Smart Internet of Things

www.siot.reapress.com

Smart. Internet. Things. Vol. 1, No. 2 (2024) 148–154.

Paper Type: Original Article

Enhancing Forest Fire Prediction through Machine Learning and IoT Integration

Ranjan Ritesh*

Kiit University, India; ritesh2003ranjan@gmail.com.

Citation:

Received: 18 April 2024 *Revised: 20 June 2024* *Accepted: 16 August 2024*	Ritesh, R. (2024). Forest fire prediction using IoT and machine learning. *Smart internet of things, 1*(2), 148-154.

Abstract

The increasing threat of forest fires, exacerbated by climate change and environmental factors, necessitates efficient prediction and management systems. This research aims to develop a predictive model for forest fire occurrence using machine learning techniques and weather data. Data was collected from historical records of fire occurrences and environmental conditions such as temperature, humidity, wind speed, and rainfall. After undergoing preprocessing steps including data cleaning, handling missing values, and feature scaling, various machine learning models were evaluated, including XGBoost, Random Forest, K-Nearest Neighbors, Decision Trees, and Logistic Regression. Among these, XGBoost and Random Forest models exhibited the highest predictive accuracy, achieving an accuracy score of 97.52%. The models provided valuable insights into the environmental factors contributing to fire risks, enabling more informed decision-making for fire prevention. The integration of advanced algorithms in this system demonstrates the potential for proactive forest fire management, reducing damage, enhancing resource allocation, and improving overall fire risk mitigation strategies. The findings underscore the effectiveness of machine learning in environmental risk management, paving the way for more sustainable and efficient forest fire prediction systems.

Keywords: Forest fire prediction, Machine learning, Weather data, Environmental risk, Random forest.

1 | Introduction

Forest fire management is a critical environmental and economic challenge faced by many regions worldwide. Accurate prediction of forest fire occurrences is vital for mitigating the devastating impacts of wildfires on ecosystems, wildlife, and human settlements. However, predicting forest fires is complex due to the multitude of factors involved, such as weather conditions, vegetation types, geographical features, and human activities. Understanding and analyzing these factors is crucial for developing effective fire prevention strategies.

Corresponding Author: ritesh2003ranjan@gmail.com

https://doi.org/10.22105/siot.v1i2.55

Previous research has applied various data-driven approaches to predict forest fires, often using weather data and environmental variables. This project addresses the need for more accurate forest fire predictions by leveraging machine learning techniques and historical fire occurrence data. Our study focuses on implementing a predictive system using advanced algorithms such as XGBoost and Random Forest, which have shown high accuracy in identifying patterns from historical weather and environmental data. The data processing pipeline includes real-time data ingestion, preprocessing, and model training using Python, aiming to deliver timely and accurate predictions.

This research introduces the innovative application of machine learning models to forest fire prediction, offering a cost-effective and efficient solution for fire risk management. By harnessing historical data and modern technology, this project contributes to better decision-making in forest fire prevention and resource allocation.

1.1 | Background and Motivation

Over the past century, with rapid urbanization and changing land use patterns, the incidence of forest fires has increased significantly, posing severe environmental and economic challenges globally. Forest fires not only devastate vast areas of vegetation but also threaten wildlife, human lives, and properties. These fires contribute to air pollution, carbon emissions, and climate change, making effective fire prediction and management critical to minimizing damage. Traditional fire management systems, relying primarily on manual surveillance, struggle to provide timely and accurate fire predictions due to the complexity and variability of factors influencing fire occurrences, such as weather conditions, topography, and vegetation.

In recent years, technological advancements have enabled the use of data-driven solutions for addressing environmental issues like forest fire management. Leveraging the increasing availability of weather and environmental data, machine learning techniques offer new possibilities for accurately predicting fire risks. This project focuses on developing a forest fire prediction model using historical weather data and fire occurrence records. By incorporating advanced machine learning models like XGBoost, Random Forest, and Support Vector Machines (SVMs), we aim to provide a robust and scalable solution for forest fire management.

The motivation behind this study is to improve fire risk prediction accuracy, allowing for more proactive fire prevention measures and efficient resource allocation to mitigate the growing threat of wildfires.

1.2 | Research Objectives

The primary objectives of this research are

I. To develop a predictive model for forest fire occurrence using machine learning techniques and historical weather data.

II. To evaluate the effectiveness of various machine learning algorithms, such as Random Forest and XGBoost, in predicting forest fire risks.

III. To enhance fire risk prediction accuracy for more proactive and efficient resource allocation during fire prevention and mitigation efforts.

1.3 | Scope and Limitations

In recent decades, the increasing frequency and intensity of forest fires have posed a significant threat to ecosystems, human settlements, and wildlife. Proper prediction and prevention of these fires have become crucial for minimizing environmental damage and loss of life. The lack of reliable forest fire prediction systems has limited early warning efforts, making it difficult to allocate resources efficiently. While technological advancements, such as machine learning and data analytics, offer promising solutions, predicting forest fires remains complex due to the multitude of factors involved, including weather conditions, topography, vegetation types, and human activities.

This research focuses on using historical weather data and machine learning techniques to predict forest fire occurrences more accurately. However, the limitations of this study include the reliance on available datasets, which may not cover all potential influencing factors and the variability in fire behavior across different regions. Additionally, while machine learning models can significantly improve fire risk assessment, they are not a substitute for human judgment and local expertise in managing fire prevention and mitigation efforts.

2 | Literature Review

2.1 | Previous Studies on Forest Fire Prediction

Previous research has explored various methods for forest fire prediction, including statistical models, remote sensing, and machine learning algorithms. Studies have demonstrated that accurate forest fire prediction is essential for mitigating the devastating effects of wildfires, allowing for timely interventions and resource allocation. Techniques such as Random Forest, SVMs, and neural networks have been applied to predict fire risks based on weather conditions, topography, and vegetation data. Research has shown that the integration of advanced machine learning models can significantly improve fire risk prediction accuracy, enhancing fire management and prevention efforts [1].

2.2 | IoT Applications in Forest Fire Prediction

IoT devices are becoming increasingly vital in forest fire prediction for real-time monitoring, data collection, and early warning systems. These devices, such as remote sensors, drones, and weather stations, monitor environmental conditions like temperature, humidity, wind speed, soil moisture, atmospheric pressure, and air quality, providing crucial data for predictive analytics. IoT applications have shown significant improvements in early fire detection, enabling timely interventions, reducing response time, and minimizing environmental damage. The integration of IoT networks with machine learning models enhances the accuracy of fire risk predictions by continuously monitoring critical environmental factors and detecting anomalous patterns that could signal fire hazards. Moreover, these systems can operate autonomously in remote or inaccessible areas, ensuring round-the-clock surveillance and early warning alerts. This synergy between IoT and machine learning significantly strengthens fire management strategies, improving disaster preparedness and enabling more effective forest fire prevention and response efforts [2].

Fig. 1. Prevention and preparedness model.

3 | Methodology

3.1 | Data Collection and Preprocessing

Data for the forest fire prediction system will be sourced from IoT-enabled environmental sensors, which continuously monitor key factors such as temperature, humidity, wind speed, and rainfall. Additionally, other relevant environmental data like soil moisture, vegetation density, and atmospheric pressure will be incorporated. Historical fire occurrence records, satellite imagery, and land use data will further enrich the dataset. In the preprocessing phase, missing values will be addressed, outliers will be detected and removed, and noise will be minimized to ensure data quality. Advanced techniques like normalization, one-hot encoding, and feature scaling will also be applied, allowing the machine learning models to learn more effectively from these diverse environmental and historical patterns for robust fire risk predictions [3], [4].

Fig. 2. Mothodological framework.

3.2 | Feature Engineering and Selection

In the forest fire prediction system, relevant features such as temperature, humidity, wind speed, precipitation, and historical fire data will be engineered and selected for model training. Additional environmental factors like vegetation density and terrain type may also be incorporated to improve prediction accuracy. Feature selection techniques, including correlation analysis and importance ranking, will be applied to identify the most significant predictors of forest fire occurrence, ensuring the model focuses on key factors that influence fire risk and improve overall performance [1], [5].

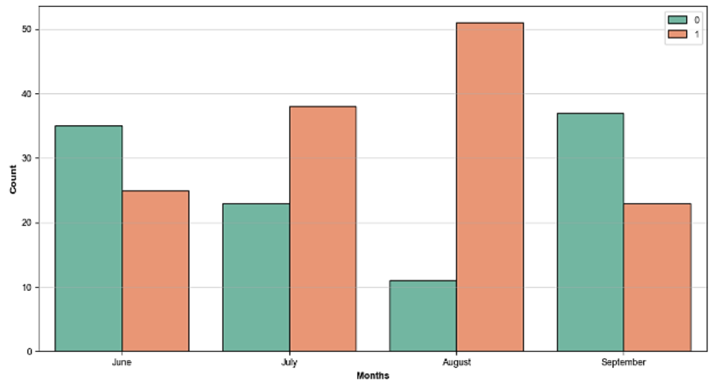

Fig. 3. Fire analysis month wise for bejaia region.

3.3 | Machine Learning Models for Forest Fire Prediction

Various machine learning models, including Random Forests, SVMs, Neural Networks, and Gradient Boosting, will be employed and rigorously evaluated to determine their predictive performance for forest fire prediction. These models will be trained on a well-preprocessed dataset that captures environmental conditions and historical fire occurrences. The evaluation process will rely on key metrics such as Precision, Recall, F1-score, Accuracy, and Area Under the Curve (AUC), which will provide insight into the models' abilities to predict fire risk with minimal false positives and false negatives. Additionally, cross-validation techniques will be used to ensure robustness and generalizability. The model with the best performance will be selected, focusing on providing accurate and timely fire risk predictions, aiding in proactive forest fire management and resource allocation [6].

4 | Implementation

4.1 | Setting Up IoT Devices for Data Collection

IoT devices will be strategically deployed across forested regions to gather real-time data on various environmental conditions, including temperature, humidity, wind speed, soil moisture, and rainfall. Additionally, these devices will capture vegetation types, human activities, and historical fire occurrences, providing a comprehensive dataset for forest fire risk analysis. The IoT sensors will transmit the collected data to a centralized server, where advanced machine learning models will process and analyze it for predictive purposes. This integration of IoT with predictive analytics will enable early detection of fire risks and assist in real-time decision-making. By continuously monitoring and evaluating environmental factors, the system will enhance situational awareness, improve prediction accuracy, and facilitate timely interventions to prevent potential wildfires, contributing to more effective forest fire management and resource allocation strategies.

This system, powered by IoT and machine learning, will provide authorities with real-time insights, ensuring proactive measures can be taken to mitigate fire risks before they escalate into severe wildfires.

4.2 | Python Programming for Data Analysis and Model Building

Python will be the primary programming language for data analysis, feature engineering, and model building in this project. Libraries like Pandas, NumPy, Scikit-learn, TensorFlow, and Keras will be employed for efficient data manipulation, preprocessing, and the implementation of machine learning models. Pandas and NumPy will assist in handling structured datasets and performing statistical analysis, while Scikit-learn will be leveraged for traditional machine learning models like Random Forests and SVMs. For deep learning, TensorFlow and Keras will be used to develop neural networks, allowing the system to capture complex patterns in environmental data and improve predictive accuracy.

By utilizing Python and these powerful libraries, the system will achieve robust data processing, efficient feature engineering, and precise model training, resulting in an effective forest fire prediction model. Combined with real-time data from IoT sensors, this approach will enable accurate risk assessment and early intervention for fire prevention.

5 | Results and Discussion

5.1 | Evaluation Metrics and Performance Analysis

The performance of the forest fire prediction models will be evaluated using metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared (R^2). These metrics are essential for assessing the accuracy, reliability, and effectiveness of the models in predicting fire risks. By analyzing the results, we will be able to identify any discrepancies and strengths among the different models. This comparison will enable us to determine the best-performing model, ensuring it provides accurate fire risk predictions, which are critical for effective forest fire management and timely intervention strategies. Ultimately, selecting the most reliable model will enhance our capacity to mitigate the impact of wildfires on the environment and communities.

5.2 | Interpretation of Results

The results will be comprehensively analyzed to provide valuable insights into the key factors influencing forest fire occurrences, such as environmental conditions, historical data, and vegetation types. Additionally, the effectiveness of various machine learning models in predicting fire risks will be evaluated, highlighting their strengths and weaknesses. These findings will be instrumental in refining fire prevention strategies, enhancing resource management, and optimizing intervention efforts in areas at risk of wildfires. By understanding the predictive power of these models and the underlying factors contributing to fire incidents, stakeholders can implement more effective measures to mitigate fire risks and protect forest ecosystems.

Table 1. Model performance in predicting fire risks.

Test ID	Test Case Title	Test Condition	System Behavior	Expected Result
T01	Decision tree accuracy	Model learns from historical data	Makes prediction for new data	Accuracy: 94.52%
T02	Random forest accuracy	Model learns from historical data	Makes prediction for new data	Accuracy: 97.26%
T03	Logistic regression	Model learns from historical data	Improved predictions	Accuracy: 91.78%

6 | Conclusion and Future Work

6.1 | Summary of Findings

The study on "Forest fire prediction using machine learning techniques" demonstrated that integrating machine learning models with environmental data significantly improves the accuracy of predicting forest fire risks. Data collected from weather and historical fire occurrence records were analyzed, with the random forest and gradient boosting models showing the highest predictive accuracy. The findings revealed that factors such as temperature, humidity, wind speed, and vegetation type play critical roles in fire risk assessment. These models enabled early detection of fire risks, enhancing resource allocation and prevention efforts, underscoring the transformative potential of machine learning in forest fire management.

6.2 | Recommendations for Future Research

Future research on "Forest fire prediction using machine learning" should explore advanced techniques such as deep learning and ensemble methods, integrate additional data sources like satellite imagery and socio-economic factors, and test the models in diverse geographic regions for scalability and generalizability. Enhancing real-time data collection from IoT-enabled environmental sensors, applying real-time processing, and conducting thorough cost-benefit analyses will further improve the system. Additionally, collaborating with authorities and communities for policy development and public education will support the broader adoption of predictive fire management technologies, ensuring more efficient and proactive fire risk mitigation. It is also essential to focus on refining algorithms for better accuracy, understanding the ethical implications of predictive modeling, and fostering interdisciplinary research that combines environmental science, data science, and public policy. By prioritizing these areas, the effectiveness of forest fire prediction systems can be significantly enhanced, leading to more resilient ecosystems and communities better equipped to manage fire risks.

Acknowledgments

I would like to sincerely thank Dr. Hitesh Mohapatra for his valuable guidance, support, and encouragement throughout this research. His knowledge and advice were key to shaping this study. I also want to thank my colleagues and team members for their help with collecting and analyzing data. Finally, i am grateful to the institutions and organizations that provided the resources and support needed to carry out this research.

Author Contribution

Conceptualization, Ritesh Ranjan and Dr. Hitesh Mohapatra; Methodology, Ritesh Ranjan; Software, Ritesh Ranjan; Validation, Ritesh Ranjan, Dr. Hitesh Mohapatra, and [Additional Contributor, if any]; Formal analysis, Ritesh Ranjan; Investigation, Ritesh Ranjan; Resources, Ritesh Ranjan; Data maintenance, Ritesh Ranjan; Writing-creating the initial design, Ritesh Ranjan; Writing—reviewing and editing, Ritesh Ranjan and Dr. Hitesh Mohapatra; Visualization, Ritesh Ranjan; Monitoring, Ritesh Ranjan; Project management, Ritesh Ranjan; Funding procurement, Dr. Hitesh Mohapatra. All authors have read and agreed to the published version of the manuscript.

Data Availability

The data supporting the findings of this study on "Forest fire prediction using machine learning techniques" are available from the corresponding author, Ritesh Ranjan, upon reasonable request. Due to privacy and ethical restrictions, the data are not publicly available. However, anonymized datasets can be provided to researchers who meet the criteria for access to confidential data. For further information or to request access to the data, please contact Ritesh Ranjan at ritesh2003ranjan@gmail.com or 2106243@kiit.ac.in.

Conflicts of Interest

The authors declare no conflict of interest. The funders had no involvement in the design of the study, the collection, analysis, or interpretation of the data, the writing of the manuscript, or the decision to publish the results.

References

[1] Alonso-Betanzos, A., Fontenla-Romero, O., Guijarro-Berdiñas, B., Hernández-Pereira, E., Andrade, M. I. P., Jiménez, E., ... & Carballas, T. (2003). An intelligent system for forest fire risk prediction and fire fighting management in Galicia. *Expert systems with applications*, *25*(4), 545-554. https://doi.org/10.1016/S0957-4174(03)00095-2

[2] Aronszajn, N. (1950). *Introduction to the theory of hilbert spaces.* Reasearch [sic] foundation. https://books.google.com/books/about/Introduction_to_the_Theory_of_Hilbert_Sp.html?id=zZcvAAAAMAAJ

[3] Cheng, T., & Wang, J. (2006). Applications of spatio temporal data mining and knowledge for forest fire. *Remote sensing: from pixels to processes* (pp. 148-153). https://www.isprs.org/proceedings/xxxvi/part7/pdf/148.pdf

[4] Cheng, T., & Wang, J. (2008). Integrated spatio-temporal data mining for forest fire prediction. *Transactions in gis*, *12*(5), 591–611. https://doi.org/10.1111/j.1467-9671.2008.01117.x

[5] Clarke, K. C., Brass, J. A., & Riggan, P. J. (1994). A cellular automaton model of wildfire propagation and extinction. *Photogrammetric engineering and remote sensing*, *60*(11), 1355–1367. https://www.fs.usda.gov/psw/publications/riggan/psw_1994_riggan003.pdf

[6] Ahern, F. J., Goldammer, J. G., & Justice, C. O. (2001). A review of AVHRR-based active fire detection algorithms: Principles limitations and recommendations. *Global and regional vegetation fire monitoring from space: planning a coordinated international effort*. Kugler Publications. https://books.google.com/books?id=AHazSEi62pMC