Integrating Remote Sensing and Machine Learning for Enhanced Wildfire Risk Assessment

Mohaddeseh Mesvari, Reza Shah-Hosseini *

School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran, (m.mesvari, rshahosseini)@ut.ac.ir

Keywords: Wildfire, Remote sensing, Machine learning models, Random Forest, Sentinel-2, Landsat-8.

Abstract

Forest fires pose a significant environmental threat, contributing to substantial ecological destruction and economic losses. With the accelerating impacts of climate change, including rising temperatures and prolonged droughts, the frequency and intensity of such fires are on the rise, raising urgent concerns for effective management and response strategies. This study employs advanced remote sensing techniques, specifically utilizing Sentinel-2 and Landsat-8 satellite imagery, to evaluate their efficacy in estimating and predicting wildfire risks. By integrating a diverse set of environmental variables—such as local meteorological conditions, vegetation indices, and topographic features—this research implements machine learning models, notably Random Forest and Extreme Gradient Boosting, to create a comprehensive wildfire risk assessment framework. Additionally, the importance of individual predictors in estimating fire risk, revealing that elevation, slope, aspect, and surface temperature substantially influence the models' predictions. The results indicate that the higher spatial resolution of Sentinel-2 data provides more accurate fire risk estimations than Landsat-8 imagery. This work lays the foundation for improved wildfire management strategies and highlights the integration of satellite data and machine learning as a powerful approach to supporting disaster preparedness and resource allocation in fire-prone regions.

1. Introduction

Forest fires are among the environmental hazards that cause extensive damage to natural resources, human infrastructure, biodiversity, and community health every year in different parts of the world, especially in areas with dense vegetation and hot and dry climate conditions (Müller et al., 2020). In recent years, with the intensification of climate change, increasing temperatures, decreasing humidity, and the spread of droughts, the occurrence and spread of fires has become an increasing trend, and this issue has become a serious concern in the field of natural resource management, environment, and food security (Xie et al., 2022). In such circumstances, accurate and timely identification and prediction of fires for the purpose of crisis management, resource planning, and reducing the damages caused by them becomes more important (Zhu et al., 2024).
One of the new and efficient approaches in this field is the use of remote sensing technology and satellite data, which have become a key tool in the analysis of environmental phenomena due to their ability to monitor regularly, wide coverage, free or low-cost access, and the ability to extract quantitative and qualitative information. Sentinel-2 satellite images belonging to the European Union's Copernicus program and Landsat-8 belonging to the United States Geological Survey are among the most widely used and reliable data sources in studies of vegetation cover, natural hazards, and especially forest fires. These two systems, with differences in spatial resolution, imaging time interval, and spectral characteristics, provide the possibility of effective comparative analysis in the field of evaluating their capabilities in detecting and predicting fires.
In addition to visual information from satellite images, the use of auxiliary data such as air temperature, wind speed and direction, relative humidity, and vegetation indices such as NDVI, EVI, as well as fire-specific indices such as NBR (Normalized Burn Ratio) and BAI (Burned Area Index) can significantly improve the accuracy and reliability of analyses. These data not only reveal the biological and climatic characteristics of the regions, but also provide a basis for examining the underlying conditions of fires, paving the way for building more realistic and predictive models.
On the other hand, recent advances in the field of artificial intelligence, especially machine learning and deep learning, have opened new horizons in modeling and predicting complex phenomena such as fires. Deep learning algorithms such as convolutional neural networks (CNN) and recurrent networks (RNN) are capable of extracting hidden patterns, understanding complex relationships between variables, and making accurate predictions based on large volumes of data. The integration of satellite images, auxiliary data, and deep learning algorithms provides a powerful platform for the development of early warning systems and risk assessment.
In this study, the main objective is to evaluate the effectiveness of Sentinel-2 and Landsat-8 satellite data in estimating and predicting forest fires, with an emphasis on comparing their spatial, temporal, and spectral characteristics. Deep learning algorithms are also used to analyze the data and develop an intelligent and accurate model for identifying and predicting fire-prone areas. The results of this research can be used to optimize management decisions, design early warning systems, reduce human and environmental damage, and improve the resilience of ecosystems against fire.

2. Study area and Data

2.1 Study area

The study area for this flood zoning analysis is located within the Santa Monica Mountains region, west of Los Angeles, California. As illustrated in Figure (1), the delineated zone covers a segment of the coastal and mountainous environment

* Corresponding author

https://doi.org/10.5194/isprs-annals-X-4-W8-2025-509-2026

extending from the Malibu coastal areas northward into the interior mountainous terrain.

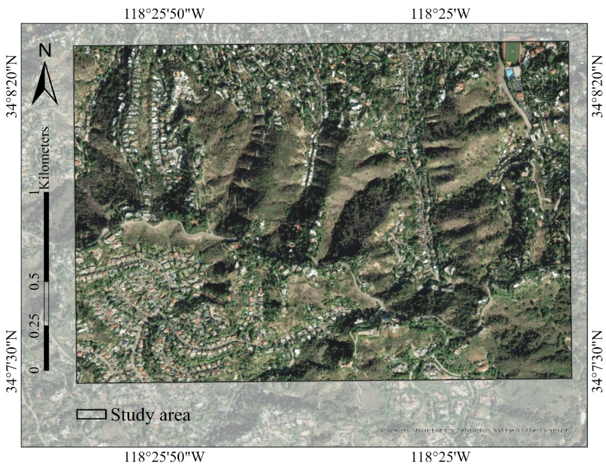


Figure 1. Study area

By integrating both static and dynamic environmental variables—ranging from vegetation condition and topography to short-term meteorological influences—the dataset enables a multifaceted analysis of fire hazards. Moreover, the use of Sentinel-2 and Landsat-8 allows for comparative analysis across different spatial resolutions, enhancing the flexibility and depth of the study. This comprehensive approach ensures that key physical and climatic drivers of fire risk are effectively captured, supporting more accurate mapping and decision-making in fire-prone landscapes.

2.2 Dataset

Accurate and comprehensive data selection is crucial for reliable wildfire risk mapping. In this study, a multi-source dataset was constructed by integrating spectral, topographic, and meteorological variables. These data sources were chosen to capture the key environmental conditions and changes associated with fire occurrence, behavior, and post-fire impacts. Data were obtained from globally recognized satellite missions and climate reanalysis datasets to ensure consistency, spatial coverage, and temporal reliability. Table (1) provides a detailed summary of all variables used in the analysis, including their source, spatial resolution, and analytical purpose. These data layers were selected for their known influence on wildfire ignition, spread, and intensity, and were harmonized to support comparative and integrative modeling using both Sentinel-2 and Landsat-8 imagery.

Data Variable	Source	Spatial Resolution	Purpose / Use
Visible and NIR bands	Sentinel-2 MSI / Landsat-8 OLI	10m (S2), 30m (L8)	Vegetation health, land cover classification
SWIR bands	Sentinel-2 MSI / Landsat-8 OLI	20m (S2), 30m (L8)	Burned area detection, soil/vegetation moisture
NDVI	Derived from (S2/L8)	10–30m	Vegetation index
NDVI before / NDVI after	Derived temporal NDVI composites	10–30m	Pre- and post-fire vegetation comparison
dNDVI	NDVI after – NDVI before	10–30m	Change in vegetation due to fire
NBR	Derived from NIR and SWIR bands	20–30m	Fire severity and burn area detection
dNBR	Pre- and post-fire NBR difference	20–30m	Fire impact severity
RBR	Derived index from NBR and dNBR	20–30m	Relative burn intensity
RdNBR	dNBR normalized by square root of pre-NBR	20–30m	Adjusted fire severity
RdNDVI	dNDVI normalized by pre-NDVI	10–30m	Relative vegetation loss
Average LST	ERA5	~10km	Land surface thermal profile
Elevation	SRTM DEM	30m	Topographic influence on fire behavior
Slope	Derived from DEM	30m	Terrain steepness, affects fire spread rate
Aspect	Derived from DEM	30m	Terrain orientation, affects dryness and sunlight exposure
Total Precipitation	ERA5	~10km	Climatic input influencing vegetation moisture
Wind Direction	ERA5	~10km	Directional influence on fire propagation
Wind Speed	ERA5	~10km	Fire spread potential

Table 1. Dataset used in the research

This extensive and harmonized dataset forms a robust analytical foundation for wildfire risk modelling. Figure 2 shows some of the main spectral indices obtained from the two satellites Sentinel-2 and Landsat-8.

https://doi.org/10.5194/isprs-annals-X-4-W8-2025-509-2026 |

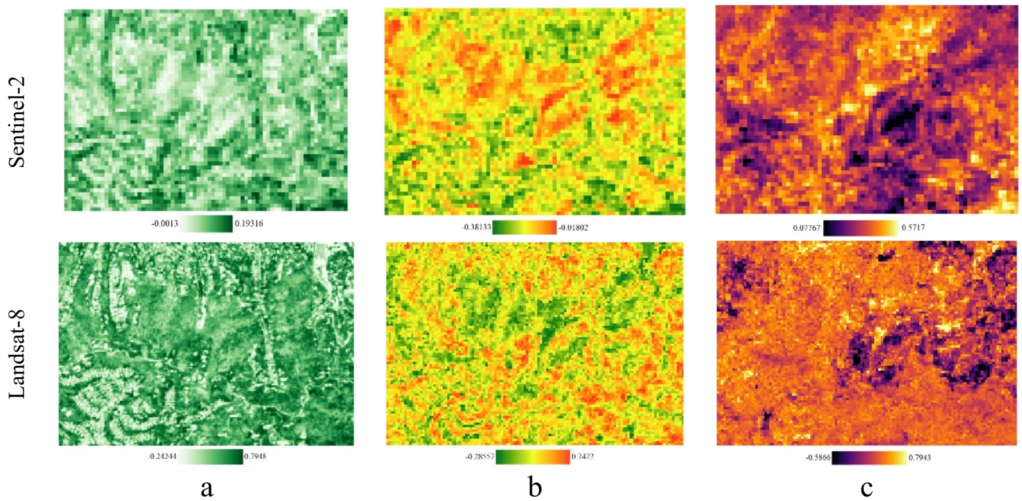


Figure 2. Spectral indices obtained from Sentinel-2 and Landsat-8 images, a) NDVI, b) NBR, and c) RBR.

3. Methodology

In this study, two advanced machine learning models, Random Forest (RF) (Breiman 2001; Cutler et al., 2007) and Extreme Gradient Boosting (XGBoost) (Ramraj et al., 2016), were employed to estimate and map the fire hazard zones in the study area. Both models are ensemble learning techniques capable of handling complex, nonlinear relationships between multiple input variables and the target variable, making them suitable for environmental hazard mapping where multiple interacting factors contribute to fire risk.

Fire Occurrence Data:

Fire occurrence records were obtained from the Fire Information for Resource Management System (FIRMS), which delivers active fire detections derived from MODIS and VIIRS satellite sensors. To improve spatial granularity and accuracy, Landsat-derived burned area products were also utilized. The fire dataset covers a temporal span from 2016 to 2025 and was used to label the presence or absence of fire at specific points.

Sample Labelling and Fire Probability Estimation:

Each point in the study area was labelled based on its spatial relationship to observed fire events. A circular buffer was generated around each point, and fire presence was determined by detecting any FIRMS or Landsat fire pixels within this proximity threshold. The probability of fire occurrence was then calculated based on the distance to the nearest fire event, allowing for a spatially weighted assessment of fire risk.

Dataset Compilation:

All relevant features, including spectral indices (e.g., NDVI, NBR, RdNDVI), land surface temperature (LST), wind speed and direction, precipitation, elevation, slope, and aspect, were extracted and compiled for each sample point. The dataset was then temporally segmented as follows.

- Training period: 2016–2024
- Evaluation period: 2025

This temporal division enabled the development of predictive models using past data and their validation on future, unseen conditions, simulating operational forecasting scenarios.

3.1 Model Development and Assessment

3.1.1 Random Forest (RF)

Random Forest is an ensemble learning algorithm that combines multiple decision trees trained on random subsets of data and features through bootstrap aggregating (bagging). This randomness enhances model diversity, reducing overfitting and improving generalization. Final predictions are obtained by averaging outputs for regression or majority voting for classification. In fire hazard zoning, Random Forest uses environmental, climatic, topographic, and vegetation variables—such as NDVI, NBR, land surface temperature, elevation, slope, aspect, precipitation, and past fire events—to model fire occurrence patterns and predict relative hazard levels. The method effectively handles high-dimensional data, tolerates missing values, and provides feature importance insights to identify key fire risk drivers.

3.1.2 Extreme Gradient Boosting (XGBoost)

XGBoost is an efficient and scalable implementation of gradient boosting that enhances traditional methods through regularization, optimized tree learning, and effective handling of missing data. It constructs an ensemble of decision trees sequentially, where each tree corrects the errors of its predecessors, achieving high predictive accuracy and reduced overfitting. For fire hazard zoning, the XGBoost model was trained with the same input variables as the Random Forest model and optimized by minimizing appropriate loss functions. Its ability to model complex nonlinear relationships among environmental variables makes it suitable for capturing the multifactorial nature of fire hazards. Furthermore, XGBoost provides feature importance metrics, aiding the interpretation of key drivers of fire susceptibility. The combined use of Random Forest and XGBoost enables robust and interpretable fire hazard zoning, supporting improved wildfire risk assessment and management.

4. Results and Discussion

The main objective of this research is to estimate a wildfire risk zoning map and to examine the factors influencing wildfire risk using satellite data, including climatic data and images as well as indices related to the Sentinel-2 and Landsat-8 sensors. In

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-W8-2025-509-2026

this context, the impact of satellite image resolution on the estimation of wildfire risk will also be analyzed. Initially, a dataset comprising factors that may affect wildfire risk will be created, and then two machine learning models will be employed to estimate the wildfire risk for each point.

In this paper, we evaluated the effectiveness of the model using a comprehensive set of evaluation metrics. These included the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) (Chai and Draxler 2014; Chicco, Warrens, and Jurman 2021), and R-squared (R^2)(Cameron and Windmeijer 1997). Each of these metrics was chosen to provide unique insights into the model's prediction accuracy and dependability.

	Models	R^2	RMSE	MAE	Max error
Sentinel-2	RF	97.92	0.022	0.007	0.454
	XGBoost	99.78	0.007	0.002	0.238
Landsat-8	RF	93.77	0.038	0.015	0.479
	XGBoost	97.41	0.024	0.005	0.424

Table 2. Evaluation criteria values.

The high accuracy observed in this study may be attributed to the inclusion of training samples located in close spatial and temporal proximity to the 2025 fire point. This unintended correlation, likely resulting from the extensive fire event in 2024, may have contributed to the elevated accuracy of the model's predictions.

The results of the forest fire risk zoning will be visually presented. As demonstrated in the figure below, both models provide acceptable outcomes in the context of fire risk zoning and have successfully estimated the fire risk map for the year 2025. As shown in the figure below, both models have generally performed well in estimating the probability of forest fires. Additionally, when comparing the two deep learning models, the XGBoost model has outperformed the RF model, demonstrating a higher accuracy in estimating the probability of forest fires. Furthermore, when comparing the results obtained from the two datasets, Sentinel-2 and Landsat-8, it is evident that the results derived from the Sentinel-2 satellite have yielded better outcomes than those from Landsat-8. This indicates that the difference in spatial resolution between these two sensors can significantly impact the estimation process of forest fire probabilities.

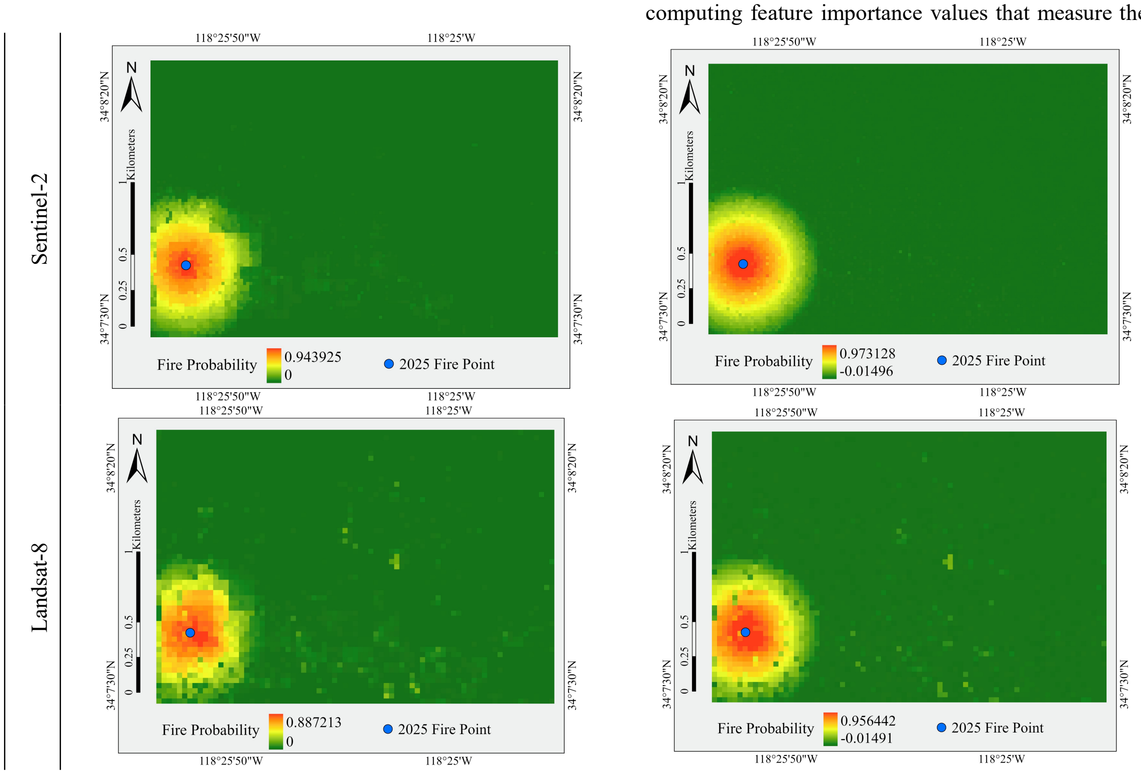
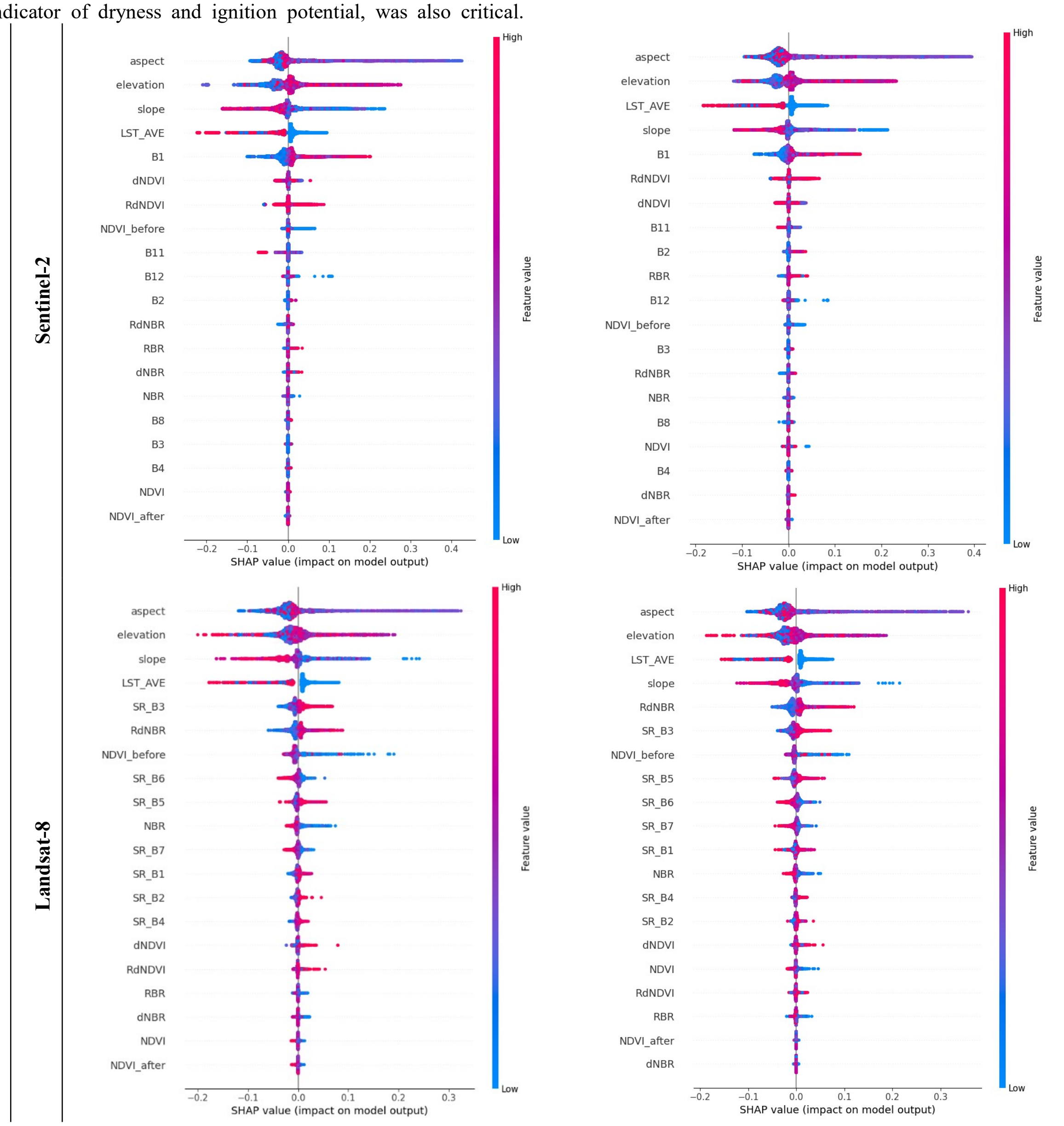


Figure 3. Forest fire risk zoning map, results from the a) RF and, b) XGBoost models.

This study not only compares the performance of Sentinel-2 and Landsat-8 imagery but also evaluates the influence of various factors on forest fire risk estimation. Climatic parameters, satellite-derived variables, and relevant indices were used as model inputs. The SHAP library (Hu and Wang 2023) was applied to quantify the contribution of each parameter by computing feature importance values that measure their impact on model predictions.

The analysis revealed that topographic and climatic factors, particularly elevation, slope, aspect, and surface temperature, play major roles in predicting fire risk. Elevation strongly influenced fire susceptibility through its effects on vegetation type, temperature, and moisture conditions. Slope demonstrated high importance, as steeper terrains promote faster fire spread via preheating of vegetation. Aspect analysis showed that south-facing slopes exhibited greater fire risk due to higher solar exposure and drier conditions. Surface temperature, as a direct indicator of dryness and ignition potential, was also critical. These findings highlight the necessity of integrating topographic and thermal parameters for accurate wildfire risk assessment.



a b

Figure 4. The importance of each of the input parameters of the a) RF and, b) XGBoost models.

The findings of this research highlight that the primary contribution of the study lies in the comprehensive comparative analysis of Sentinel-2 and Landsat-8 satellite data, including their derived spectral and climatic indices, as well as the evaluation of two robust machine learning algorithms, Random Forest and XGBoost, for wildfire estimation. This comparative approach provides valuable insights into the relative performance, advantages, and limitations of each sensor–model combination, thereby contributing to the optimization of data and model selection strategies for operational wildfire risk assessment.

5. Conclusion

This study highlights the transformative potential of integrating remote sensing technology with advanced analytical methods to enhance understanding and prediction of wildfire risks. Through a detailed comparative analysis of Sentinel-2 and Landsat-8 satellite data, it was found that the former significantly outperforms the latter in estimating fire susceptibility due to its enhanced spatial resolution. The incorporation of auxiliary datasets—featuring both climatic and topographic variables—into machine learning models such as Random Forest and Extreme Gradient Boosting proved crucial in achieving high accuracy and interpretability in predictions of fire probability. The use of the SHAP library provided critical insights into the importance of various environmental factors, demonstrating that elevation, slope, aspect, and surface temperature are significant predictors of fire risk. Elevation emerged as a paramount determinant, influencing vegetation types and moisture conditions, while slope and aspect played critical roles in fire dynamics, facilitating the speed of fire spread and contributing to localized fire susceptibility. The findings necessitate the continued integration of these environmental parameters into fire risk assessment frameworks. Ultimately, the predictive models developed in this research offer valuable insights for policymakers and resource managers, enabling more informed decisions to mitigate the adverse effects of forest fires. As the threat of wildfires continues to rise with climate change, the establishment of robust early warning systems and risk management plans becomes imperative. Future investigations should focus on refining these models further, exploring the application of real-time data inputs, and continuing to leverage tools like SHAP for greater interpretability, thereby enhancing predictive capabilities and operational efficiency in wildfire management strategies.

References

Breiman, Leo. 2001: 'Random forests', *Machine learning*, 45: 5-32.

Cameron, A Colin, and Frank AG %J Journal of econometrics Windmeijer. 1997: 'An R-squared measure of goodness of fit for some common nonlinear regression models', 77: 329-42.

Chai, Tianfeng, and Roland R %J Geoscientific model development Draxler. 2014: 'Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature', 7: 1247-50.

Chicco, Davide, Matthijs J Warrens, and Giuseppe %J Peerj computer science Jurman. 2021: 'The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation', 7: e623.

Cutler, D Richard, Thomas C Edwards Jr, Karen H Beard, Adele Cutler, Kyle T Hess, Jacob Gibson, and Joshua J Lawler. 2007: 'Random forests for classification in ecology', *Ecology*, 88: 2783-92.

Hu, Linwei, and Ke Wang. 2023: 'Computing SHAP Efficiently Using Model Structure Information', *arXiv preprint arXiv:2309.02417*.

Müller, Markus U, Nikoo Ekhtiari, Rodrigo M Almeida, and Christoph %J arXiv preprint arXiv:.00580 Rieke. 2020: 'Super-resolution of multispectral satellite images using convolutional neural networks'.

Ramraj, Santhanam, Nishant Uzir, R Sunil, Shatadeep %J International Journal of Control Theory Banerjee, and Applications. 2016: 'Experimenting XGBoost algorithm for prediction and classification of different datasets', 9: 651-62.

Xie, Lingxiao, Rui Zhang, Junyu Zhan, Song Li, Age Shama, Runqing Zhan, Ting Wang, Jichao Lv, Xin Bao, and Renzhe Wu. 2022: 'Wildfire risk assessment in Liangshan Prefecture, China based on an integration machine learning algorithm', *Remote Sensing*, 14: 4592.

Zhu, Duowang, Xiaohu Huang, Haiyan Huang, Zhenfeng Shao, and Qimin %J arXiv preprint arXiv:.12847 Cheng. 2024: 'ChangeViT: Unleashing plain vision transformers for change detection'.