to get around these restrictions because of their exceptional capacity to extract complex patterns from data and manage multiple inputs.

C) Classical Machine Learning Techniques for Fire Prediction

For predicting forest fires, multiple traditional machine learning methods have been thoroughly studied, including "Random Forests (RF), Support Vector Machines (SVM), and Decision Trees (DT)" [2]. Although decision trees are interpretable and effectively manage both continuous and categorical variables, they are prone to overfitting when working with data that is extremely variable. Support Vector Machines need careful kernel selection and parameter adjustment, but they perform well in high-dimensional feature spaces. Because of their robustness, decreased propensity for overfitting, and capacity to estimate feature importance, Random Forests - an ensemble-based model have proven to perform better in a variety of environmental prediction tasks. According to a number of studies, RF models are capable of accurately identifying areas that are prone to fire and calculating the probability of fires based on topographical factors, vegetation types, and weather.

D) Deep Learning Approaches for Wildfire Modeling

Deep learning (DL) methods have gained popularity over traditional machine learning (ML) methods for the spatiotemporal modeling of wildfire outbreaks. CNNs have been used to extract spatial features from satellite observations, remote sensing images, and vegetation indices, while RNNs and LSTM networks have been used to model temporal dependencies in fire-related time series data. These models can discover intricate spatial and temporal relationships that conventional machine learning methods would overlook. However, their effectiveness is sometimes restricted by the need for large, labeled datasets, which are hard to get in environmental research due to missing records, high data collection costs, and limited data availability. Furthermore, deep learning models are computationally intensive and may lack interpretability, which restricts their applicability in certain real-world decision-making contexts.

E) Addressing Class Imbalance with SMOTE

Another major challenge in forest fire prediction is the class imbalance problem, where there are much more low-risk instances than high-risk fire episodes. Unbalanced datasets may cause classifiers to be biased toward the majority class, which could lead to insufficient detection of minority (fire) cases. To lessen this issue, resampling techniques-particularly the "Synthetic Minority Over-sampling Technique (SMOTE)" [1] - have been widely employed. SMOTE balances the class distribution and improves model performance by generating synthetic samples of the minority class through interpolation between existing instances. Numerous studies have demonstrated the effectiveness of SMOTE and its modifications in improving the expected accuracy of machine learning models based on ecological and environmental information.

F) Ensemble and Stacking-Based Learning for Improved Prediction

Researchers are increasingly using ensemble learning techniques to enhance generalization and prediction performance. By combining several base models, strategies like bagging, boosting, and stacking [4] use each model's unique advantages while lowering bias and variation. By combining several learners, including neural networks and tree-based models, into a meta-learning framework, stacking ensembles have proven to be the most flexible and resilient of these.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2025

DOI: 10.17148/IJIREEICE.2025.131113

This method enhances classification accuracy and stability by enabling the ensemble to identify both linear and nonlinear dependencies in the data.

G) Integration of RF and NN with SMOTE in the Proposed Framework

Building upon these insights, the present study integrates Random Forest (RF) and Neural Network (NN) classifiers within a stacking ensemble framework, incorporating Logistic Regression (LR) as the meta-learner. Additionally, SMOTE-based resampling is employed to address class imbalance and enhance model learning on minority (fire) instances. This combination leverages the interpretability and robustness of RF, the deep feature learning capacity of NN, and the decision optimization capability of LR, resulting in a balanced, generalizable, and reliable approach for forest fire risk prediction.

III. PROPOSED METHODOLOGY

The proposed methodology aims to provide an effective and easily comprehensible framework for forest fire risk prediction in order to handle the problem of class imbalance and ensure high accuracy and model stability. In order to handle data imbalance, the framework combines Random Forest (RF) and Neural Network (NN) classifiers in a stacking ensemble architecture with Logistic Regression (LR) and Synthetic Minority Over-sampling Technique (SMOTE). The methodical procedures for feature engineering, model creation, evaluation, and data pretreatment are described in this section.

A) Data Preprocessing

The dataset used in this work contains numerical and categorical features that describe a wide range of environmental and climatic factors, such as temperature, humidity, wind speed, rainfall, and temporal variables like day and month. To ensure compatibility with machine learning methods, label encoding was utilized to convert categorical data (such as day and month) into numerical form. Imputation techniques were used to manage any missing values in order to maintain the data's completeness and integrity.

B) Feature Scaling and Engineering

By fitting each feature to a similar distribution, StandardScaler was used to standardize the data and make sure all numerical variables were on the same scale. This procedure increases training efficiency and speeds up convergence, especially for neural network models.

In addition to the existing parameters, several derived features were engineered to improve the model's interpretability and performance:

- ***Temperature-Humidity Ratio:*** Represents the combined effect of temperature and humidity on fire ignition likelihood.
- ***Wind-Moisture Effect:*** Captures the relationship between wind speed and relative moisture, two critical factors influencing fire spread.
- ***Fire Weather Index (FWI):*** A composite indicator that integrates meteorological variables to estimate overall fire danger levels.

Richer contextual information is provided by these extra features, which improves the model's ability to understand intricate environmental relationships.

C) Handling Class Imbalance

It was discovered that the minority (high-risk) and majority (low-risk) fire classes were out of proportion. This issue was resolved by applying the SMOTE. By creating artificial samples for the minority class by interpolating between preexisting examples, SMOTE balances the distribution of classes. This procedure reduces bias toward the majority class and improves the classifier's capacity to identify high-risk fire incidents.

D) Model Architecture

With Random Forest (RF) and Neural Network (NN) models acting as base learners and Logistic Regression acting as the meta-learner, the suggested framework uses a stacking ensemble technique.

- ***Random Forest (RF):*** This group of decision trees ensures resilience and less overfitting by capturing non-linear interactions and estimating feature importance through aggregated judgments. ***Neural Network (NN):*** The Adam optimizer is used to optimize the NN model, which consists of fully connected layers with Rectified Linear Unit (ReLU) activation functions. It captures deeper representations and effectively represent complex dependencies in the dataset.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2025

DOI: 10.17148/IJIREEICE.2025.131113

The probabilistic outputs from both base models are input to the Logistic Regression meta-model, which learns an optimal decision boundary that enhances overall classification accuracy. This hierarchical learning mechanism leverages RF's interpretability and NN's learning depth, resulting in improved generalization on unseen data.

E) Model Evaluation

Area, F1-Score, Accuracy, Recall and Precision. The performance of the proposed model was assessed using typical classification measures that fall under the ROC Curve (AUC). Together, these measures evaluate the model's ability to forecast majority and minority classes. Cross-validation was also used to guarantee uniformity and reduce overfitting, offering a trustworthy indicator of the model's resilience and capacity for prediction across various data subsets.

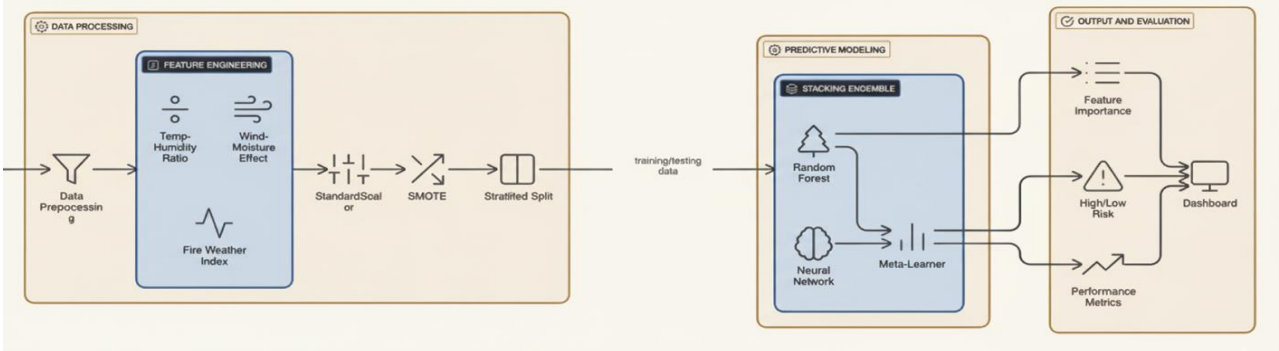


Fig. 1 Architecture Diagram

IV. EXPERIMANETAL SETUP AND RESULT

The Forest Fires dataset, which was meticulously preprocessed, feature-engineered, and balanced using the SMOTE to address the issue of class imbalance, was used for the experimental evaluation. To guarantee fair model evaluation and gauge the suggested framework's capacity for generalization, the dataset was split into training (70%) and testing (30%) subsets.

The Random Forest (RF) classifier was configured with 200 estimators to capture nonlinear interactions and maximize its resilience. The Neural Network (NN) component consisted of two hidden layers with 32 and 16 neurons, respectively, along with Rectified Linear Unit (ReLU) activation functions and the Adam optimizer for efficient convergence. The probabilistic outputs of RF and NN models were combined in the hybrid stacking framework, and Logistic Regression was used as the meta-learner to improve classification performance and final predictions.

Standard measures, including Recall, F1-score, Precision, Accuracy, and Area Under the Curve (AUC), were used to assess the model's performance in order to offer a thorough evaluation of predictive quality. Table 1 provides a summary of the comparison outcomes between the training and testing datasets.

TABLE 1: EVALUATION METRICS

Dataset	Accuracy	Precision	Recall	F1-Score
Train	0.992	0.991	0.994	0.992
Test	0.872	0.904	0.832	0.866

With a test accuracy of 87.2% and an F1-score of 0.866, the metrics show that the suggested hybrid stacking model performed well in generalization, demonstrating a well-balanced trade-off between precision and recall. Robust predictive stability and little overfitting are demonstrated by the close alignment of training and testing metrics.

- The Confusion Matrix (Figure 2) shows that there were few false positives and false negatives and that most high-risk and low-risk cases were accurately identified.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2025

DOI: 10.17148/IJIREEICE.2025.131113

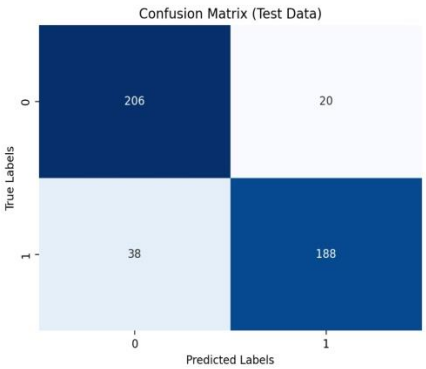
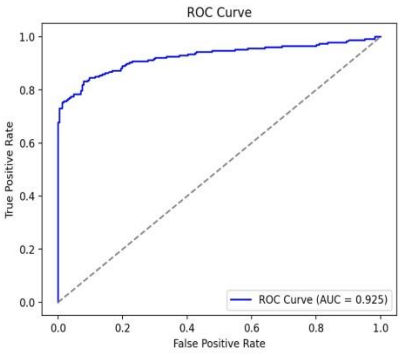


Fig 2. Confusion Matrix

- The high AUC value of the ROC Curve (Figure 3) confirms the model's good discriminative capacity to distinguish between high-risk and low-risk fire incidents.



s

Fig 3. ROC Curve

Overall, the experimental results demonstrate how well the proposed hybrid stacking technique enhances forest fire risk prediction through balanced learning, enhanced generalization, and reliable decision-making.

V. DISCUSSION

A) Model Performance and Effectiveness

For estimating the danger of forest fires, the suggested hybrid stacking model that combines Random Forest (RF) and Neural Network (NN) classifiers turned out to be a dependable and effective method. The ensemble outperformed the individual classifiers by utilizing the complementing strengths of both models, namely NN's capacity to model intricate nonlinear interactions and RF's capacity to capture feature-level dependencies. The decision fusion procedure was further improved by adding a Logistic Regression meta-learner, which increased overall classification accuracy and model stability.

B) Addressing Class Imbalance

The SMOTE was used to successfully correct for class imbalance, which is common in environmental datasets. By balancing the minority and majority classes, the model showed a significant improvement in memory for high-risk fire events. This improvement is crucial for real-world applications since it can help with timely responses and preventative measures by identifying rare but serious fire events.

C) Experimental Evaluation

With an F1-score of 0.866, an Area Under the Curve value of 0.925, and a test accuracy of 87.2%, experimental study showed the hybrid model's good predictive performance. These measures show that the suggested framework is resilient against overfitting and has good generalization. The outcomes validate the stacking approach's capacity to efficiently integrate several learning algorithms for improved prediction stability and accuracy.

DOI: 10.17148/IJIREEICE.2025.131113

D) Feature Importance Analysis

Rainfall, the Initial Spread Index (ISI), and the Fine Fuel Moisture Code (FFMC) were found to be among the most important predictors of forest fire risk using feature importance analysis. These variables demonstrate how important fuel-related and climatic factors are in determining fire vulnerability. Comprehending their respective contributions facilitates the creation of more focused fire prevention tactics and enhances interpretability.

E) Future Enhancements

Despite its good performance, the suggested model's predictive ability should be enhanced. A more thorough grasp of the environmental elements affecting fire behavior would be possible by include further variables like vegetation indices, soil moisture, and elevation data. Additionally, combining temporal sequence modeling, remote sensing images, and optimization-based feature selection methods may improve computational efficiency and spatial-temporal accuracy.

F) Overall Implications

This study demonstrates that a potent framework for forest fire prediction is created by combining ensemble learning, feature engineering, and data balancing. The proposed stacking strategy is a practical, scalable, and data-driven approach to real-time forest fire risk assessment. Early warning systems, proactive catastrophe mitigation initiatives, and sustainable forest management can all benefit greatly from its deployment.

VI. CONCLUSION

This study uses a hybrid stacking model that combines “Random Forest (RF), Neural Network (NN), and Logistic Regression” [4] classifiers to propose an efficient method for forest fire risk prediction. To improve overall decision fusion and model stability, the combination of RF and NN makes use of their complimentary strengths: NN can model complicated nonlinear patterns, while RF can capture feature-level correlations. Logistic Regression acts as the meta-learner.

The “Synthetic Minority Over-sampling Technique (SMOTE)” [1], which enhanced the model's sensitivity to high-risk fire scenarios, was successfully used to address class imbalance. In line with established environmental findings, feature importance analysis revealed that the Initial Spread Index (ISI), rainfall, and the Fine Fuel Moisture Code (FFMC) were significant contributors in fire incidence.

Overall, the integration of SMOTE-based data balancing, feature engineering, and ensemble learning through Logistic Regression provides a reliable and scalable framework for forest fire prediction, contributing to proactive wildfire management and sustainable forest conservation.

REFERENCES

[1]. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” *J. Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[2]. L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[3]. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning Representations by Back-Propagating Errors,” *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[4]. D. H. Wolpert, “Stacked Generalization,” *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992.