



## OPEN ACCESS

## EDITED BY

Ana Cristina Russo,  
University of Lisbon, Portugal

## REVIEWED BY

Huilin Huang,  
Pacific Northwest National Laboratory (DOE),  
United States  
Sergey Venevsky,  
Southern Scientific Center (RAS), Russia

## \*CORRESPONDENCE

Qingqing Yang  
✉ 624767243@qq.com

RECEIVED 13 September 2024

ACCEPTED 05 May 2025

PUBLISHED 22 May 2025

## CITATION

Chen X, Yang Q, Chen Z, Lei J, Wu T, Li Y and  
Pan X (2025) The interaction between  
temperature and rainfall determines the  
probability of tropical forest fire occurrence in  
Hainan Island.

*Front. For. Glob. Change* 8:1495699.

doi: 10.3389/ffgc.2025.1495699

## COPYRIGHT

© 2025 Chen, Yang, Chen, Lei, Wu, Li and  
Pan. This is an open-access article distributed  
under the terms of the [Creative Commons  
Attribution License \(CC BY\)](#). The use,  
distribution or reproduction in other forums is  
permitted, provided the original author(s) and  
the copyright owner(s) are credited and that  
the original publication in this journal is cited,  
in accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# The interaction between temperature and rainfall determines the probability of tropical forest fire occurrence in Hainan Island

Xiaohua Chen<sup>1,2</sup>, Qingqing Yang<sup>1,2\*</sup>, Zongzhu Chen<sup>1</sup>, Jinrui Lei<sup>1</sup>,  
Tingtian Wu<sup>1</sup>, Yuanling Li<sup>1</sup> and Xiaoyan Pan<sup>1</sup>

<sup>1</sup>Hainan Academy of Forestry (Hainan Academy of Mangrove), Haikou, China, <sup>2</sup>Key Laboratory of Tropical Forestry Resources Monitoring and Application of Hainan Province, Haikou, China

Severe forest fires have erupted in numerous tropical regions globally, threatening carbon storage in tropical ecosystems, the survival of plant species, and human health. Consequently, developing more precise prediction models for tropical forest fire hazards is essential for establishing effective fire prevention and management strategies. Although traditional logistic regression is widely employed for mapping forest fire probabilities, machine learning methods such as random forest have become more prevalent over the past decade. The applicability of random forest and logistic regression in predicting tropical forest fire probabilities has not been explored, leading to insufficient understanding of the driving factors of tropical forest fires on this tropical continental island with diverse forest types. This study integrated ground-based fire statistics from the Hainan Forestry Department and moderate resolution imaging spectroradiometer (MODIS) fire point data to create a highly accurate forest fire dataset for Hainan Island, spanning 20 years (2000–2020). Both logistic regression and random forest were used to develop tropical forest fire hazard models and explore the driving mechanisms of fires on Hainan Island. The results show that: (1) climatic factors contribute most significantly to the tropical forest fire probability, followed by human activities and topography, while vegetation factors (i.e., normalized difference vegetation index) made no significant contribution; (2) temperature and rainfall are the dominant factors influencing fire probability, with rising temperatures and decreasing rainfall substantially increasing the forest fire hazard; and (3) both logistic regression and random forest are reliable for predicting tropical forest fire hazards, but random forest demonstrates greater adaptability. In conclusion, our evidence suggests that the probability of tropical forest fires will increase under global warming and drought. The logistic regression and random forest models developed in this study provide valuable insights for identifying high-hazard forest fire areas in tropical regions. These findings have important implications for global tropical forest management and fire prevention, aiding in the formulation of targeted control strategies.

## KEYWORDS

tropical forest fires, forest fire ecology, fire drivers, climate change impact, global warming

# 1 Introduction

Tropical forests are among the most biodiverse and productive terrestrial ecosystems on Earth, harboring over 50% of the planet's plant and animal species while contributing more than one-third of global oxygen production (Gatti et al., 2021; Lewis et al., 2015). Although tropical forests are located in humid environments, wildfires remain highly prevalent in these ecosystems (Juárez-Orozco et al., 2017; Ngoc Thach et al., 2018; Sze et al., 2019). According to reports, forest fires (including wildfires and those associated with deforestation activities) accounted for 62% of the total burned area across all land-use types in equatorial Asia during the period from July 1996 to August 2012 (Giglio et al., 2013). These fires not only disrupt natural vegetation succession and carbon sequestration capacity in ecosystems, but also generate substantial amounts of harmful gases, posing significant threats to both the environment and human health (Brando et al., 2019; Trang et al., 2023a; Withey et al., 2018). Under the background of ongoing climate change, with increasing drought frequency and rising mean temperatures, the likelihood of wildfires is expected to rise substantially (Goss et al., 2020). In this context, it has become critically urgent to identify threshold values of wildfire drivers and establish efficient hazard prediction models for tropical forests.

Wildfires are commonly associated with meteorological conditions, vegetation types, topography, and human activities (Guo et al., 2020; Haas et al., 2024; Juárez-Orozco et al., 2017; Mohammadi et al., 2014). Globally, researchers have employed a variety of quantitative methods to examine the drivers of forest fires. Early studies predominantly relied on traditional statistical and knowledge-based methods, such as logistic regression (Mohammadi et al., 2014) and generalized linear models (Guo et al., 2015). With technological advances, more powerful machine learning algorithms, including random forests, support vector machines, and artificial neural networks, have been applied to this field (Abid, 2021; Alkhatib et al., 2023). In Canada, scholars utilized an improved logistic regression model to predict coniferous crown fire probability, demonstrating that the final model incorporating wind speed, fuel layer spacing, litter moisture, and surface fuel consumption factors achieved superior accuracy (Perrakis et al., 2023). In Jilin Province, northern China, Gao et al. compared three traditional methods (logistic regression, geographically weighted logistic regression, and Lasso regression) with two machine learning approaches (random forest and support vector machines) for temperate forest fire prediction, revealing that the machine learning models outperformed the generalized linear regression models in predictive accuracy (Gao et al., 2024). In the Mediterranean region, Milanovic et al. applied random forest and logistic regression models to assess forest fire occurrence probability in eastern Serbia, finding that the random forest model exhibited superior overall predictive capability (Milanovic et al., 2021). In Vietnam, Tehrany et al. implemented the LogitBoost ensemble-based decision tree method for spatial prediction of tropical forest fire susceptibility, demonstrating optimal performance (92% predictive accuracy) across both training and validation datasets (Tehrany et al., 2019). Despite these advances in fire risk modeling across various ecosystems, current research lacks a systematic comparison of the applicability of Random Forest and Logistic Regression models for predicting fire occurrence probability in tropical moist forests.

Data regarding forest fires and their drivers are critical for drawing accurate conclusions on forest fire prediction modeling. Major sources

of such data include ground-based surveys and remote sensing. Ground surveys, which depend on fire monitoring stations equipped with sensors, collect data on meteorological conditions, fuel moisture, fire locations, and fire extent (Barmpoutis et al., 2020). While ground-based surveys provide high precision, they are often spatially limited due to funding constraints. In contrast, satellite remote sensing offers a faster, more cost-effective way to collect fire event, climate, vegetation, topography, and socioeconomic data over larger areas (Szpakowski and Jensen, 2019). However, the accuracy of satellite remote sensing is constrained in tropical regions due to frequent cloud cover (Miettinen et al., 2013). As ground fire data are often unavailable, most existing studies rely solely on remote sensing to explore the drivers of tropical forest fires (Su et al., 2021; Trang et al., 2022; Trang et al., 2023b). This limits the reliability of research findings. Therefore, we argue that the combined analysis of both ground and satellite data is essential for obtaining a comprehensive understanding of the causes of tropical forest fires. Furthermore, most existing studies on tropical forest fire drivers focus on single forest types, such as natural forests in the Amazon, plantations in southeast Asia, or savannas in central Africa (Dwomoh et al., 2019; Fonseca et al., 2017; Kouassi et al., 2020; Trang et al., 2023a). A comprehensive analysis of ecosystems with multiple forest types is lacking, hindering a full understanding of global tropical forest fire dynamics. Hainan Island, a continental island located at the intersection of the Asian tropical rainforest and the global monsoon evergreen broad-leaved forest, hosts diverse tropical forest types, including primary natural forests, plantations, shrublands, and mangroves (Huang et al., 2021; Jiang et al., 2016). This mosaic of forest ecosystems interspersed with agricultural land and a dense population makes Hainan Island an ideal location for the comprehensive study of the drivers of tropical forest fires. However, the knowledge of the driving factors of forest fires on Hainan Island remains limited. Thus, a detailed study of forest fire drivers on Hainan Island could provide valuable insights for global tropical forest fire prevention and management.

Here, we integrated fire statistics from the Hainan Forestry Department and moderate resolution imaging spectroradiometer (MODIS) fire point data to develop a comprehensive 20-year (2000–2020) forest fire dataset for Hainan Island. Two approaches were employed to predict the probability of forest fire occurrence: a traditional statistical method and an emerging machine-learning technique. Initially, this study utilized logistic regression, a widely adopted method in forest fire prediction research that is renowned for its interpretability and robust performance in predicting binary outcomes. To achieve enhanced predictive capacity and capture the complex nonlinear relationships among variables, this study subsequently implemented a random forest model. The two main objectives of the present study were to (1) construct higher-accuracy predictive models for fire occurrence probability in tropical forests, and (2) identify threshold values of tropical forest wildfire drivers. To this end, this study addressed the following research questions: (1) which method, logistic regression or random forest, demonstrates greater applicability in constructing predictive models for the probability of tropical forest fires in Hainan Island? (2) What are the critical thresholds of key environmental drivers affecting tropical forest wildfire probability? We consider that this information holds significant value in the context of increasing wildfire incidence driven by climate change and rapid socioeconomic development in tropical regions.

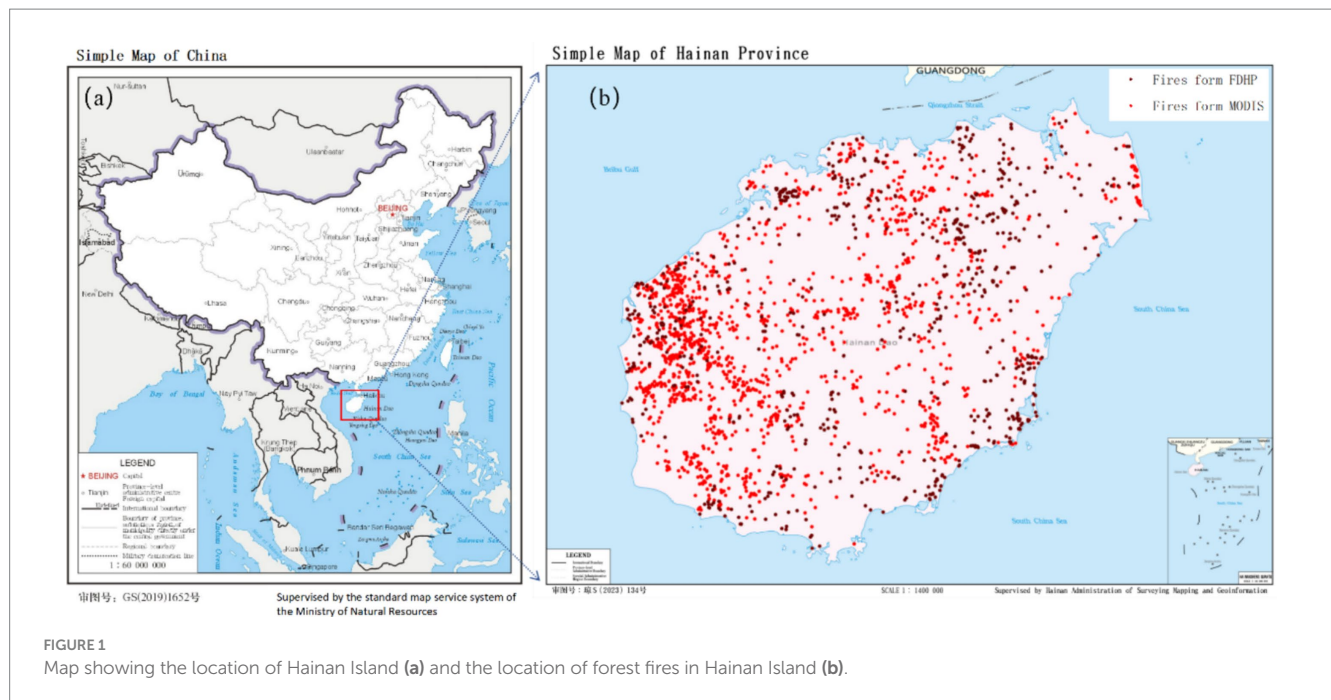


FIGURE 1

Map showing the location of Hainan Island (a) and the location of forest fires in Hainan Island (b).

## 2 Materials and methods

### 2.1 Study area

The tropical island of Hainan, located in China, was selected as the study area (Figure 1a). Hainan Island is a continental island situated between 18.80°N–20.10°N latitude and 108.37°E–111.03°E longitude, covering an area of 33,900 km<sup>2</sup>. The island belongs to the monsoon tropical climate zone, with an annual average temperature of 25°C and annual average rainfall exceeding 1,600 mm. The southwestern part of the island tends to be drier, with the annual rainfall being around 1,000 mm. The island experiences distinct wet and dry seasons, with the dry season spanning from November to April and the wet season spanning from May through October. Rainfall during the wet season accounts for 70–90% of the total annual precipitation.

Hainan Island's topography is characterized by higher elevations at the center, with the terrain gradually sloping downwards toward the coastal plains. The landscape includes mid-mountains, low mountains, and coastal plains. Forests are mainly concentrated in the central part of the island, covering over 60% of the total area and amounting to 21,976 km<sup>2</sup>. Hainan Island has diverse forest types, including primary tropical rainforests, mangroves, and various types of economically cultivated plantations. The government places great importance on forest fire prevention and management in this area. The designated forest fire prevention period runs from December 1 through May 31, with a high-hazard period from March 1 to May 31.

### 2.2 Data

#### 2.2.1 Forest fire data

All forest fire statistical data for the past two decades (2000–2020) on Hainan Island were obtained from the Forestry Department of

Hainan Province (FDHP). Experts from the FDHP collected data on the time and location of each forest fire occurrence in the field from 2000 to 2020. Following each fire incident, FDHP experts used handheld global positioning system (GPS) devices (UniStrong, China) to record the location of each fire as the center of the burned area. Furthermore, to supplement potentially missed recorded fire point data, MODIS fire locations were downloaded from the National Aeronautics and Space Administration (NASA). Only forest fires occurring on Hainan Island were extracted from the MODIS data to ensure that fires in agricultural fields and industrial areas were not included in the analysis. This study compared the two datasets and prioritized using the fire point data from the FDHP. Finally, a forest fire dataset for Hainan Island over the past two decades (2000–2020) was compiled. This dataset consists of 1,526 fire points, with 744 fire points from the FDHP and 782 fire points from MODIS data (Figure 1b). This dataset represents the closest approximation to the actual occurrences of forest fires.

#### 2.2.2 Climate, topography, and human activity data

For this study, climate data were selected based on four indicators (the average temperature, average wind speed, average rainfall, and minimum relative humidity). The selection of these climate indicators was based on previous research findings and data availability (Eskandari et al., 2020; Wang et al., 2023). These climate indicator data were provided by the National Meteorological Information Center (NMIC) and were collected from meteorological stations located on Hainan Island. The inverse distance weight (IDW) method in ArcMap 10.8<sup>1</sup> was employed to perform spatial interpolation on the obtained meteorological data (Xiao et al.,

<sup>1</sup> <https://www.esri.com/en-us/home>

2016), thereby obtaining spatial distribution data for the entire island (Supplementary Figure S1). The formulas used for inverse distance weighting are as follows:

$$\hat{Z}(s_o) = \sum_j^n \lambda_j Z(s_j) \quad (1)$$

$$\lambda_j = \frac{d_j^{-p}}{\sum_j^n d_j^{-p}} \quad (2)$$

$$\sum_j^n \lambda_j = 1 \quad (3)$$

where  $\hat{Z}(s_o)$  represents the interpolated result at point  $s_o$ ;  $Z(s_j)$  represents the observed value at point  $s_j$ ;  $\lambda_j$  represents the weight of each sampling point; and  $d_j$  represents the distance between the interpolation point and the sampling point. As shown in Equations 1–3, the sum of the weights of all sampling points is equal to 1.

In addition to climate variables, this study selected other indicators that may affect forest fire hazards, including the altitude above sea level, degree of slope, aspect of slope, normalized difference vegetation index (NDVI), proximity to roads, and proximity to residence zones (Supplementary Figure S2). ArcMap 10.8 was employed to merge, clip, and process the downloaded data to make them usable for analysis (Table 1).

## 2.3 Forest fire driving factor analysis method

One of the key focuses of this study was to investigate the effects of climate, vegetation, topography, and human activities on the probability of forest fire occurrences. The present study utilized forest fire points in Hainan Island as the dependent variable and indicators of climate, vegetation, topography, and human activities as independent variables to determine the relative importance of different forest fire drivers using the GeoDetector method and the mean decrease accuracy (MDA) index.

The GeoDetector model is a statistical tool designed to explore spatial heterogeneity and identify its underlying driving forces (Wang et al., 2010). Because the GeoDetector method requires categorical data for independent variables, the continuous independent variables were discretized. The optimal number of intervals for discretization was determined based on the maximum Q value, a metric computed using the GeoDetector model (Supplementary Figure S3). In this study, the factor detection function of GeoDetector was employed to explain the relative importance of wildfire driving factors, with the level of explanation measured based on the Q value. The Q value is calculated using the following formula (Equation 4):

$$Q = 1 - \frac{\sum_{h=1}^L N_h \sigma_h^2}{N \sigma^2} \quad (4)$$

where  $h = 1, \dots, L$  represents the strata of variable Y or factor X;  $N_h$  and  $N$  are the strata  $h$  and the number of units in the whole area,

TABLE 1 List of data used in this study.

Data	Abbreviation	Units	Spatial resolution	Temporal coverage	Source
Forest fire data based on field investigation				2000–2020	Forestry Department of Hainan Province (FDHP) <a href="https://lyj.hainan.gov.cn/">https://lyj.hainan.gov.cn/</a>
Forest fire data based on moderate resolution imaging spectroradiometer			1 km	2000–2020	NASA fire information for Resource Management System (FIRMS) <a href="https://firms.modaps.eosdis.nasa.gov">https://firms.modaps.eosdis.nasa.gov</a>
Average temperature	AveTemp	°C		Monthly from 2000 to 2020	National Meteorological Information Center <a href="https://data.cma.cn/">https://data.cma.cn/</a>
Average wind speed	AveWind	m/s			
Average rainfall	AveRain	mm			
Minimum relative humidity	MiniHumi	%			
Altitude above sea level	Altitude	m	90 m	2020	NASA SRTM Digital Elevation 30 m Google Earth Engine (GEE) <a href="https://developers.google.cn/earth-engine/">https://developers.google.cn/earth-engine/</a>
Degree of slope	Slope	°	90 m		
Aspect of slope	Aspect		90 m		
Normalized difference vegetation index	NDVI		1 km	Monthly from 2000 to 2020	Landsat series data Google Earth Engine (GEE) <a href="https://developers.google.cn/earth-engine/">https://developers.google.cn/earth-engine/</a>
Proximity to road	ProxiRoad	m	1:1,000,000	Daily from 2020	Open street map <a href="https://www.openstreetmap.org/">https://www.openstreetmap.org/</a>
Proximity to residence zone	ProxiResidence	m	1:1,000,000	Daily from 2020	Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-1C Google Earth Engine (GEE) <a href="https://earthengine.google.com/">https://earthengine.google.com/</a>

respectively; and  $\sigma_h^2$  and  $\sigma$  are the variances of the Y value of strata  $h$  and the whole area, respectively. The value range of Q is [0,1], and a larger value of Q indicates a stronger determination ability or relative importance of the independent factor X on the attribute Y. The GeoDetector calculations were performed using the “GD” package in RStudio version 4.3.1.

The MDA index is a metric used to evaluate the importance of each independent variable (feature) in a random forest model. This index measures the impact of randomly permuting each variable on the model’s performance. In the present study, the MDA computations were implemented using the “RandomForest” package in RStudio version 4.3.1.

## 2.4 Construction of forest fire occurrence probability prediction model

### 2.4.1 Logistic regression model

Logistic regression modeling has become one of the most common techniques in forest fire research due to its excellent interpretability and predictive accuracy. Before utilizing the logistic regression model, this study conducted Spearman’s correlation coefficient tests to assess the strengths of the relationships among all factors, aiming to determine the degree of correlation between them. Furthermore, the variance inflation factor (VIF) was employed to detect multicollinearity among the independent variables, thereby determining which factors could ultimately be utilized in the model as independent variables. The formula used to calculate the VIF is as follows (Equation 5):

$$VIF = \frac{1}{1 - R^2} \quad (5)$$

Only variables with a  $VIF \leq 10$  and a Spearman’s correlation coefficient below 0.7 were considered for model construction (Dormann et al., 2013). Spearman’s correlation coefficient tests and multicollinearity checks were performed using the “car” and “corrplot” packages in RStudio 4.3.1 software.

A logistic regression model was used to predict the probability of forest fire occurrence in the study area. Assuming that the probability of a forest fire at a certain point is  $P$ ; then, the probability of no forest fire is  $(1 - P)$ . The formula for calculating  $P$  is as follows (Equation 6):

$$P = \frac{e^{(a_0 + a_1x_1 + \dots + a_mx_m)}}{1 + e^{(a_0 + a_1x_1 + \dots + a_mx_m)}} = \frac{1}{1 + e^{-(a_0 + a_1x_1 + \dots + a_mx_m)}} \quad (6)$$

where  $P$  represents the probability of forest fire occurrence;  $x_1, x_2, \dots, x_m$  represent the driving factors influencing forest fire points on Hainan Island;  $a_0$  is the intercept; and  $a_1, a_2, \dots, a_m$  represent the coefficients corresponding to each driving factor in the logistic regression model.

This study randomly generated 2,289 non-fire points within the forest distribution range of the study area, based on 1.5 times the quantity of forest fire points. These non-fire points, along with 1,526 forest fire points, were binary coded in ArcMap 10.8 software, with a value of 1 representing fire occurrence and a value of 0 representing no fire occurrence. The “caTools” package in RStudio 4.3.1 software was utilized to create this model.

### 2.4.2 Random forest model

Random forest is an ensemble learning method that builds a model comprising multiple decision trees. These trees are constructed through bootstrap sampling of the training data and random feature selection, and classification or regression is performed by aggregating predictions through voting or averaging. In the present study, the “caret” package in R was employed to optimize the parameters. The initial value for the number of features (mtry) was set to 2, with an increment of 2 for each iteration, and the maximum value was set to 16. Ultimately, this study selected the number of features (mtry) that yielded the highest accuracy to construct the random forest model. This model was created using the “RandomForest” package in RStudio 4.3.1.

## 2.5 Forest fire susceptibility map generation and accuracy assessment methodology

To create the fire susceptibility map, 70% of the fire points were utilized as training samples and the remaining 30% were used as validation samples. The logistic regression model constructed in this study was employed to compute the predicted probabilities of fire occurrence based on the validation samples. Subsequently, the Kriging interpolation tool in ArcMap 10.8 software was used to generate the fire susceptibility map.

The accuracy of the fire susceptibility map was validated based on the receiver operating characteristic (ROC) curve. The area under the curve (AUC) of the ROC curve indicates the model’s accuracy. The AUC between two points can be calculated using definite integration, providing a value ranging from 0.5 to 1. An AUC value between 0.5 and 0.7 reflects poor accuracy, values between 0.7 and 0.8 indicate acceptable accuracy, values between 0.8 and 0.9 represent high accuracy, and values above 0.9 indicate outstanding model accuracy (Hosmer Jr et al., 2013). The ROC curve plotting and calculations were performed using the “pROC” package in RStudio 4.3.1 software.

## 3 Results

### 3.1 Relationship between forest fire incidence and climate, topography, vegetation, and human activities

The findings demonstrated that the number of forest fires exhibited a trend of initial increase followed by a decrease with rising annual average temperature, which suggests that a higher temperature does not necessarily correlate with a higher frequency of forest fires (Figure 2a). The highest number of forest fire occurrences was observed when the annual average temperature ranged from 24.5°C to 24.9°C, accounting for 44.4% of the total forest fire occurrences. The maximum number of forest fire points was observed at an average wind speed of 1.9–2.7 m/s, constituting 32.8% of the total. This phenomenon may be attributed to Hainan Island’s tropical maritime monsoon climate, where strong winds during the monsoon season are often accompanied by rainfall, which further reduces fire risks by

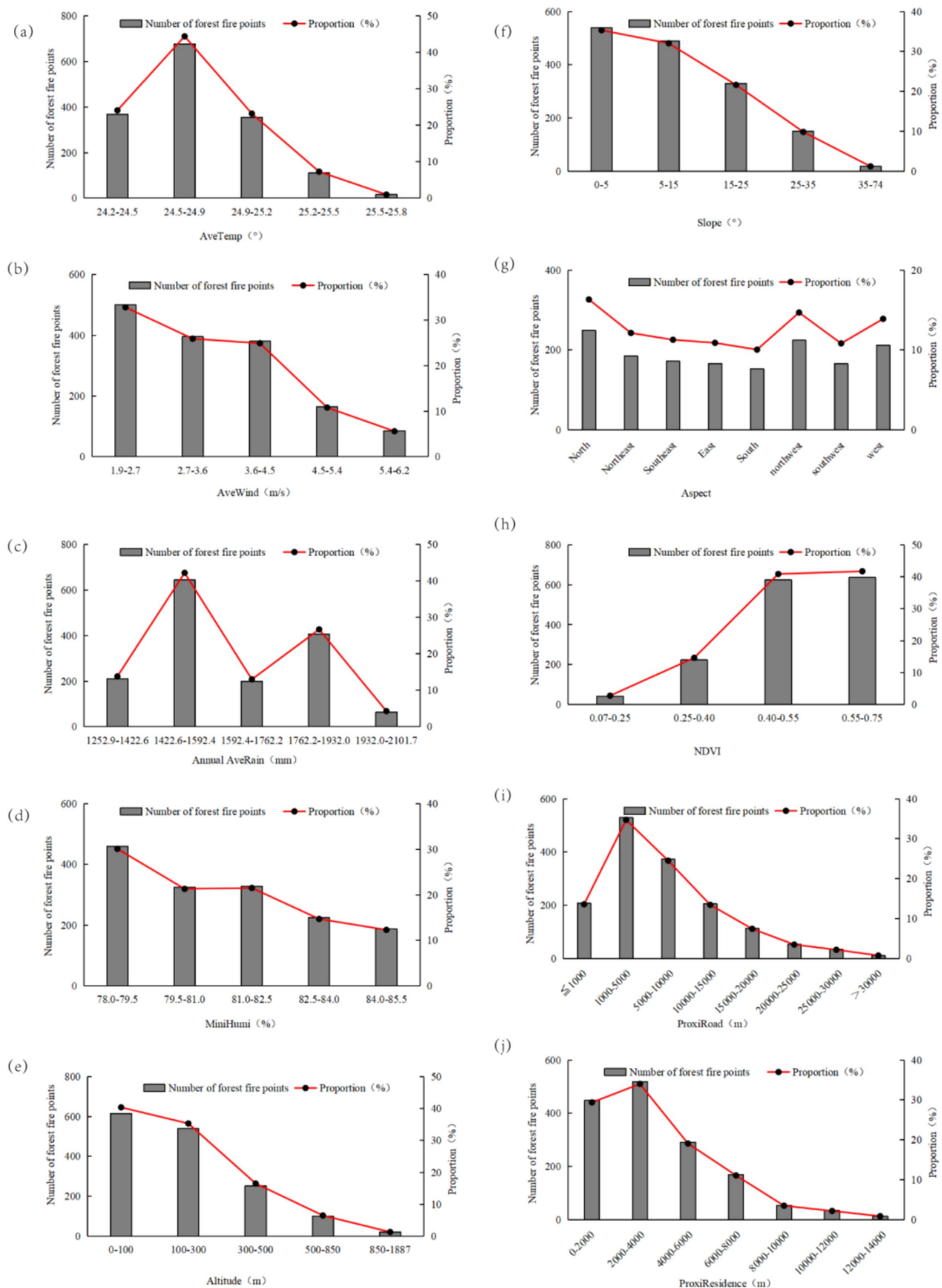


FIGURE 2

Forest fire incidence and overlay of climatic factors (a–d), topographic factors (e–g), vegetation factors (h), and human activity factors (i,j) in Hainan Island from 2000 to 2020. Note the difference in y-scale between the left and right panels.

moistening vegetation. Forest fire incidence exhibited a decreasing trend with increasing average wind speed (Figure 2b). The peak number of forest fire points occurred when the annual rainfall ranged from 1422.6 to 1592.4 mm, representing 42.2% of the total (Figure 2c). Overall, there was a decreasing trend in forest fire points with elevated annual minimum relative humidity (Figure 2d). As the altitude increased, the number of forest fire points significantly decreased (Figure 2e). Forest fire points were concentrated between slopes of 0°–25°, with fewer occurrences as the slope steepness increased (Figure 2f). The distribution across different aspects of slopes was relatively uniform (Figure 2g). Regions with a higher NDVI tended to have more forest fire points (Figure 2h). The forest fire incidence notably decreased in areas farther from roads and residential zones (Figures 2i,j).

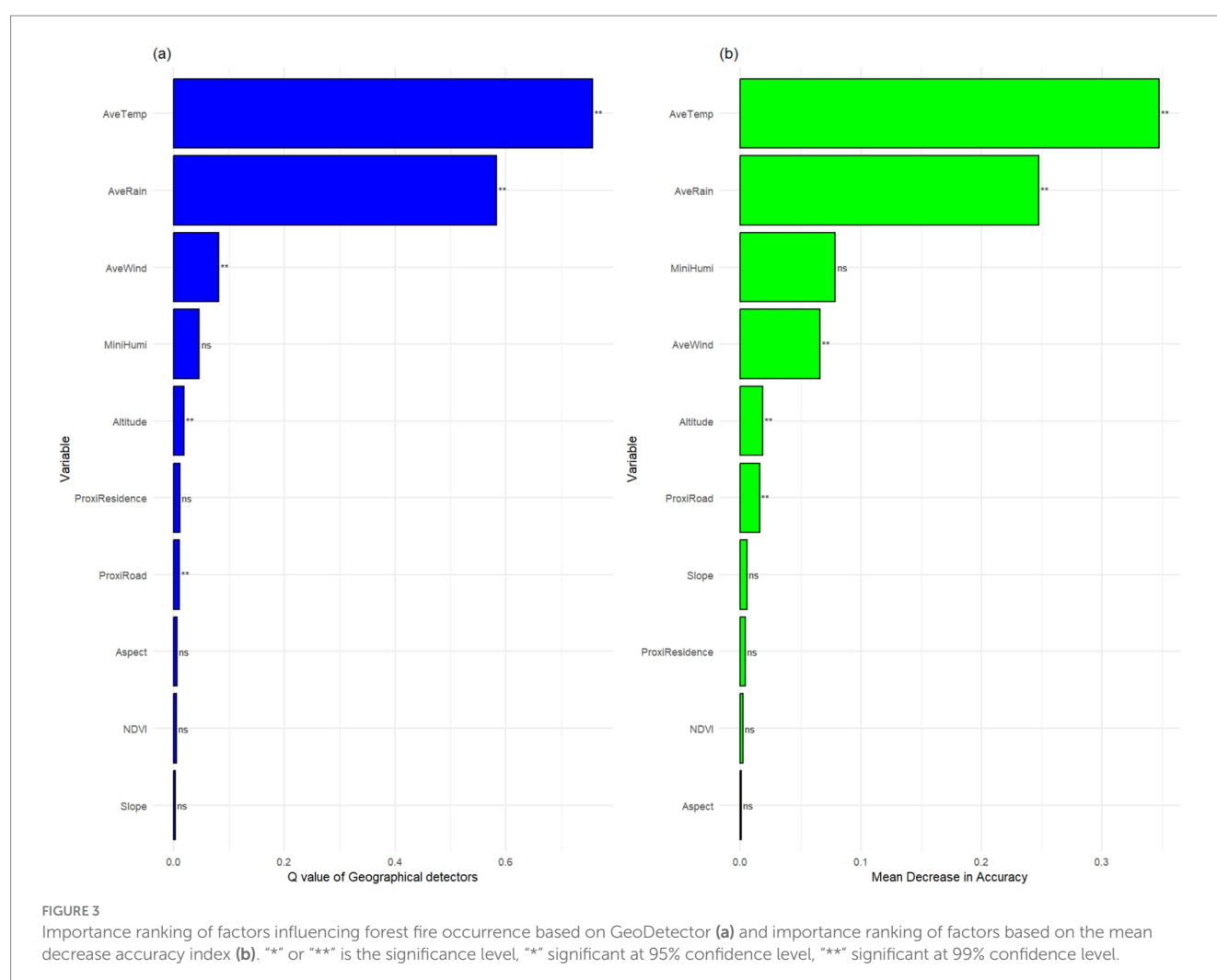
### 3.2 Analysis of driving factors for forest fire incidence

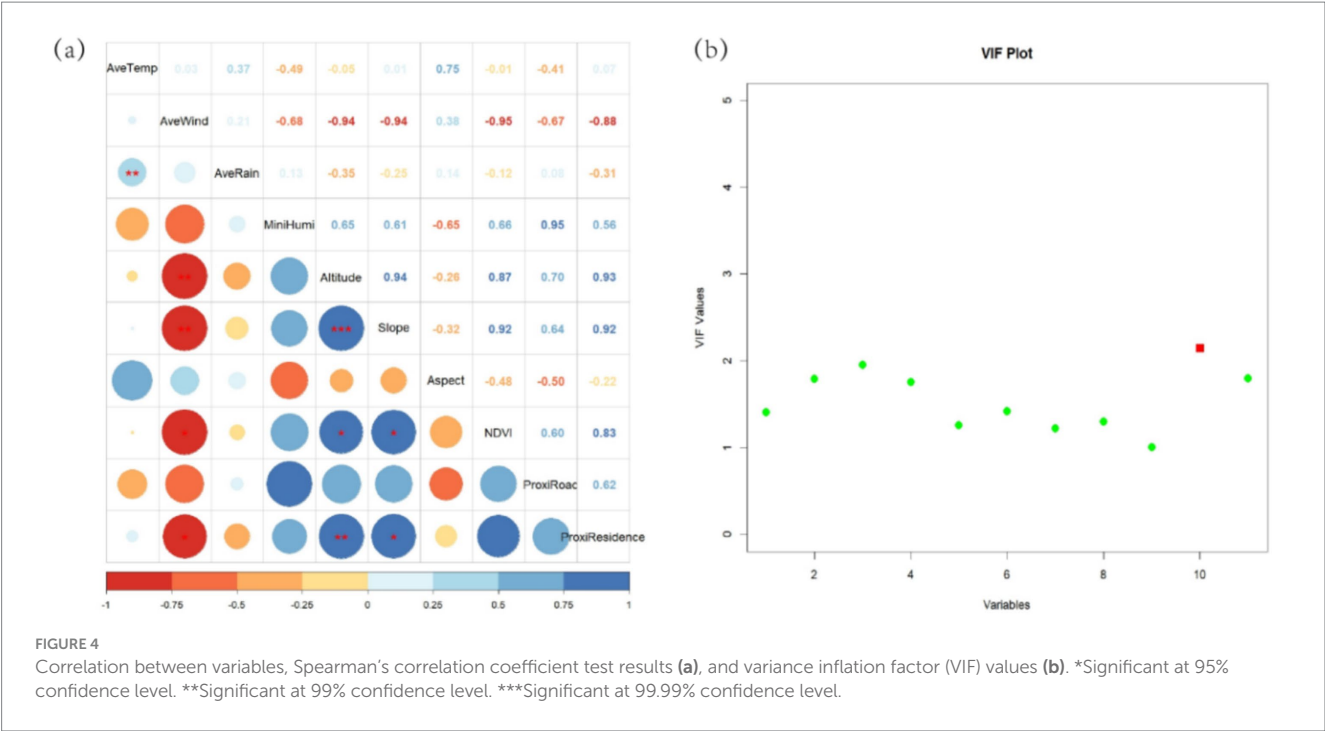
The factor detection function was applied to determine the relative impact of individual drivers on wildfire occurrence

(Figure 3a). Only the average temperature, average rainfall, average wind speed, proximity to roads, and altitude above sea level were found to significantly influence the occurrence of tropical forest fires, with the q-values ranked from largest to smallest. Among all factors, average temperature had the highest q-value ( $q = 0.4733$ ), followed by average rainfall ( $q = 0.4449$ ). The MDA index values (Figure 3b) further validated this finding. These results indicate that climatic factors are the primary drivers of the spatial patterns of tropical forest fires in the study area, particularly the average temperature and average rainfall.

### 3.3 Interrelationships among factors

Spearman's correlation analysis revealed that climate variables exhibited low correlation, while the altitude, slope, NDVI, proximity to roads, and proximity to residential areas were highly correlated (Figure 4a). In particular, the correlation between altitude and slope was the highest and most significant. However, all variables had VIF values below 10, even below 2, which indicated that there was no collinearity among these variables (Figure 4b).





3.4 Development of forest fire Hazard models using logistic regression and random forest

The fitting results of the logistic regression model showed that the annual average temperature, annual average wind speed, annual average rainfall, altitude, and proximity to roads were significant ( $p > 0.01$ ). The average annual temperature and proximity to roads were positively correlated with the number of forest fires, while the average annual wind speed, average annual rainfall, and altitude were negatively correlated (Table 2). The fire sensitivity maps developed based on the logistic regression and random forest models showed that the areas with high forest fire risk were primarily concentrated in southwestern Hainan Island (Figure 5).

3.5 Validation of model accuracy

The confusion matrix generated in this study indicated that the prediction accuracy of the logistic regression model was 82.79% (Figure 6a), while the prediction accuracy of the random forest model was 98.60% (Figure 6b). The ROC curve analysis showed that the AUC of the logistic regression model was 0.79 (Figure 7), which fell within an acceptable accuracy range. In contrast, the AUC of the random forest model was 0.99, indicating perfect classification. Although both models demonstrated reliability, the fire sensitivity map generated by the random forest model exhibited higher applicability for predicting forest fire occurrence probabilities on Hainan Island compared to the logistic regression model.

TABLE 2 Logistic regression model fitting results.

Variable name	Estimate	Significance
(Intercept)	−9.697	**
AveTemp	4.433	**
AveWind	−1.527	**
AveRain	−3.323	**
MiniHumi	5.218	ns
Altitude	−1.499	**
Slope	−8.313	ns
Aspect	−2.845	ns
NDVI	−1.926	ns
ProxiRoad	4.19	**
ProxiResidence	2.395	ns

“\*” or “\*\*” is the significance level, “\*” Significant at 95% confidence level, “\*\*” Significant at 99% confidence level. “ns” not significant.

4 Discussion

4.1 Determinants of tropical forest fires

We found that climate factors contribute the most to the spatial pattern of fire occurrence across the entire study area, particularly temperature and rainfall, which jointly and significantly drive the probability of tropical forest fires. This finding aligns with the majority of research in other tropical regions, such as the Amazon, Southeast Asia, and Africa (Cardil et al., 2023; Su et al., 2021; Trang et al., 2022; Xu et al., 2020), indicating that temperature and precipitation are the primary drivers of forest fires in most tropical regions. We further

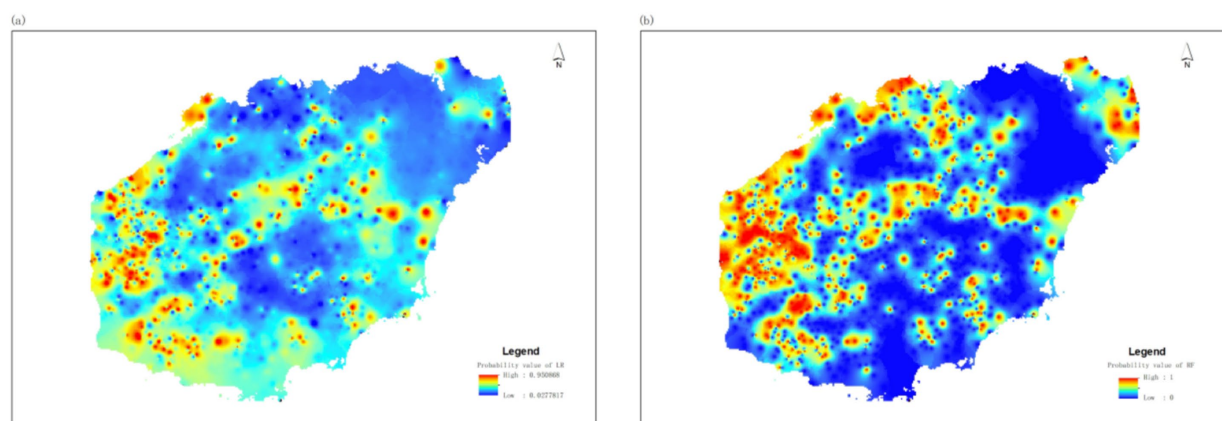


FIGURE 5  
Forest fire susceptibility map developed using the validation dataset based on the logistic regression model (a) and the random forest model (b).



FIGURE 6  
Confusion matrix of prediction results on the validation dataset based on the logistic regression model (a) and the random forest model (b).

found that rising temperatures and reduced rainfall significantly increase forest fire risk in the study area. This occurs because warmer and drier conditions substantially enhance fuel aridity in tropical forests, thereby creating more favorable conditions for fire ignition and spread. Similar conclusions have been drawn from studies in tropical forests such as the Brazilian Pantanal, northern Vietnam, and the Caribbean (Ribeiro et al., 2022; Trang et al., 2022; Van Beusekom et al., 2018). However, in tropical savannas, some studies have found

that increased drought and high temperatures do not necessarily lead to more fires (Alvarado et al., 2020). This may be attributed to the lower fuel loads in savannas compared to tropical forests, as drought and heat primarily influence fire frequency by enhancing fuel flammability. Our study revealed that when the mean annual temperature exceeds 24.5°C and the mean annual precipitation falls below 1,592.4 mm, the number of fire hotspots in the study area increases sharply. Therefore, the evidence presented in this study

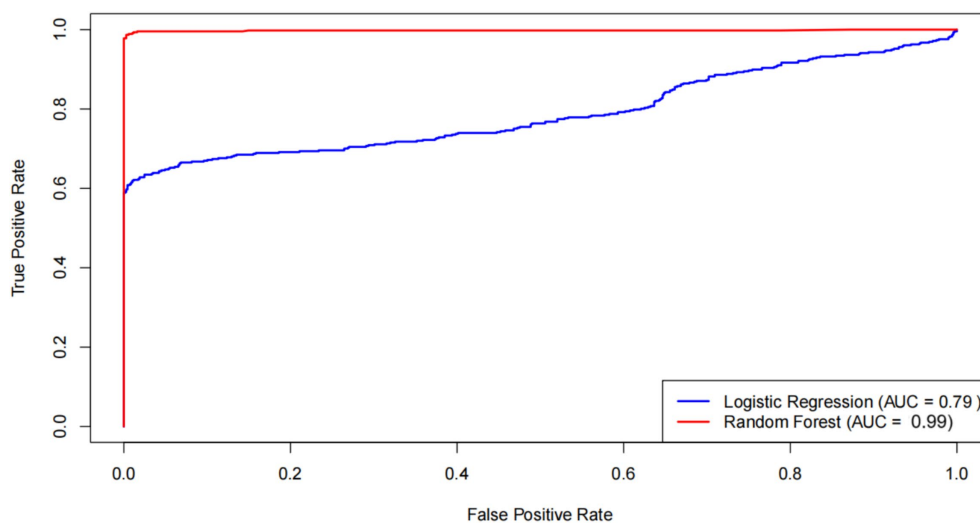


FIGURE 7

Receiver operating characteristic (ROC) curve of prediction results compared with actual values based on the logistic regression model and the random forest model.

suggests that the trends of global warming and drying may lead to more frequent tropical forest fires globally.

Although the results showed that human activities were not the primary drivers of forest fire occurrence, human activities still significantly affected the probability of tropical forest fires, albeit with a lower contribution compared to climatic factors. This finding is consistent with studies from neighboring tropical regions in southeast Asia (Trang et al., 2022; Trang et al., 2023a). This may be due to the higher population pressure and greater demand for forest land use in southeast Asian tropical regions, leading to more intense human disturbance (Gareth, 2019; Taufik, 2017). Surprisingly, the results of this study showed that vegetation factors (normalized difference vegetation index) did not significantly contribute to fire occurrence probability, which was inconsistent with findings from subtropical and temperate regions (Guo et al., 2020; Ma et al., 2020). The present study area is located in a tropical maritime monsoon climate zone, and although the biomass of tropical rainforests is abundant, this biomass only burns under conditions of water depletion, such as prolonged droughts (Trang et al., 2022). Moreover, the distribution maps (Supplementary Figures S2d, S3d) show that areas with higher NDVIs often correspond to regions with elevated air humidity, suggesting that fire occurrence is limited by fuel moisture.

## 4.2 Predictive model for tropical forest fire Hazard

Currently, there is a significant lack of development in tropical forest fire hazard prediction models, which severely hampers precise fire prevention and control efforts in these regions. Based on our identification of the primary drivers of tropical forest fires, we fitted logistic regression models and random forest models to predict the probability of forest fire occurrence. We found that both models demonstrated reliability in predicting tropical forest fire risks, but the random forest model achieved higher accuracy. This result is consistent with findings reported in studies from Vietnam (Truong

et al., 2023). In the present study, five factors—temperature, rainfall, wind speed, altitude, and proximity to roads—were significant at the 99% confidence level. Most studies on tropical forest fire hazard prediction models have also identified climate, topography, and human activities as important variables (Ngoc Thach et al., 2018; Truong et al., 2023). Although the two models showed some differences in prediction accuracy in the present study, both were able to effectively identify high-hazard areas for forest fires in tropical regions. These results are of significant practical importance for forest fire management in tropical areas. Identifying high-hazard zones will enable the development of more targeted fire prevention and control strategies, as well as the optimization of resource allocation.

## 4.3 Practical applications

This work provides a deeper understanding of the mechanisms underlying wildfire occurrence in tropical forests and supports wildfire prevention and mitigation efforts. First, by employing the logistic regression and random forest models we developed, relevant authorities can identify high-hazard areas for wildfires based on both past and current data and deploy disaster prevention and mitigation measures in advance. Currently, some wildfire management planners are hesitant to use machine learning models, as they are often considered “black box” models (Li et al., 2024). However, the present work also provides the most classic logistic regression model, which is easier for wildfire management planners to use. Second, this work provides an effective wildfire feature attribution method that reveals the driving factors of wildfires in tropical continental islands for the first time. This study explains the contribution of each feature to the occurrence of tropical forest fires and determines the specific numerical values of these contributions, which can be compared with values from other regions in the future. Understanding the driving factors of wildfires will be beneficial for the establishment of wildfire early warning systems. When these key driving factors reach critical thresholds, authorities should be alert and take proactive measures to

enhance monitoring. Furthermore, this study presents the fire hazard index of the study area in the form of a map. This will help with targeted resource planning and allocation within the region, allowing for the development of disaster prevention and mitigation strategies specific to certain areas.

## 5 Conclusion

This study integrated ground-based fire statistics and MODIS fire point data to construct a 20-year (2000–2020) forest fire dataset for Hainan Island. Using traditional logistic regression and emerging random forest models, we developed a tropical forest fire hazard model for Hainan Island and explored the main driving factors of forest fires in the region. The analysis leads to the following conclusions: (1) climate factors make the greatest contributions to the probability of tropical forest fire occurrence, followed by human activities and topography, while vegetation factors have no significant contribution; (2) temperature and rainfall dominate the probability of tropical forest fire occurrence, with rising temperatures and decreasing rainfall significantly increasing the risk of forest fires; and (3) both logistic regression and random forest models are reliable in predicting tropical forest fire risk, but the random forest model has greater applicability. The fire susceptibility maps generated using both models reveal high-risk areas in Hainan Island, providing a promising approach for local forest fire early warning and monitoring systems.

In summary, the findings indicate that the probability of tropical forest fires will increase under global warming and drought conditions. This study contributes to a deeper understanding of the factors driving tropical forest fires and provides theoretical support for the prevention and mitigation of wildfires in tropical regions worldwide.

## Data availability statement

The original contributions presented in the study are included in the article/[Supplementary material](#), further inquiries can be directed to the corresponding author.

## Author contributions

XC: Writing – original draft, Conceptualization, Formal analysis, Methodology. QY: Conceptualization, Project

administration, Resources, Writing – review & editing. ZC: Funding acquisition, Resources, Writing – review & editing. JL: Writing – review & editing. TW: Visualization, Writing – review & editing. YL: Software, Writing – review & editing. XP: Investigation, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by Hainan Provincial Natural Science Foundation of China (Grant No. 425RC845), Hainan Province Belongs to the Scientific Research Institute Technology Innovation Project (Grant No. SQKY2022-0021), and Forestry Department of Hainan Province.

## Acknowledgments

We thank them for providing the forest fire statistical data used in this research. We thank LetPub ([www.letpub.com.cn](http://www.letpub.com.cn)) for its linguistic assistance during the preparation of this manuscript.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/ffgc.2025.1495699/full#supplementary-material>

## References

- Abid, F. (2021). A survey of machine learning algorithms based Forest fires prediction and detection systems. *Fire. Technol.* 57, 559–590. doi: 10.1007/s10694-020-01056-z
- Alkhatib, R., Sahwan, W., Alkhatieb, A., and Schuett, B. (2023). A brief review of machine learning algorithms in Forest fires science. *Appl. Sci.* 13:8275. doi: 10.3390/app13148275
- Alvarado, S. T., Andela, N., Silva, T. S. F., and Archibald, S. (2020). Thresholds of fire response to moisture and fuel load differ between tropical savannas and grasslands across continents [Article]. *Glob. Ecol. Biogeogr.* 29, 331–344. doi: 10.1111/geb.13034
- Barmpoutis, P., Papaioannou, P., Dimitropoulos, K., and Grammalidis, N. (2020). A review on early Forest fire detection systems using optical remote sensing. *Sensors* 20:6442. doi: 10.3390/s20226442
- Brando, P. M., Paolucci, L., Ummenhofer, C. C., Ordway, E. M., Hartmann, H., Cattau, M. E., et al. (2019). Droughts, Wildfires, and Forest Carbon Cycling: A Pantropical Synthesis. In R. Jeanloz and K. H. Freeman (Eds.), *Annual Review of Earth and Planetary Sciences*, Vol 47 (pp. 555–581). doi: 10.1146/annurev-earth-082517-010235
- Cardil, A., Rodrigues, M., Tapia, M., Barbero, R., Ramirez, J., Stoof, C. R., et al. (2023). Climate teleconnections modulate global burned area. *Nature Commun.* 14:427. doi: 10.1038/s41467-023-36052-8
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating

- their performance. *Ecography* 36, 27–46. doi: 10.1111/j.1600-0587.2012.07348.x
- Dwomoh, F. K., Wimberly, M. C., Cochrane, M. A., and Numata, I. (2019). Forest degradation promotes fire during drought in moist tropical forests of Ghana. *For. Ecol. Manag.* 440, 158–168. doi: 10.1016/j.foreco.2019.03.014
- Eskandari, S., Miesel, J. R., and Pourghasemi, H. R. (2020). The temporal and spatial relationships between climatic parameters and fire occurrence in northeastern Iran. *Ecol. Indic.* 118:106720. doi: 10.1016/j.ecolind.2020.106720
- Fonseca, M. G., Anderson, L. O., Arai, E., Shimabukuro, Y. E., Xaud, H. A. M., Xaud, M. R., et al. (2017). Climatic and anthropogenic drivers of northern Amazon fires during the 2015–2016 El Niño event. *Ecol. Appl.* 27, 2514–2527. doi: 10.1002/eap.1628
- Gao, B., Shan, Y., Liu, X., Yin, S., Yu, B., Cui, C., et al. (2024). Prediction and driving factors of forest fire occurrence in Jilin Province, China [Article]. *J. For. Res.* 35:21. doi: 10.1007/s11676-023-01663-w
- Gareth, T. (2019). New satellite observations to improve monitoring and forecasting of severe smoke pollution over SE Asia caused by Indonesian landscape burning [awarded Grant].
- Gatti, L. V., Basso, L. S., Miller, J. B., Gloor, M., Domingues, L. G., Cassol, H. L. G., et al. (2021). Amazonia as a carbon source linked to deforestation and climate change. *Nature* 595:388. doi: 10.1038/s41586-021-03629-6
- Giglio, L., Randerson, J. T., and van der Werf, G. R. (2013). Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4). *J. Geophys. Res. Bioge.* 118, 317–328. doi: 10.1002/jgrg.20042
- Goss, M., Swain, D. L., Abatzoglou, J. T., Sarhadi, A., Kolden, C. A., Williams, A. P., et al. (2020). Climate change is increasing the likelihood of extreme autumn wildfire conditions across California. *Environ. Res. Lett.* 15:94016. doi: 10.1088/1748-9326/ab83a7
- Guo, F., Wang, G., Innes, J. L., Ma, X., Sun, L., and Hu, H. (2015). Gamma generalized linear model to investigate the effects of climate variables on the area burned by forest fire in Northeast China. *J. For. Res.* 26, 545–555. doi: 10.1007/s11676-015-0084-2
- Guo, X.-Y., Zhang, H.-Y., Wang, Y.-Q., Zhao, J.-J., and Zhang, Z.-X. (2020). The driving factors and their interactions of fire occurrence in greater Khingan Mountains, China. *J. Mountain Sci.* 17, 2674–2690. doi: 10.1007/s11629-020-6036-0
- Haas, O., Keeping, T., Gomez-Dans, J., Prentice, I. C., and Harrison, S. P. (2024). The global drivers of wildfire. *Front. Environ. Sci.* 12:1438262. doi: 10.3389/fevs.2024.1438262
- Hosmer, D. W. Jr., Lemeshow, S., and Sturdivant, R. X. (2013). Applied logistic regression. Hoboken, NJ: John Wiley & Sons.
- Huang, C., Hu, L., Jiang, Y., Xu, Y., He, J., Lin, S., et al. (2021). A 150-year avian bio-inventory on a global biodiversity hotspot island. *Global Ecol. Conserv.* 27:e01578. doi: 10.1016/j.gecco.2021.e01578
- Jiang, Y., Zang, R., Letcher, S. G., Ding, Y., Huang, Y., Lu, X., et al. (2016). Associations between plant composition/diversity and the abiotic environment across six vegetation types in a biodiversity hotspot of Hainan Island, China. *Plant and Soil* 403, 21–35. doi: 10.1007/s11104-015-2723-y
- Juárez-Orozco, S. M., Siebe, C., and Fernández, D. F. Y. (2017). Causes and effects of Forest fires in tropical rainforests: a bibliometric approach. *Trop. Conserv. Sci.* 10, 1–14. doi: 10.1177/1940082917737207
- Kouassi, J.-L., Wandan, N., and Mbow, C. (2020). Predictive modeling of wildfire occurrence and damage in a tropical savanna ecosystem of West Africa. *Fire* 3:42. doi: 10.3390/fire3030042
- Lewis, S. L., Edwards, D. P., and Galbraith, D. (2015). Increasing human dominance of tropical forests. *Science* 349, 827–832. doi: 10.1126/science.aaa9932
- Li, H., Vulova, S., Rocha, A. D., and Kleinschmit, B. (2024). Spatio-temporal feature attribution of European summer wildfires with explainable artificial intelligence (XAI). *Sci. Total Environ.* 916:170330. doi: 10.1016/j.scitotenv.2024.170330
- Ma, W., Feng, Z., Cheng, Z., Chen, S., and Wang, F. (2020). Identifying Forest fire driving factors and related impacts in China using random Forest algorithm. *Forests* 11:507. doi: 10.3390/f11050507
- Miettinen, J., Hyer, E., Chia, A. S., Kwok, L. K., and Liew, S. C. (2013). Detection of vegetation fires and burnt areas by remote sensing in insular Southeast Asian conditions: current status of knowledge and future challenges. *Int. J. Remote Sens.* 34, 4344–4366. doi: 10.1080/01431161.2013.777489
- Milanovic, S., Markovic, N., Pamucar, D., Gigovic, L., Kostic, P., and Milanovic, S. D. (2021). Forest Fire Probability Mapping in Eastern Serbia: Logistic Regression versus Random Forest Method [Article]. *Forests*, 12:5. doi: 10.3390/f12010005
- Mohammadi, F., Bavaghar, M. P., and Shabani, N. (2014). Forest fire risk zone modeling using logistic regression and GIS: an Iranian case study. *Small Scale Forestry* 13, 117–125. doi: 10.1007/s11842-013-9244-4
- Ngoc Thach, N., Bao-Toan Ngo, D., Xuan-Canh, P., Hong-Thi, N., Hang Thi, B., Nhat-Duc, H., et al. (2018). Spatial pattern assessment of tropical forest fire danger at Thuan Chau area (Vietnam) using GIS-based advanced machine learning algorithms: a comparative study. *Eco. Inform.* 46, 74–85. doi: 10.1016/j.ecoinf.2018.05.009
- Perrakis, D. D. B., Cruz, M. G., Alexander, M. E., Hanes, C. C., Thompson, D. K., Taylor, S. W., et al. (2023). Improved logistic models of crown fire probability in Canadian conifer forests [Article]. *Int. J. Wildland Fire*. Wf23074. doi: 10.1071/wf23074
- Ribeiro, A. F. S., Brando, P. M., Santos, L., Rattis, L., Hirschi, M., Hauser, M., et al. (2022). A compound event-oriented framework to tropical fire risk assessment in a changing climate. *Environ. Res. Lett.* 17:065015. doi: 10.1088/1748-9326/ac7342
- Su, Z., Zheng, L., Luo, S., Tigabu, M., and Guo, F. (2021). Modeling wildfire drivers in Chinese tropical forest ecosystems using global logistic regression and geographically weighted logistic regression. *Nat. Hazards* 108, 1317–1345. doi: 10.1007/s11069-021-04733-6
- Sze, J. S., Jefferson, and Lee, J. S. H. (2019). Evaluating the social and environmental factors behind the 2015 extreme fire event in Sumatra, Indonesia. *Environ. Res. Lett.* 14:15001. doi: 10.1088/1748-9326/aaee1d
- Szapkowski, D. M., and Jensen, J. L. R. (2019). A review of the applications of remote sensing in fire ecology. *Remote Sens.* 11:2638. doi: 10.3390/rs11222638
- Taufik, M. (2017). Hydrological drought and wildfire in the humid tropics. Wageningen: Wageningen University.
- Tehrany, M. S., Jones, S., Shabani, F., Martinez-Alvarez, F., and Dieu Tien, B. (2019). A novel ensemble modeling approach for the spatial prediction of tropical forest fire susceptibility using LogitBoost machine learning classifier and multi-source geospatial data [Article]. *Theor. Appl. Climatol.* 137, 637–653. doi: 10.1007/s00704-018-2628-9
- Trang, P. T., Andrew, M. E., Chu, T., and Enright, N. J. (2022). Forest fire and its key drivers in the tropical forests of northern Vietnam. *Int. J. Wildland Fire* 31, 213–229. doi: 10.1071/wf21078
- Trang, P. T., Andrew, M. E., and Enright, N. J. (2023a). Burn severity and proximity to undisturbed forest drive post-fire recovery in the tropical montane forests of northern Vietnam. *Fire Ecology* 19:47. doi: 10.1186/s42408-023-00205-3
- Trang, P. T., Enright, N. J., Chu, T., and Andrew, M. E. (2023b). Drivers of wildfire burn severity in the montane rainforests of northern Vietnam. *Int. J. Wildland Fire* 32, 1773–1787. doi: 10.1071/wf23026
- Truong, T. X., Nhu, V.-H., Phuong, D. T. N., Nghi, L. T., Hung, N. N., Hoa, P. V., et al. (2023). A new approach based on TensorFlow deep neural networks with ADAM optimizer and GIS for spatial prediction of Forest fire danger in tropical areas. *Remote Sens.* 15:3458. doi: 10.3390/rs15143458
- Van Beusekom, A. E., Gould, W. A., Monmany, A. C., Khalyani, A. H., Quiñones, M., Fain, S. J., et al. (2018). Fire weather and likelihood: characterizing climate space for fire occurrence and extent in Puerto Rico. *Clim. Chang.* 146, 117–131. doi: 10.1007/s10584-017-2045-6
- Wang, J.-F., Li, X.-H., Christakos, G., Liao, Y.-L., Zhang, T., Gu, X., et al. (2010). Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun region, China. *Int. J. Geogr. Inf. Sci.* 24, 107–127. doi: 10.1080/13658810802443457
- Wang, W., Zhao, F., Wang, Y., Huang, X., and Ye, J. (2023). Seasonal differences in the spatial patterns of wildfire drivers and susceptibility in the southwest mountains of China. *Sci. Total Environ.* 869:161782. doi: 10.1016/j.scitotenv.2023.161782
- Withey, K., Berenguer, E., Palmeira, A. F., Espírito-Santo, F. D. B., Lennox, G. D., Silva, C. V. J., et al. (2018). Quantifying immediate carbon emissions from El Niño-mediated wildfires in humid tropical forests. *Philos. Trans. Royal Soc. B* 373, 373:20170312. doi: 10.1098/rstb.2017.0312
- Xiao, Y., Gu, X., Yin, S., Shao, J., Cui, Y., Zhang, Q., et al. (2016). Geostatistical interpolation model selection based on ArcGIS and spatio-temporal variability analysis of groundwater level in piedmont plains, northwest China [Article]. *Springerplus*, 5:425. doi: 10.1186/s40064-016-2073-0
- Xu, X. Y., Jia, G. S., Zhang, X. Y., Riley, W. J., and Xue, Y. (2020). Climate regime shift and forest loss amplify fire in Amazonian forests. *Glob. Chang. Biol.* 26, 5874–5885. doi: 10.1111/gcb.15279