



## Article

# Machine Learning for Predicting Forest Fire Occurrence in Changsha: An Innovative Investigation into the Introduction of a Forest Fuel Factor

Xin Wu <sup>1,2</sup> , Gui Zhang <sup>1,2,\*</sup>, Zhigao Yang <sup>1,2</sup>, Sanqing Tan <sup>1,2</sup>, Yongke Yang <sup>1,2</sup> and Ziheng Pang <sup>1,2</sup>

<sup>1</sup> School of Forestry, Central South University of Forestry and Technology, Changsha 410004, China; csuftfire@csuft.edu.cn (X.W.); zgyang@csuft.edu.cn (Z.Y.); t19920990@csuft.edu.cn (S.T.); yky@csuft.edu.cn (Y.Y.); pang\_ziheng@csuft.edu.cn (Z.P.)

<sup>2</sup> Yuelushan Laboratory Carbon Sinks Forests Variety Innovation Center, Changsha 410004, China

\* Correspondence: zgui@csuft.edu.cn

**Abstract:** Affected by global warming and increased extreme weather, Hunan Province saw a phased and concentrated outbreak of forest fires in 2022, causing significant damage and impact. Predicting the occurrence of forest fires can enhance the ability to make early predictions and strengthen early warning and responses. Currently, fire prevention and extinguishing in China's forests and grasslands face severe challenges due to the overlapping of natural and social factors. Existing forest fire occurrence prediction models mostly take into account vegetation, topographic, meteorological and human activity factors; however, the occurrence of forest fires is closely related to the forest fuel moisture content. In this study, the traditional driving factors of forest fire such as satellite hotspots, vegetation, meteorology, topography and human activities from 2004 to 2021 were introduced along with forest fuel factors (vegetation canopy water content and evapotranspiration from the top of the vegetation canopy), and a database of factors for predicting forest fire occurrence was constructed. And a forest fire occurrence prediction model was built using machine learning methods such as the Random Forest model (RF), the Gradient Boosting Decision Tree model (GBDT) and the Adaptive Augmentation Model (AdaBoost). The accuracy of the models was verified using Area Under Curve (AUC) and four other metrics. The RF model with an AUC value of 0.981 was more accurate than all other models in predicting the probability of forest fire occurrence, followed by the GBDT (AUC = 0.978) and AdaBoost (AUC = 0.891) models. The RF model, which has the best accuracy, was selected to predict the monthly forest fire probability in Changsha in 2022 and combined with the Inverse Distance Weight Interpolation method to plot the monthly forest fire probability in Changsha. We found that the monthly spatial and temporal distribution of forest fire probability in Changsha varied significantly, with March, April, May, September, October, November and December being the months with higher forest fire probability. The highest probability of forest fires occurred in the central and northern regions. In this study, the core drivers affecting the occurrence of forest fires in Changsha City were found to be vegetation canopy evapotranspiration and vegetation canopy water content. The RF model was identified as a more suitable forest fire occurrence probability prediction model for Changsha City. Meanwhile, this study found that vegetation characteristics and combustible factors have more influence on forest fire occurrence in Changsha City than meteorological factors, and surface temperature has less influence on forest fire occurrence in Changsha City.

**Keywords:** early warning of forest fire risk; random forest; live fuel moisture content; forest fire drivers



**Citation:** Wu, X.; Zhang, G.; Yang, Z.; Tan, S.; Yang, Y.; Pang, Z. Machine Learning for Predicting Forest Fire Occurrence in Changsha: An Innovative Investigation into the Introduction of a Forest Fuel Factor. *Remote Sens.* **2023**, *15*, 4208. <https://doi.org/10.3390/rs15174208>

Academic Editor: Quazi K. Hassan

Received: 28 June 2023

Revised: 16 August 2023

Accepted: 24 August 2023

Published: 27 August 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In recent years, extreme weather events such as high temperatures, droughts, high winds and dry thunderstorms have been frequent, leading to the proliferation and reoccurrence of forest fires worldwide, with fire mega disasters that have shocked the world [1].

China is facing the same disaster situation and risk, and last summer and autumn saw a succession of phased concentrations of forest fires in Chongqing, Hunan, Guangxi and Jiangxi, indicating that China has also entered this trending cycle. Extreme weather events exacerbate transpiration in the vegetation canopy and reduce the live fuel moisture content (LFMC), creating excellent conditions for forest fires to occur [2].

It is important to understand the probability of forest fires in real time and on a large scale to provide an objective understanding of the level of forest fire risk and to provide a scientific basis for decision making to effectively prevent forest fires [3]. There are many drivers of forest fires, including fuel moisture content, meteorological, topographic and human activity factors, etc. [4]. Existing forest fire prediction models mostly take into account vegetation, topographic, meteorological and human activity factors, but the occurrence of forest fires is closely related to the amount of forest fuels [2,5,6]. However, the occurrence of forest fires is closely related to the live fuel moisture content and evaporation from the top of forest vegetation canopy before the occurrence of fires [7]. It is important to accurately obtain the live fuel moisture content and evaporation from the top of the forest vegetation canopy before a fire occurs and add them to the database of forest fire drivers [8]. It is also important to quantify the degree of influence of different drivers on the occurrence of forest fires, identify the main drivers affecting the occurrence of forest fires and build a highly accurate forest fire occurrence prediction model [9].

Fuel moisture content is an important driver of combustible ignition and forest fire spread rates [1,10,11]. It is usually divided into the combustible moisture content of growing vegetation and the combustible moisture content of dead vegetation [12–14]. Previous studies have demonstrated that the frequency of forest fires and the area burned tend to increase as the moisture content of combustible material decreases [15,16]. This is due to the fact that forest fuels with high moisture content requires more energy to evaporate the water, reducing the probability of forest fires and their rate of spread [17].

Topography is an important parameter for topographic analysis of forest fires by affecting airflow and local microclimate, changing the spatial distribution of forest fuel and influencing the occurrence of forest fires. In the study of the correlation between forest fires and topography, the influence of topography on forest fire occurrence was obtained based on a comprehensive analysis of several factors such as topography, climate and vegetation on forest fire occurrence and development [18,19]. Fang et al. found that the topography factor explained 29.2% of the heterogeneity in the spatial distribution of fire intensity through an enhanced regression number model and concluded that vegetation and topography had a greater influence on fire intensity [20]. Kong et al. suggested that changes in elevation, slope and slope orientation lead to changes in regional surface moisture content and temperature and that such changes affect the spatial distribution of surface vegetation and decomposition of surface deadfall, which in turn affects forest fire occurrence [21]. Climatic factors affect the occurrence of forest fires by influencing the flammability of forest fuel and the amount of combustible material accumulated on the ground. Some studies have demonstrated that surface temperature and air temperature are positively correlated, while there is a strong correlation between surface temperature and forest fire occurrence [22,23], and the spatial distribution pattern of forest fire events changes with inter-annual fluctuations in precipitation and air temperature [24]. Persistent high temperatures cause surface soil moisture and combustible moisture content to decrease rapidly, and surface combustibles continue to accumulate, providing a material basis for forest fires to occur. Wind can trigger changes in the direction and speed of forest fire spread, and sustained strong winds can also lead to evaporation of water, resulting in a decrease in the fuel moisture content. Studies have shown that as the average annual temperature rises, the effect of average annual temperature on the overfire area tends to increase and then decrease [25], and the effect of average annual precipitation on the overfire area is the same as the average annual temperature [26].

As the intensity and frequency of forest fires increase, their impacts worldwide are elevating at an alarming rate [27]. It is important to analyse the process of the occurrence

and development of forest fire activities under the influence of forest fire drivers and their potential spatial interactions and feedback mechanisms in different environmental contexts [28,29]. Modelling and outputting the probability of forest fire occurrence in an explicit spatial form and constructing early warning models for forest fire occurrence prediction can provide new insights in response to ongoing climate change and widespread human activities [6]. Since the 1990s, regression models (e.g., linear regression and logistic regression) have been widely used in forest fire probability modelling. And as an alternative to regression models, binary statistical models have gradually started to be applied in the field of forest fire prediction, such as frequency ratio methods (FR), weights of evidence, deterministic factors and evidence-based belief functions (EBF) [30]. However, it has been noted that these models are very sensitive to the quality of the input data and often mask the true relationship between forest fires and their drivers [31]. Artificial intelligence (AI) has been on the rise in recent years and has also proven to be efficient and accurate in the field of predicting natural hazards [32,33]. Among these AI methods, Random Forest (RF), Artificial Neural Network (ANN), Gradient Boosting Decision Tree (GBDT) and AdaBoost have been shown to outperform traditional statistical methods in forest fire modelling and its applications [5,6,25,28,34–43]. Forest fire predictive modelling, by linking the occurrence of forest fires to changes in the variables that drive them, such as climate, fuels, terrain and even human activity, has become an important part of the field and promises to improve the success of forest fire prevention and control and commanded suppression [44,45]. In addition, another advantage of AI methods is their ability to integrate seamlessly with many other methods, thereby increasing the level of model performance [46]. Similarly, AI methods in forest fire prediction modelling can provide detailed information such as remote sensing image recognition, correlation analysis, and spatial pattern recognition of fire occurrence, which can be used as input parameters for key components such as the construction of forest fire risk early warning models.

Recent comparative studies have addressed traditional regression models, which assume a linear relationship between forest fire occurrence and its drivers [47]. However, such models are unable to accurately describe the currently widely accepted complex non-linear relationship between drivers and forest fires on spatial and temporal scales, making it difficult to accurately assess the risk of forest fire occurrence [48]. Fuel moisture content is one of the important factors influencing forest fire occurrence, and for the selection of forest fire drivers, some recent studies have not included it in forest fire prediction models [5,6,49]. At the same time, most of the related studies have made risk zoning of regional forest fires on a quarterly basis, while the monthly divergence of fires in the study area is obvious, and the seasonal changes are difficult to guide specific forest fire prevention and suppression management [5,41]. While previous studies have constructed forest fire prediction models, no substantial short-term predictions have been carried out in a particular region [5]. The thematic maps produced for risk zoning did not utilise land classes to filter the interpolation results, resulting in the generation of high-risk zones for the occurrence of forest fires falling within Dongting Lake in Hunan Province, a rather obvious loophole [5,6].

To address these issues, this study was based on data from 20,269 forest fire hotspots in Hunan Province from 2004 to 2021. While integrating the traditional forest fire driving factors such as meteorology, topography, human activities and vegetation cover, we innovatively introduced combustible-related factors such as vegetation canopy water content, forest vegetation canopy evapotranspiration and soil surface water content to construct a database of forest fire driving factors. By screening the important forest fire driving factors, three machine learning methods, namely AdaBoost, GBDT and RF, were used to construct a forest fire prediction model. The optimal model is selected based on the results of model accuracy evaluation, and the month-by-month forest fire occurrence probability prediction map for 2022 is drawn using Changsha City as an application case. An objective understanding of the risk level of forest fire occurrence in Changsha City in the short term is of great significance for the scientific allocation of fire prevention resources and effective forest fire prevention work [50].

Therefore, the objectives of this study were (1) identification of the main driving factors influencing the occurrence of forest fires in Hunan Province; (2) construction of a forest fire occurrence prediction model based on AdaBoost, GBDT and RF; (3) mapping the risk of forest fires in Changsha by month in 2022.

## 2. Materials and Methods

### 2.1. Study Area

Hunan Province is located between longitude 108°47'–114°15'E and latitude 24°38'–30°08'N, with a horseshoe-shaped topography surrounded by mountains on three sides and opening towards the north, and a subtropical monsoon climate with simultaneous rain and heat [5,51,52], as shown in Figure 1. As of 2022, the forest coverage rate of Hunan Province reaches 59.98%, the forest storage volume reaches 664 million cubic metres and the forest fire damage rate is controlled at 0.129%. Changsha, the capital of Hunan Province, is located in the north-eastern part of Hunan Province and belongs to the same subtropical monsoon zone with the same climatic background [53]. As of 2022, Changsha's forest coverage rate is stable at 55%, ranking in the top three of China's provincial capital cities, and the forest storage volume reaches 30.86 million m<sup>3</sup>, a year-on-year increase of 4.05%, or 1.2 million m<sup>3</sup>. Changsha has conducted a great deal of publicity and education work in promoting the construction of ecological civilisation and the management of forest fire prevention and suppression, etc. In 2022, Changsha's GDP was CNY 139.611 billion, an increase of 4.5% year-on-year, with a population inflow of 181.3 million people, which is ranked first in China [54]. Due to the undulating terrain, high forest cover, high combustible load, frequent anthropogenic activities and low rainfall in autumn and winter resulting in low water content of combustible materials, there are significant monthly differences in the frequency of forest fires in the region, as shown in Figure 2. Especially during the local forest fire protection period (October to May), forest fire risk warning and control need to be particularly strengthened.

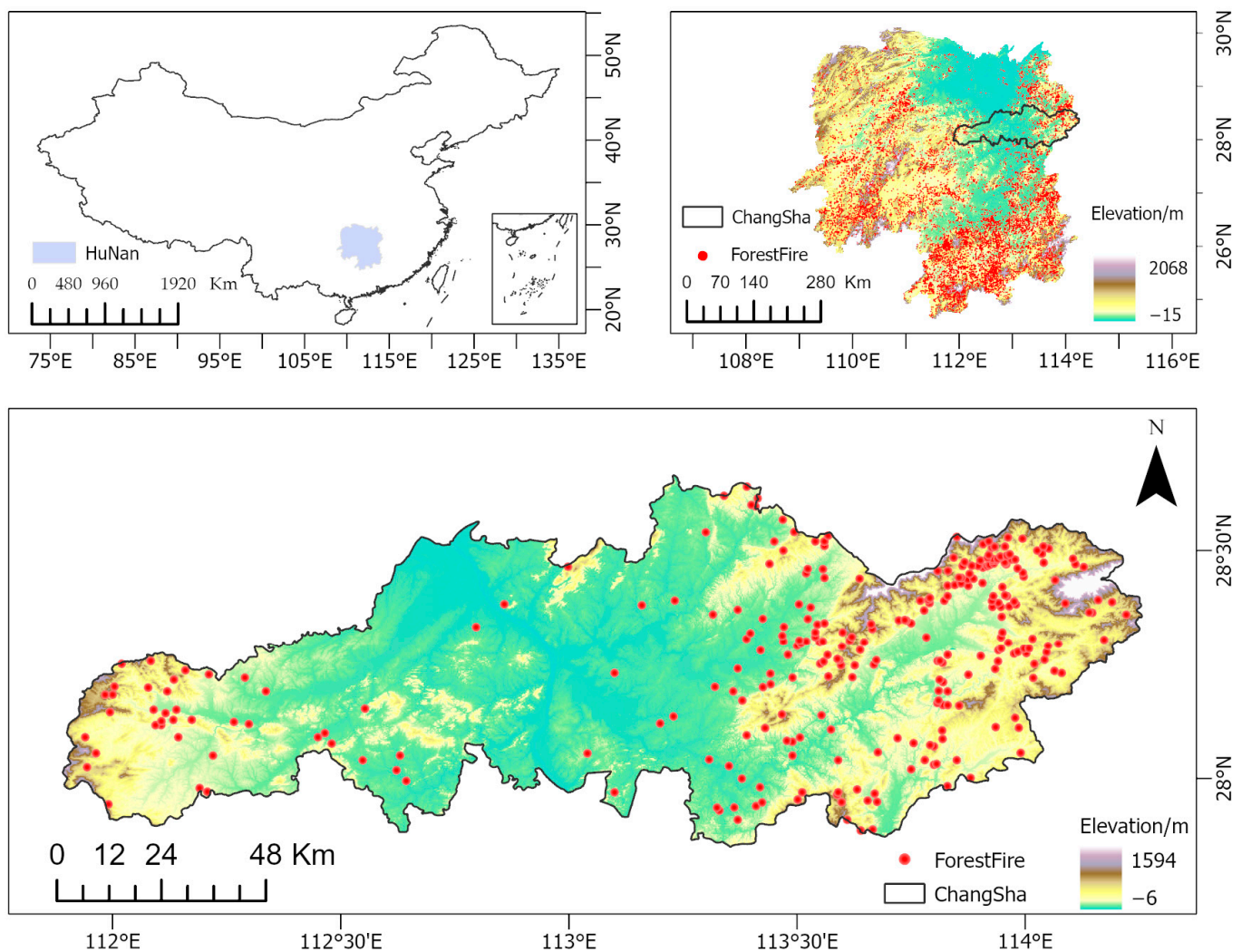
### 2.2. Data Sources and Pre-Processing

#### 2.2.1. Fire Point Data

The forest fire satellite hotspot data used in this study were obtained from the 2004–2021 satellite monitoring hotspot database provided by the Forest Fire Early Warning and Monitoring Information Centre of the Ministry of Emergency Management of China. This database records a large amount of attribute information of satellite hotspots, such as longitude, latitude, time of occurrence, land type and the type of hotspots returned from field surveys [5,6]. As shown in Figure 3, the number of historical forest fires in Hunan Province has an obvious monthly divergence pattern, and the current research related to forest fire occurrence trend prediction is mostly based on quarterly analysis, which has certain limitations [5]. Generally speaking, the special protection period for forest fires in Hunan Province is from October to May of the following year, but the frequency of forest fires in different months also varies to some extent. To ensure data quality, this study first screened satellite hotspot data for land types of forested land based on the type of hotspot feedback from field surveys and removed anomalous samples from the original dataset, such as attribute missing data and duplicate monitoring of satellite hotspot samples, to create fire spot data for Hunan Province. Then, based on the fire point data of Hunan Province from 2004 to 2021, random non-fire points were created in the ratio of 1:1 between the number of fire points and non-fire points within the forest land coverage of Hunan Province as data samples without fires, and the generated random non-fire points were ensured to have both temporal and spatial randomness [4,41]. It should be noted that since the spatial resolution of the thermal infrared band of the MODIS, NOAA and FY3 series satellite images used for satellite hotspot data extraction is 1 km, and the latitude and longitude of the acquired satellite hotspots are the latitude and longitude information of the central image element of the images, to ensure that the random non-fire points do not overlap spatially with the fire points, a buffer zone analysis is performed on the



fire points, and a 500 m fire point range is created for each fire point. A buffer zone was created for each fire point, and random points that fell into the buffer zone were removed to ensure the accuracy of the data. Finally, a new ForestFire attribute field was added to the satellite monitoring hotspot database, setting the attribute value of real fire points to 1 and the attribute value of random non-fire points to 0. The forest land cover data of Hunan Province used in this study to create the random non-fire point data were obtained from the global ground cover dataset GlobeLand30 with a spatial resolution of 30 m for extraction (<http://www.globallandcover.com/> accessed on 31 December 2022) [5,6]. The final number of fire and random points obtained for Hunan Province was 20,269, as shown in Figure 4.

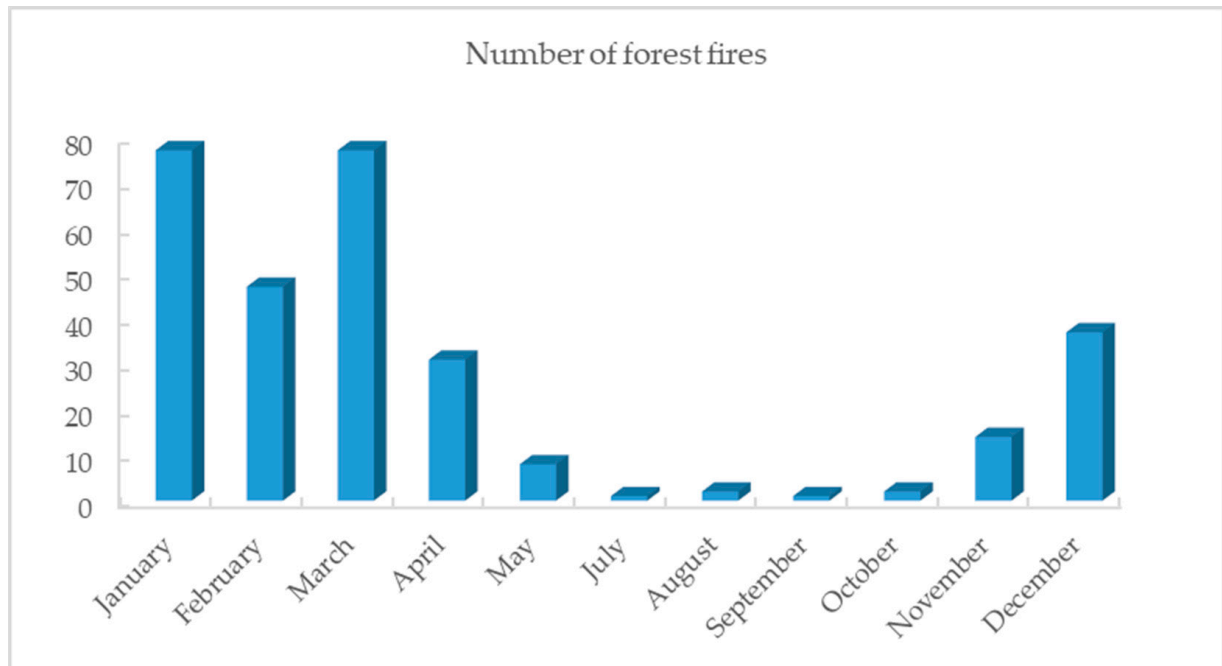


**Figure 1.** The top left corner indicates the geographical location of Hunan in China; the top right corner indicates the geographical location of Changsha in Hunan; the map below indicates the topographical distribution of Changsha in the study area; the red circles indicate the spatial distribution of satellite monitoring hotspots from 2004 to 2021.

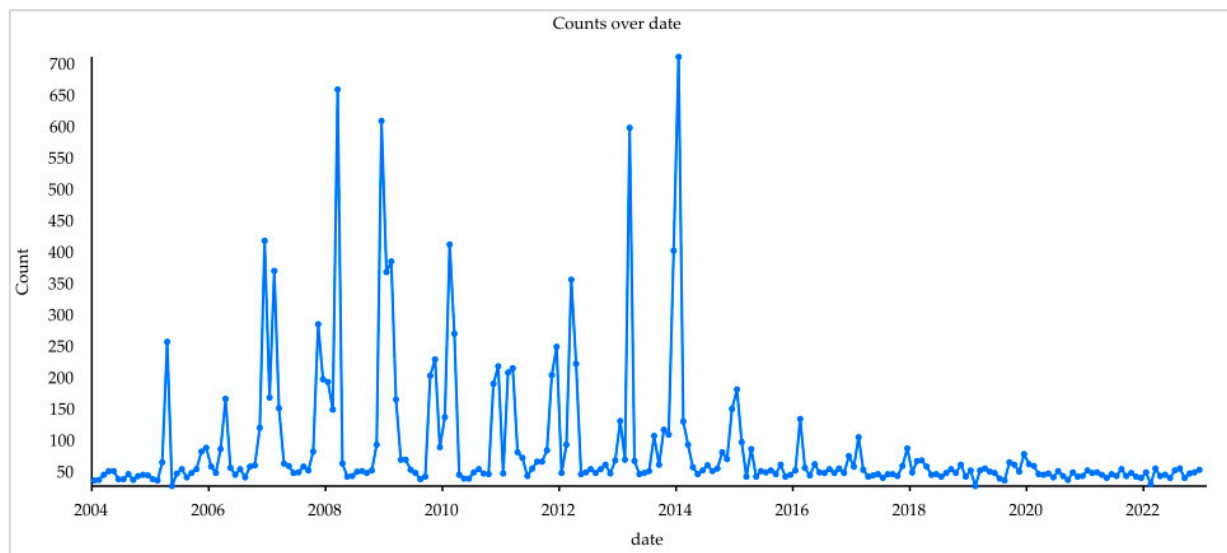
### 2.2.2. Data on Forest Fuels and Vegetation

Forest fuels data are an important factor in the occurrence and development of fires, and vegetation canopy moisture content and vegetation canopy evaporation, which characterise the state of living combustible material, are also important indicators in assessing fire risk ratings [52,55]. Vegetation canopy moisture content refers to the moisture content of the leaves, branches and other parts of vegetation and is an important indicator for assessing the flammability of vegetation to fire. Vegetation canopy evapotranspiration

is the amount of water consumed by vegetation for transpiration, which can influence the degree of vegetation dieback and the fire risk rating. These two parameters can be estimated from the `skin_reservoir_content` and `evaporation_from_the_top_of_canopy_sum` provided in the ERA5-Land dataset, and these data can effectively represent the forest fuel moisture content, which may have a significant impact on the prediction of forest fire occurrence [56].

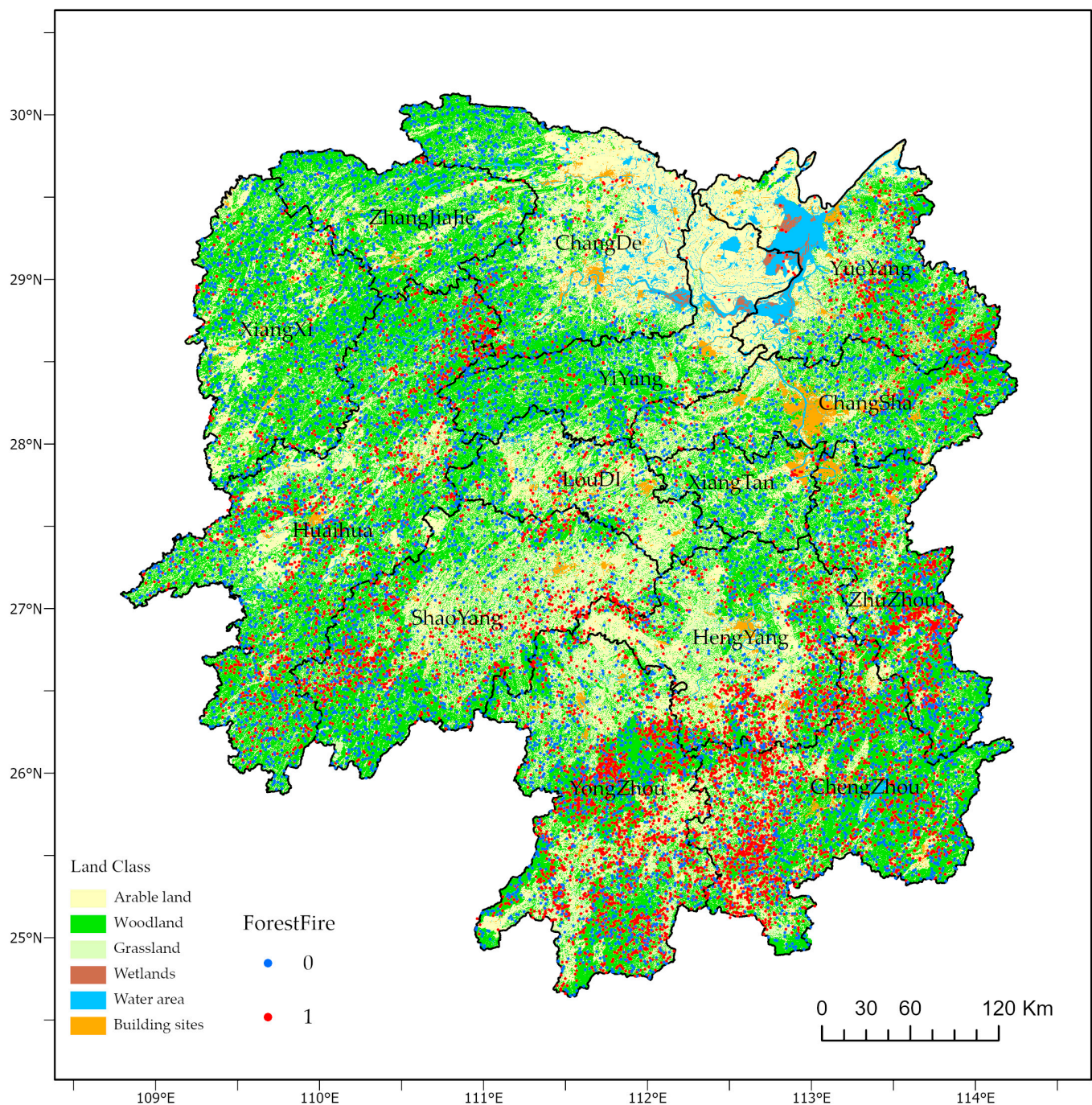


**Figure 2.** Monthly number of forest fire occurrences in Changsha, 2004–2021. where the horizontal coordinate is the month, and the vertical coordinate is the number of forest fires.



**Figure 3.** Spatial distribution of forest fires in Hunan Province from 2004 to 2022, with the horizontal coordinate being the month and the vertical coordinate being the number of real fires.





**Figure 4.** Spatial distribution of forest fires in Hunan Province, 2004–2021, where dark green is woodland, light green is grassland, blue dots are non-fire points and red dots are fire points.

Land cover data were obtained from GlobeLand30 data [57,58]. The Normalized Difference Vegetation Index (NDVI) is a common quantitative remote sensing indicator used not only to assess the spectral reflectance properties of vegetation but also to indicate changes in moisture availability and vegetation growth. NDVI values are therefore important indicators for assessing the moisture content of surface vegetation and the extent of vegetation cover. At the same time, the location of the fire and the size of the overburdened area are closely related to the condition of the vegetation at that time. Therefore, the assessment of the interaction between vegetation and fire occurrence needs to rely on NDVI data. In this study, we used the latitude and longitude of satellite hotspot information and the time of fire occurrence in Hunan Province to obtain immediate NDVI information for

each fire and non-fire site in Hunan Province from 2004 to 2021 in the Google Earth Engine platform. Meanwhile, to predict the relationship between vegetation condition and fire events in Changsha, we obtained the monthly mean NDVI information at 1 km interval random points in Changsha for 2022.

### 2.2.3. Meteorological Data

In order to obtain more comprehensive and accurate data, in this study, with the help of the Google Earth Engine (GEE) platform, based on the latitude, longitude and fire occurrence time of satellite hotspot information in Hunan Province from 2004 to 2021, the corresponding meteorological information at the time of fire occurrence was obtained by batch matching in the ERA5-Land dataset [59,60]. Based on this, we carried out detailed pre-processing of the meteorological data, such as removing missing data points and normalising the data. Nine meteorological drivers were selected, namely dew point temperature, net surface heat radiation, runoff, surface temperature, evapotranspiration, eastward wind speed, precipitation, northward wind speed and soil moisture content [61,62]. These factors are critical in the study of forest fire occurrence and development.

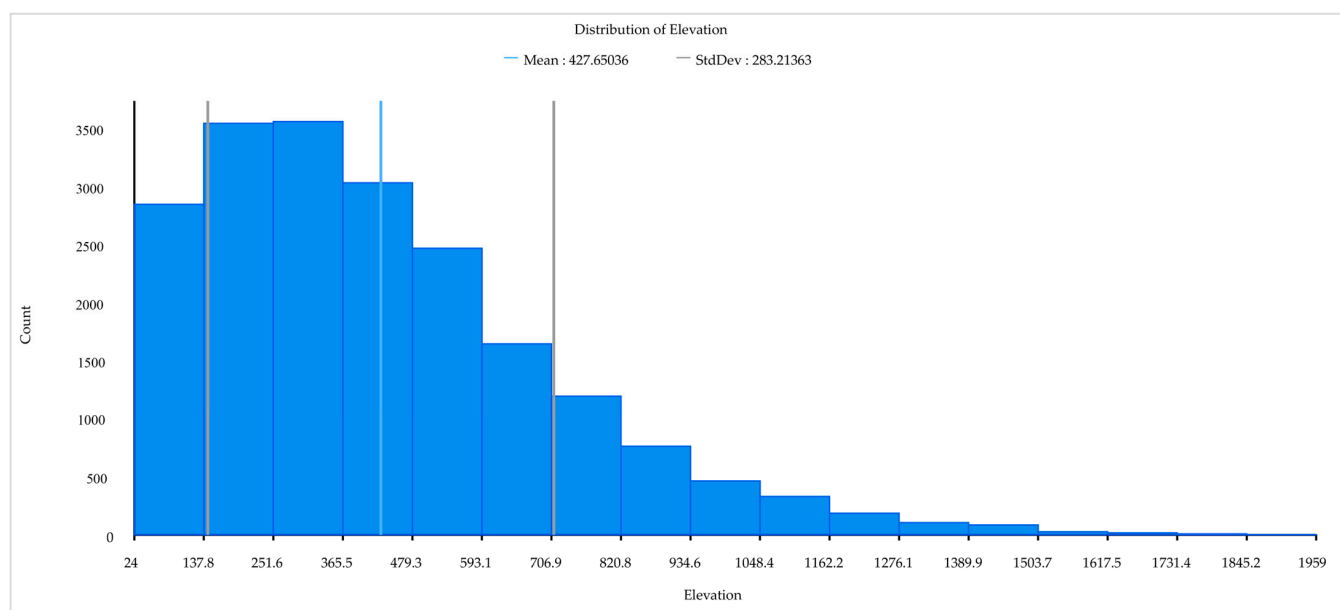
At the same time, we obtained data from 13,423 distribution points within the administrative area of Changsha at 1 km intervals and obtained monthly average meteorological drivers from January to December 2022 using the ERA5-Land dataset to construct a database of drivers for predicting the probability of forest fires in Changsha in 2022.

### 2.2.4. Topographical Data

Topographic differences have a direct impact on the composition of vegetation types and the spatial distribution of combustible material, as shown in Figure 5, with forest fires occurring more frequently at elevations below 700 m [63,64]. The topographic features of the Dongting Lake region in northern Hunan Province, the Wuling Mountains–Xuefeng Mountains in western Hunan, the Mofu Luoxiao Mountains in the east and the Nanling Mountains in the south differ significantly in terms of topographic features and vegetation types leading to large differences in the state of combustible material. Even at the same altitude, the slope and aspect of the slope can lead to significant differences in vegetation growth and local climatic conditions, which can affect the occurrence of fires, with the northern Nanling Mountains experiencing significantly more fires than the other regions. In addition, topography affects geographical accessibility, transportation accessibility, population distribution and economic development, with important implications for forest fire prevention and management. The research data for this paper were extracted from the United States Geological Survey (USGS) website (<https://earthexplorer.usgs.gov/> accessed on 31 December 2022) by downloading Digital Elevation Model (DEM) data for Hunan Province at a spatial resolution of 30m for elevation, slope and slope orientation [52].

### 2.2.5. Anthropogenic Activity Data

The data on the distribution of road networks and settlements used in this study were obtained from the 1:250,000 National Basic Geographic Database on the website of the National Basic Geographic Information Centre (<http://www.webmap.cn> accessed on 31 December 2022) [52]. In general, the closer the distance to roads, railways and settlements, the more frequent the anthropogenic activities [65]. As for population density and Gross Domestic Product (GDP), the data are obtained from the Resource and Environment Science and Data Center (<https://www.resdc.cn> accessed on 31 December 2022) with a resolution of 1 km in 2019 [66,67]. The more densely populated and economically developed the area is, the more activities such as trekking and travelling around the area, and the probability of anthropogenic fire use increases the risk of forest fires [5,6,52].



**Figure 5.** Number of forest fires at different altitudes, where the horizontal coordinates are altitude ranges and the vertical coordinates are the number of forest fires, the dark blue line represents the mean and the grey line represents the standard deviation.

#### 2.2.6. Predictors of Forest Fire Occurrence

In this study, twenty-one forest fire occurrence prediction impact factors were selected from the above five categories. Their names, abbreviations, resolution/units and data sources are shown in Table 1:

**Table 1.** Predictors of forest fire occurrence.

Name of Predictor	Abbreviations	Resolution, Units	Data Sources
Vegetation canopy water content	Vcwc	m of water equivalent	<a href="https://cds.climate.copernicus.eu">https://cds.climate.copernicus.eu</a> accessed on 30 December 2022
evaporation from the top of canopy	Eftc	m of water equivalent	<a href="https://cds.climate.copernicus.eu">https://cds.climate.copernicus.eu</a> accessed on 30 December 2022
Volume of water in soil layer 1 (0–7 cm)	Vws	1 (volume fraction)	<a href="https://cds.climate.copernicus.eu">https://cds.climate.copernicus.eu</a> accessed on 30 December 2022
Type of land use	Ldu	30 m	<a href="http://www.globallandcover.com">http://www.globallandcover.com</a> accessed on 31 December 2022
Easterly wind speed	Ews	m/s	<a href="https://cds.climate.copernicus.eu">https://cds.climate.copernicus.eu</a> accessed on 30 December 2022
northerly wind speed	Nws	m/s	<a href="https://cds.climate.copernicus.eu">https://cds.climate.copernicus.eu</a> accessed on 30 December 2022
Total evaporation	Tev	m of water equivalent	<a href="https://cds.climate.copernicus.eu">https://cds.climate.copernicus.eu</a> accessed on 30 December 2022
Surface temperature 2 m	Stem	K	<a href="https://cds.climate.copernicus.eu">https://cds.climate.copernicus.eu</a> accessed on 30 December 2022
Normalized Difference Vegetation Index	Ndvi	250 m	<a href="https://cds.climate.copernicus.eu">https://cds.climate.copernicus.eu</a> accessed on 30 December 2022
Dew point temperature	Dwp	K	<a href="https://cds.climate.copernicus.eu">https://cds.climate.copernicus.eu</a> accessed on 30 December 2022
Total precipitation	Tprp	m	<a href="https://cds.climate.copernicus.eu">https://cds.climate.copernicus.eu</a> accessed on 30 December 2022
Aspect	Asp	30 m	<a href="https://earthexplorer.usgs.gov">https://earthexplorer.usgs.gov</a> accessed on 31 December 2022



Table 1. Cont.

Name of Predictor	Abbreviations	Resolution, Units	Data Sources
Slope	Slo	°	<a href="https://earthexplorer.usgs.gov">https://earthexplorer.usgs.gov</a> accessed on 31 December 2022
Elevation	Eva	m	<a href="https://earthexplorer.usgs.gov">https://earthexplorer.usgs.gov</a> accessed on 31 December 2022
Nearest road and railway	nrar	km	<a href="http://www.webmap.cn">http://www.webmap.cn</a> accessed on 16 December 2022
Gross domestic product	GDP	RMB/km <sup>2</sup>	<a href="https://www.resdc.cn">https://www.resdc.cn</a> accessed on 16 December 2022
Population density	Popd	persons/km <sup>2</sup>	<a href="https://www.resdc.cn">https://www.resdc.cn</a> accessed on 16 December 2022
Closest distance to population centres	Cdtp	km	<a href="http://www.webmap.cn">http://www.webmap.cn</a> accessed on 16 December 2022
Longitude of the fire point	Lon	°	<a href="https://slcyfh.mem.gov.cn">https://slcyfh.mem.gov.cn</a> accessed on 31 December 2022
Latitude of the fire point	Lat	°	<a href="https://slcyfh.mem.gov.cn">https://slcyfh.mem.gov.cn</a> accessed on 31 December 2022
Fire point date	Fpd	date	<a href="https://slcyfh.mem.gov.cn">https://slcyfh.mem.gov.cn</a> accessed on 31 December 2022

### 2.3. Research Method

For modelling, the Random Sampling mode is used, in which the training set size is 80% and the repeat train parameter is set to 100, while both model training and prediction processes are in stratified. The role of the dependent variable is set as target, the role of the 20 independent variables is set as feature and the rest of the variables are set as skip, and the machine learning models such as Random Forest, AdaBoost, and Gradient Boosting are imported for training, respectively, and the accuracy is measured by five indexes, namely AUC, CA, F1, Precision and Recall five indicators for accuracy evaluation.

#### 2.3.1. Adaptive Boosting Algorithm

Adaptive Boosting (AdaBoost) is an integrated learning method. The scientific basis for choosing the AdaBoost model for forest fire occurrence prediction is (1) Strong prediction ability: AdaBoost forms a strong classifier by combining multiple weak classifiers. It is able to better capture and utilise the relationship between features to improve prediction accuracy. (2) Adaptive learning: AdaBoost adjusts the sample weights according to misclassified samples in each iteration. This enables the model to pay more attention to samples that are difficult to classify, thus adapting to datasets with different sample distributions and complexities. (3) Not easy to cause overfitting: As AdaBoost adopts an iterative training method, only some samples are selected for training in each iteration, so it can effectively prevent the occurrence of the overfitting phenomenon [35].

In forest fire occurrence prediction, the AdaBoost model can be used to build a powerful classifier to identify and predict the occurrence of forest fires [36]. The basic principle of AdaBoost is to train a series of weak classifiers through iterations and combine them into a strong classifier. In each iteration, AdaBoost looks at samples that were misclassified in the previous round and re-weights these samples so that the misclassified samples receive more attention in the next round. In this way, after multiple iterations, AdaBoost can bring the performance of a weak classifier up to the level of a strong classifier.

In forest fire occurrence prediction, fire events can be considered as a positive category and non-fire events as a negative category [6]. First, a large amount of data on forest fires is collected, including information on features such as combustibles, meteorological factors, geography and vegetation conditions. These features are then used to construct weak classifiers, each of which classifies a sample based on the features and generates a prediction result. Next, the AdaBoost algorithm is used for iterative training. In each iteration, AdaBoost adjusts the weights of the samples based on the classification results

of the previous round, increasing the weights of misclassified samples and decreasing the weights of correctly classified samples. The adjusted sample weights are then used to train the next round of the weak classifier. Through multiple rounds of iterations, the set of weak classifiers is continuously optimised to form a strong classifier. Finally, new samples can be predicted based on the strong classifier obtained from the training. Inputting the features of the sample to be predicted, the strong classifier will classify it based on the previously learned knowledge and determine whether it is a fire event or not. The AdaBoost model can be expressed as follows:

$$h(x) = \begin{cases} 1 & \sum_{j=1}^J \alpha_j h_j(x) \geq \frac{1}{2} \sum_{j=1}^J \alpha_j \\ \text{otherwise} & \end{cases} \quad (1)$$

$$\alpha_j = \log \frac{1}{\beta_j} \quad (2)$$

### 2.3.2. Gradient Boosting Decision Tree Algorithm

The Gradient Boosting Decision Tree (GBDT) is an algorithm for classifying or regressing data by using a linear combination of basis functions and continuously reducing the residuals generated by the training process. GBDT is also an iterative integrated learning method that gradually improves prediction accuracy by weighting repeated training data samples and optimising the model through a gradient descent algorithm. The scientific basis for choosing it as the prediction of forest fire occurrence includes (1) Gradient descent: GBDT fits the residuals iteratively and optimises the model gradually through the gradient descent algorithm, which reduces the loss function and improves the prediction accuracy of the model. (2) Weak Classifier Integration: GBDT adopts the integration of multiple weak classifiers to obtain the final prediction result through weighted voting, which improves the generalisation ability of the model. (3) Regularisation: GBDT introduces a regularisation term in each iteration to prevent overfitting, which improves the stability and generalisation ability of the model [5,6,36].

In forest fire prediction, GBDT can build a powerful regression model using historical weather data, fuel features, human activities and other relevant information [41,68]. By learning the forest fire occurrence status (yes or no) from the training samples, GBDT can learn the patterns and characteristics of forest fire occurrence and then make predictions about future forest fires. GBDT can also provide an assessment of feature importance to help identify factors that have a significant impact on the occurrence of forest fires and provide decision support for fire prevention efforts. The training process can be represented by the following equation [43]:

$$f_0(x) = \operatorname{argmin} \sum_{i=1}^N L(y_i, c) \quad (3)$$

This step is to initialise the weak learner, where  $f_0(x)$  is the regression tree function for the current iteration, and the next step is to calculate the residuals for each sample.

$$r_{im} = - \left[ \frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f(x)=f_{m-1}(x)} \quad (4)$$

Negative degrees are calculated for each sample, where  $r_{im}$  is the residual value.

$$f(x) = f_0(x) + \sum_{m=1}^M \sum_{j=1}^J R_{jm} I(x \in R_{jm}) \quad (5)$$

The learner is updated by calculating the best-fit value and the resulting  $f(x)$  is the final learner, where  $R_{jm}$  is the leaf node region.

### 2.3.3. Random Forest Algorithm

Random Forest (RF) is a classifier that uses multiple trees to train and predict samples. It is widely used in fields such as medicine, genetics, ecology and remote sensing because of its ability to select important feature variables and automatically identify the importance of the feature variables [62]. The scientific rationale for selecting the RF model for forest fire prediction includes (1) Random sampling: the RF uses Bootstrap sampling to take samples from the original training set with put-backs. This allows each decision tree to have slightly different training data, increasing model diversity and reducing variance. (2) Random feature selection: At each node, RF randomly selects a portion of features for partitioning, which helps to reduce the correlation between features and improves the robustness of the model. (3) Voting Integration: RF decides the final prediction result using the voting results of multiple decision trees, which reduces the errors that may be brought by a single decision tree [4,5,34,41,62,63,66,67].

The RF model can also be trained in parallel, thus improving operational efficiency and maintaining good accuracy even when the training set is missing data. However, RF model training samples require large data size and cannot make over-range predictions. In this paper, fire point training samples were collected from a large range of Hunan Province data, while a smaller range of 1 km random point data from Changsha City was used for the test set. This approach helps to overcome the disadvantage of small training data and low accuracy and uses the RF model to perform importance evaluation to screen forest fire drivers. At the same time, adjusting the core parameters affecting the accuracy of the RF model is important to improve the model stability and generalization ability. In this paper, we set the decision tree trees ( $n_{\text{estimators}}$ ) to 600, used a random sample of 80% of the forest fire samples as the training set and determined the optimal parameters of the model by setting up five repetitions of the training. This approach helps to improve the accuracy and robustness of the model.

Training phase: For each tree: (1) A random set of samples with  $N$  samples is obtained by performing a put-back sampling from the original forest fire occurrence training dataset. (2) At each node, a subset of  $m$  features is randomly selected ( $m < \text{total number of features}$ ), and then the best features and division points are selected for division by some division criterion (e.g., Gini coefficient or information gain). (3) Step (2) is repeated recursively until stopping when a stopping condition is reached (e.g., maximum depth is reached, or the number of node samples is less than a threshold). The completed random forest model is obtained.

Prediction phase: For each sample: (1) Forest fire occurrence is predicted on each tree in the random forest. (2) For classification problems, a voting mechanism is used to determine the final prediction category; for regression problems, the average of all tree prediction outputs is taken as the final prediction.

### 2.3.4. Inverse Distance Weight Interpolation Algorithm

The Inverse Distance Weighting (IDW) interpolation method involves assigning different weights to the data points according to their distance from the position to be interpolated and then performing a weighted average. The probabilities of the obtained 1 km  $\times$  1 km grid points in Changsha City were interpolated using the inverse distance weight model, and the mask extraction was performed using the Changsha City forest land cover data. The predicted distribution of monthly probability of forest fire occurrence in Changsha City's forest cover area was plotted. This method is more accurate for data with a relatively dense and uniform distribution of point sets. The points sampled in the test dataset for this study are equally spaced at 1 km intervals and are well suited to the IDW method. The core idea of IDW is to assume that things that are closer together are more correlated and are given more weight. In the analysis of the drivers affecting forest fire occurrence, IDW uses measurements around the predicted location when predicting other locations that have not been sampled. Sampled values closest to the predicted location have a greater influence on the predicted value than sampled values further away from the predicted location. The

inverse distance weighting method assumes that each sampling point has a local influence which decreases with increasing distance. Its formula can be expressed as follows:

$$\hat{Z}(x_0) = \frac{\sum_{i=1}^n Z(X_i) d_{ij}^{-r}}{\sum_{i=1}^n d_{ij}^{-r}} \quad (6)$$

### 2.3.5. Accuracy Evaluation

The classification ability of different machine learning methods is evaluated using five metrics: Accuracy, Precision, Recall, F1 (H-mean) and AUC (Area Under Curve). Accuracy is the assessment of the proportion of correct predictions, precision is the assessment of the detection rate of predicted positive cases, recall is the assessment of the detection rate of true positive cases and F1 is used to assess the precision and recall rates. The more the area enclosed by the ROC curve and the horizontal axis is close to 1, the better the prediction of the model. Accuracy, Precision, Recall and F1 can be expressed by the following equations:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (10)$$

where  $TP$  is the number of positive samples for which the model predicts a positive class,  $FP$  is the number of negative samples for which the model predicts a positive class,  $TN$  is the number of negative samples for which the model predicts a negative class and  $FN$  is the number of positive samples for which the model predicts a negative class.

### 2.3.6. Evaluation of the Importance of Characteristic Factors

After selecting the optimal model for accuracy, the relative importance of feature variables needs to be evaluated to achieve important feature selection. In this paper, the ranking is performed by using Gain Ratio and Gini, and the GINI ranking uses the GINI index in the decision tree to calculate the importance of features. Specifically, it randomly rewashes and divides the dataset, then calculates the GINI index of each feature in different divisions and finally ranks the features according to the size of the GINI index such that the features with higher importance are ranked higher. The formulas of Gain Ratio and Gini index are as follows:

$$GainRatio(D, a) = \frac{Gain(D, a)}{IV(a)} \quad (11)$$

$$IV(a) = - \sum_{v=1}^V \frac{|D^v|}{|D|} \log_2 \frac{|D^v|}{|D|} \quad (12)$$

$$Gini(D) = 1 - \sum_{i=0}^n \left( \frac{D_i}{D} \right)^2 \quad (13)$$

$$Gini(D|A) = \sum_{i=0}^n \frac{D_i}{D} Gini(D_i) \quad (14)$$

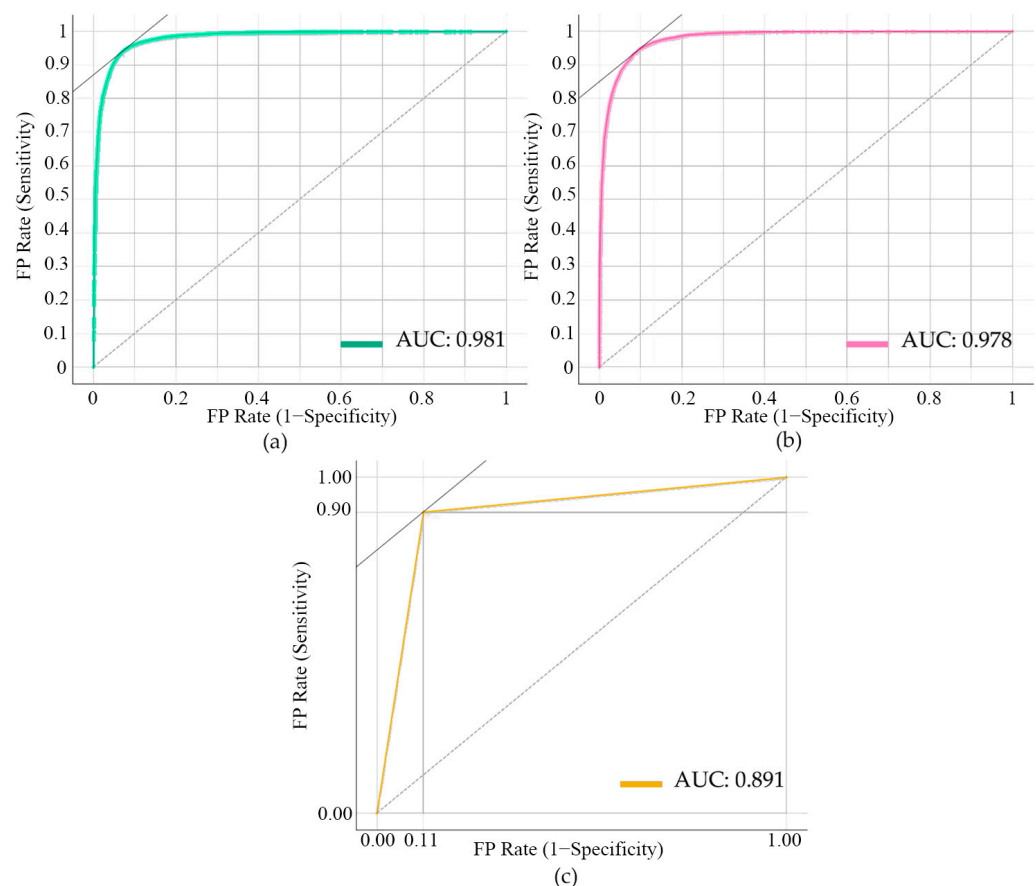
The information gain rate penalises attributes with more values by introducing a penalty term called Split information (SI), where  $IV(a)$  is determined by the number of eigenvalues of attribute  $A$ . The greater the number, the larger the  $IV$  value and the smaller the information gain rate. This prevents the model from preferring attributes with more eigenvalues, and if it is simply segmented according to this rule, the model will again prefer features with fewer eigenvalues. Therefore, the attributes with higher-than-average information gain are first identified from the candidate segmentation attributes, and the ones with the highest gain rate are selected from them.

The Gini coefficient is used as a criterion to select the optimal segmentation attribute, which can be used for classification and regression. *Gini* describes the purity, similar to the meaning of information entropy.  $Gini(D)$  reflects the purity of the dataset  $D$ , and the smaller the value, the higher the purity. We select the attribute in the candidate set that makes the minimum *Gini* index after division as the optimal division attribute.

### 3. Results

#### 3.1. Model Accuracy Validation

We used five indicators to analyse the accuracy of the three models, and the results show that the accuracy of the three models is above 89%, among which the RF model (AUC = 98.1%), with better accuracy in all indicators, is more suitable for constructing the prediction model for the occurrence of forest fires in Hunan Province. It is followed by Gradient Boosting (AUC = 97.8%) and AdaBoost (AUC = 89.1%) (Figure 6, Table 2).



**Figure 6.** Plotting ROC curves and corresponding AUC values using different machine learning algorithms. (a) is the R-curve for the RF model, (b) is the R-curve for the GBDT model and (c) is the R-curve for the Adaboost model.



**Table 2.** Model accuracy comparison.

Machine Learning Algorithm	Category	Accuracy	Precision	Recall	F1	AUC
SVM	Forest fire	0.779	0.768	0.801	0.784	0.847
	Non-forest fire		0.791	0.756	0.773	
GBDT	Forest fire	0.899	0.882	0.921	0.901	0.962
	Non-forest fire		0.917	0.876	0.896	
RF	Forest fire	0.908	0.896	0.925	0.910	0.965
	Non-forest fire		0.922	0.892	0.907	

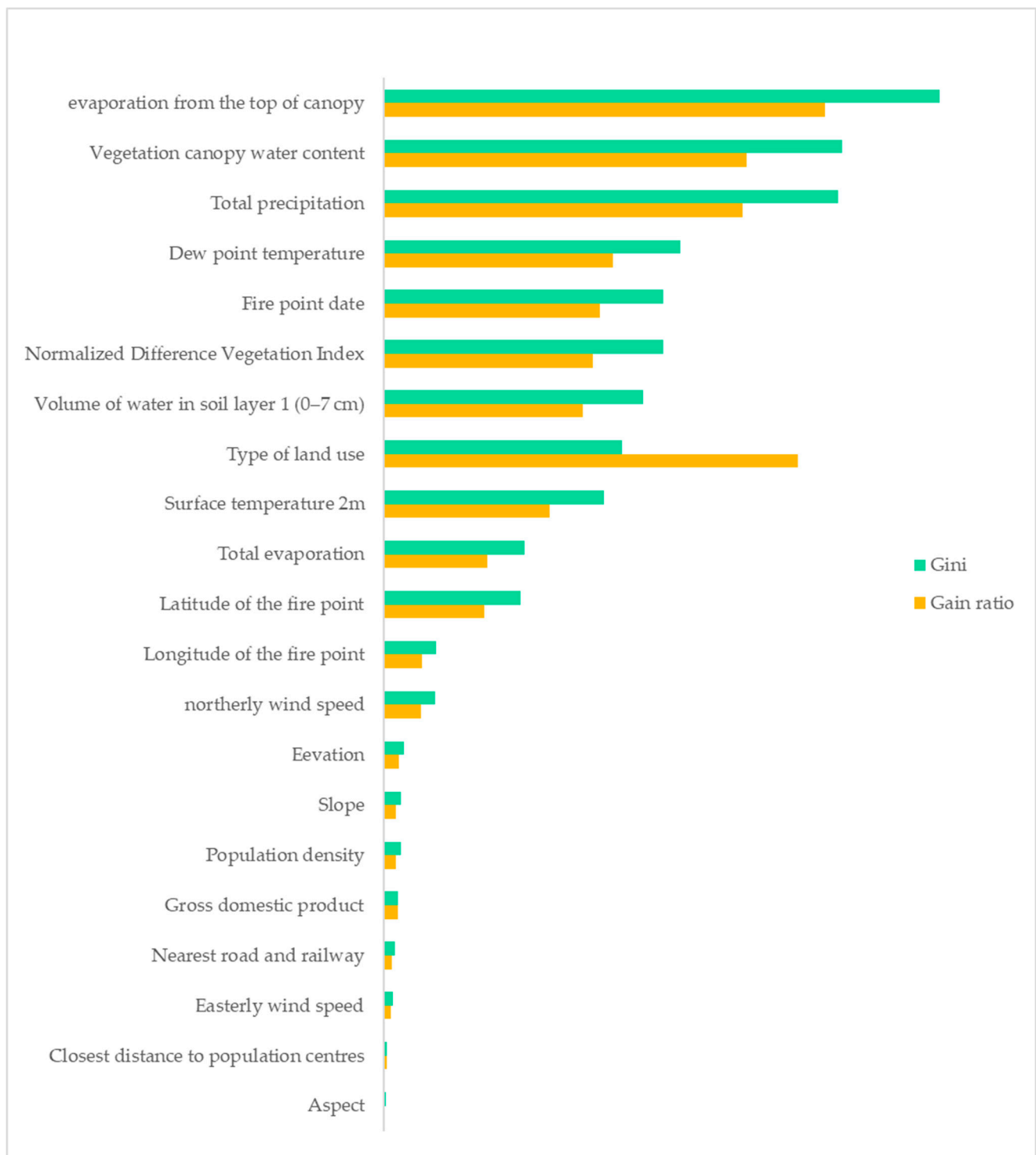
### 3.2. Characteristic Factor Importance Evaluation

Among the 21 forest fire factors, evapotranspiration from the top of the vegetation canopy was the most important (weight: 15.4%), followed by vegetation canopy water content (weight: 12.7%), closely followed by cumulative rainfall (weight: 12.58%), dew-point temperature (weight: 8.2%), date (weight: 7.7%), NDVI (weight: 7.7%) and volume of surface water in the soil (weight: 7.2%), etc. (Table 3, Figure 7).

**Table 3.** Ranking table for evaluating the importance of predictors.

Name of Predictor	Gain Ratio	Gini
Evaporation from the top of canopy	12.22%	15.40%
Vegetation canopy water content	10.05%	12.70%
Total precipitation	9.95%	12.58%
Dew point temperature	6.35%	8.21%
Fire point date	5.98%	7.74%
Normalized Difference Vegetation Index	5.79%	7.74%
Volume of water in soil layer 1 (0–7 cm)	5.52%	7.19%
Type of land use	11.47%	6.61%
Surface temperature 2 m	4.61%	6.10%
Total evaporation	2.86%	3.90%
Latitude of the fire point	2.79%	3.80%
Longitude of the fire point	1.06%	1.46%
northerly wind speed	1.04%	1.42%
Elevation	0.40%	0.56%
Slope	0.34%	0.48%
Population density	0.33%	0.46%
Gross domestic product	0.39%	0.40%
Nearest road and railway	0.23%	0.32%
Easterly wind speed	0.18%	0.25%
Closest distance to population centres	0.07%	0.10%
Aspect	0.04%	0.06%

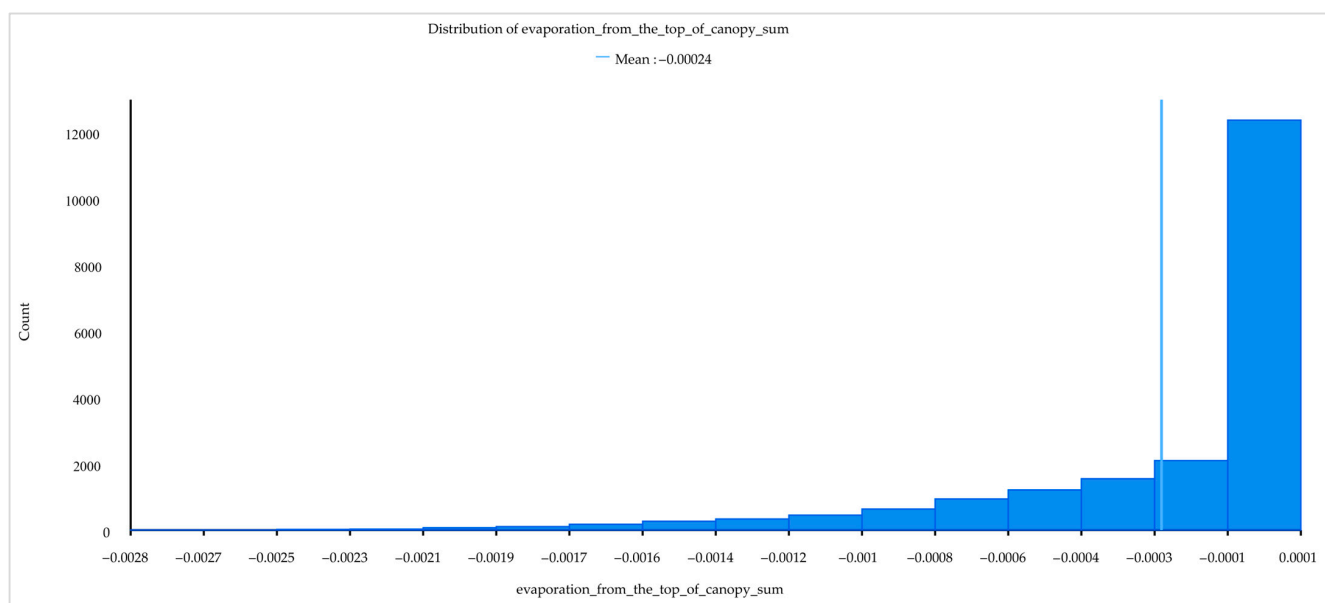
Evaporation from the top of the vegetation canopy (weight: 15.4%) is related to factors such as forest soil moisture and atmospheric humidity. When evaporation from the top of the vegetation canopy is low, the moisture content of the forest soil decreases and the environment becomes drier, thus increasing the probability of fire. Evapotranspiration from the top of the vegetation canopy affects meteorological factors such as temperature, relative humidity and wind speed, which in turn further affect the probability of fire occurrence. For example, low relative humidity and strong winds will reduce the forest fuel moisture, increasing the probability of fire. Evaporation from the top of the vegetation canopy is influenced by the degree and type of vegetation cover. Dense cover and flammable vegetation types can increase the rate and intensity of fire spread. As shown in Figure 8, evaporation from the top of the vegetation canopy reached nearly 12,000 fires in a relatively small interval, with a significant impact.



**Figure 7.** Plot of the results of the feature factor importance assessment, with the results of the Gini algorithm evaluation in green and the results of the gain ratio algorithm evaluation in orange.

The canopy water content of vegetation (weight: 12.7%) is the amount of water in the plant canopy and is one of the most important factors affecting the probability of forest fires. Changes in the moisture content of the plant canopy can affect the occurrence and spread of fires and are therefore of great importance for forest management and fire prevention. The higher the canopy moisture content of vegetation, the more difficult it is to start a fire before it burns and the lower the burning rate when it burns, as having enough moisture will weaken burning and fire spread. Therefore, if the vegetation is sufficiently

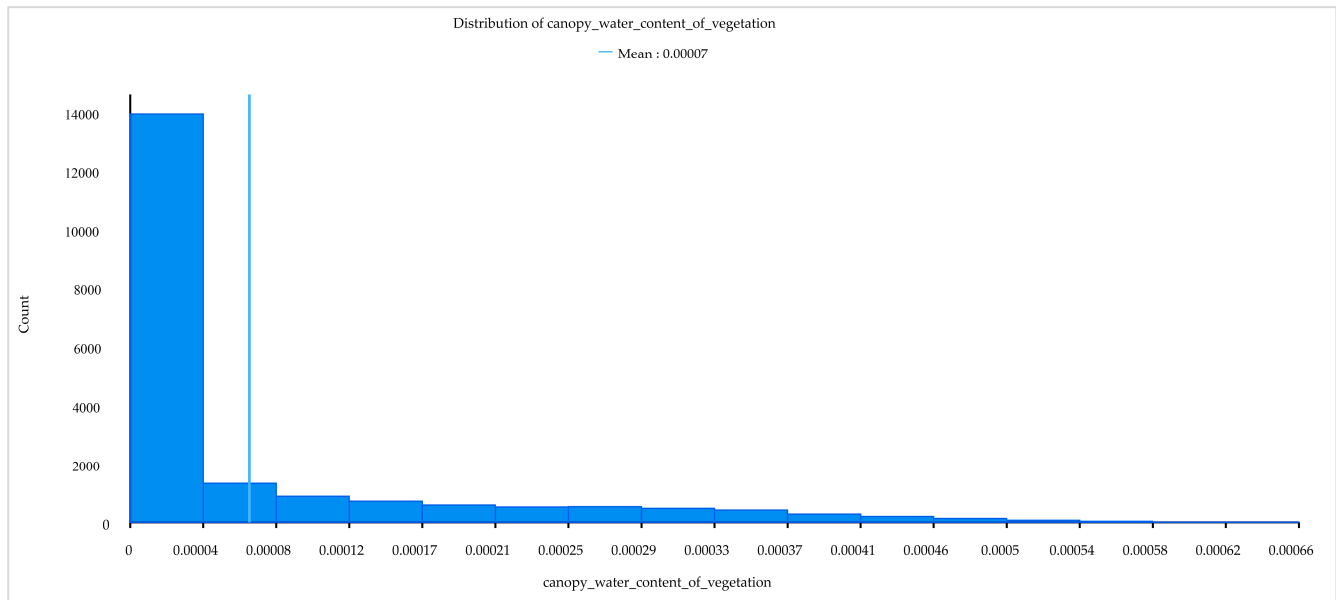
rich in canopy water content, it will reduce the probability of forest fires. On the other hand, when the canopy water content of vegetation is too low, the plants will lose their natural fire protection capacity. Plants lose moisture in dry weather and become dry and flammable, and fires can easily spread once the source of the fire comes into contact with the vegetation. Therefore, the probability of forest fires increases when the canopy moisture content of vegetation is too low. The canopy moisture content of vegetation is a key factor in the probability of forest fires. Keeping the vegetation canopy moisture content within an appropriate range will reduce the risk of fire occurrence and spread and enable better forest management and fire prevention. As shown in Figure 9, the number of forest fires occurring at low vegetation canopy moisture content reached over 14,000, with particularly significant impacts.



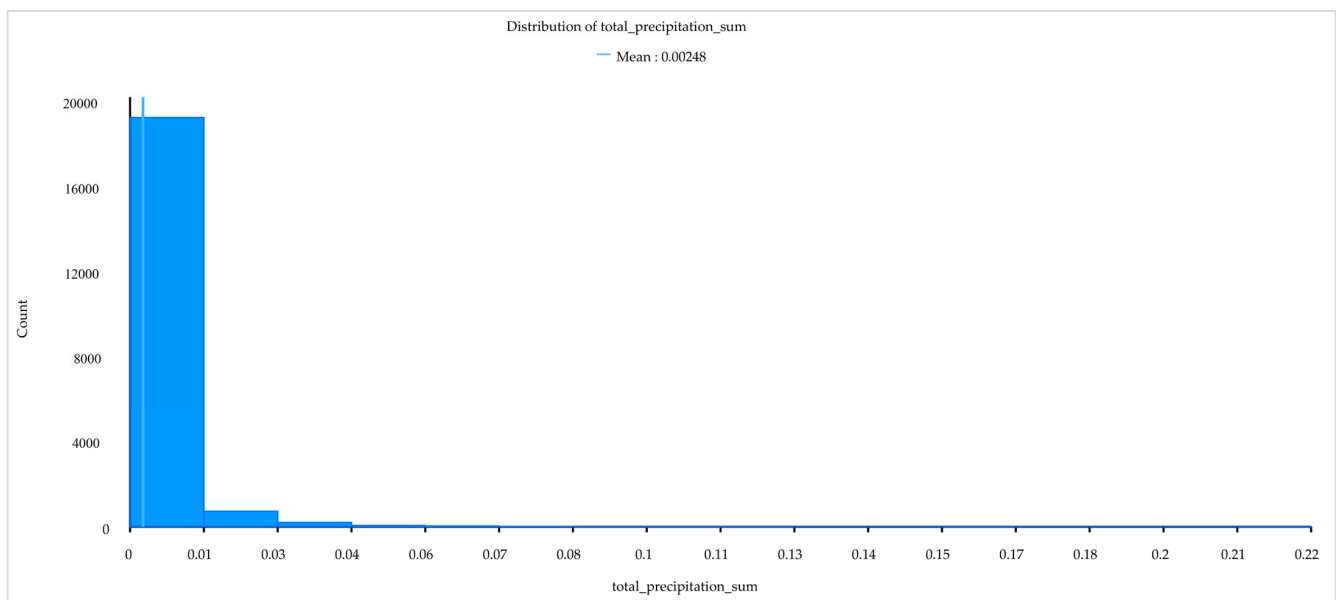
**Figure 8.** Statistical plot of the number of forest fires under different vegetation canopy evapotranspiration conditions, where the horizontal coordinate is the vegetation canopy evapotranspiration interval, the vertical coordinate is the number of forest fires, and the dark blue line represents the mean value.

Total precipitation (weight: 12.58%) is the total amount of rainfall over a period of time. It is one of the most important factors affecting the probability of forest fires. The higher the total precipitation, the greater the moisture content in the vegetation, thus reducing the probability of forest fires. Firstly, rainfall increases the moisture content of vegetation. In times of drought, the moisture in the vegetation gradually evaporates away, making it easier to burn and increasing the probability of fire. However, if there is sufficient rainfall to continuously recharge the water, the vegetation will be wetter and harder to catch fire, thus reducing the probability of fires. Secondly, rainfall can control the occurrence and spread of fires. Fires are often difficult to control during periods of drought. Because dry vegetation burns easily, fires tend to spread quickly. However, in the event of rainfall, the rain can reduce the ambient temperature, slowing down the fire and aiding in fire suppression efforts. At the same time, rainfall also makes the vegetation around a fire more moist and less likely to catch fire or spread. However, a relatively water-poor situation can be predicted by a pre-rainfall analysis, and fire prevention efforts can be advanced by increasing artificial rainfall, for example. Conversely, if rainfall reaches a certain level, it may also cause natural disasters such as flooding, which can create new problems. In summary, total precipitation is one of the most important factors affecting the probability of forest fires. By consistently increasing total precipitation, the moisture content in vegetation can be increased, reducing the distribution of drought-prone flammable vegetation and

effectively preventing forest fires from occurring. As shown in Figure 10, when the total precipitation is small, the number of forest fires exceeds 18,000, indicating that this indicator has a significant impact on the occurrence of forest fires and that measures should be taken to take special precautions when the total precipitation is in this range.



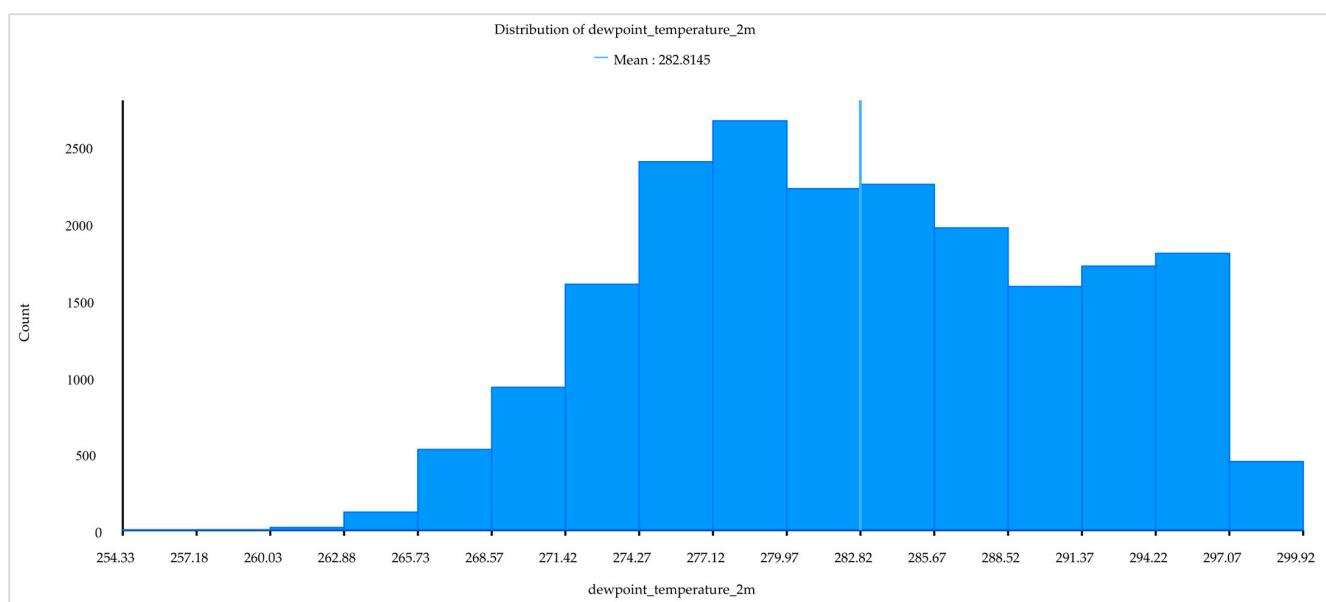
**Figure 9.** Statistics on the number of forest fires under different canopy water content of vegetation, where the horizontal coordinates are the canopy water content of vegetation intervals, the vertical coordinates are the number of forest fires and the dark blue line represents the mean value.



**Figure 10.** Statistical map of the number of forest fires under different total precipitation conditions, where the horizontal coordinate is the vegetation canopy moisture content interval, the vertical coordinate is the number of forest fires and the dark blue line represents the mean value.

The dew point temperature (weight: 8.21%) is a measure of the moisture content of the air and has a significant impact on the probability of forest fires. The lower the dew point temperature, the lower the moisture content of the air and the more likely vegetation will dry out, thus increasing the likelihood of forest fires. Specifically, when there is insufficient

moisture in the air, the wetness of the vegetation surface is reduced and the vegetation itself is more likely to dry out, making it easier for a fire to spread if it encounters a source of ignition. In addition, the lower the moisture content in the air, the faster the fire spreads and the more severe the damage caused by the fire. Conversely, when there is sufficient moisture in the air, the vegetation surface is relatively more moist and therefore less likely to dry out and catch fire. In the event of a fire, fire spread is also somewhat inhibited as higher air humidity slows the spread of fire. Thus, higher dew point temperatures are more conducive to reducing the probability of forest fires. In conclusion, the effect of dew point temperature on the occurrence of forest fires depends on the moisture content of the air. When the moisture content of the air is low, the lower the dew point temperature, the greater the probability of fire occurrence; conversely, the lower the probability of fire occurrence. Therefore, in forest fire prevention and monitoring, attention should be paid to the moisture content of the air and reasonable control of the moisture content of the air to maintain the moisture balance of the vegetation, thus reducing the probability of forest fires. As shown in Figure 11, the influence of dew point temperature on the number of forest fires spans a wide range and the degree of influence is not very concentrated compared to combustible related factors.



**Figure 11.** Statistical plot of the number of forest fires at different dew point temperatures, where the horizontal coordinates are the dew point temperature intervals, the vertical coordinates are the number of forest fires and the dark blue line represents the mean.

According to the results, forest fuels have a strong influence on the occurrence of forest fires, with meteorological factors having a secondary effect on forest fires. In general, long periods of increased air temperature and decreased relative humidity lead to increased transpiration and decreased fuel moisture content in forest areas, therefore increasing the flammability of fuel in forest areas and leading to an increased probability of forest fires. In contrast, an increase in rainfall and relative humidity increases the water content of combustible material and reduces the risk of forest fires. Furthermore, in addition to the important indicator of daily rainfall, the effect of rainfall on forest fires is also related to total precipitation. In other words, even though vegetation has the ability to regulate and adapt, its flammability gradually increases if there is a prolonged lack of rainfall. Wind speed has been used in forest fire-related research mainly to study its effect on the direction and speed of spread and propagation of forest fires. However, its effect is due to its ability to change the relative humidity of the air and heat, among other things.



Vegetation factors have a greater impact on the probability of forest fires compared to topography and human activity factors. Areas with high forest cover tend to have more dry, dead wood and vegetation, which can burn easily and spread more quickly, so the probability of forest fires is higher in areas with high forest cover. Areas with high grass cover have a relatively lower probability of forest fires because grass has a shorter growth cycle, is easier to clear and manage, and does not create contiguous burning areas as forests do. In addition, accelerated urbanisation can upset the original ecological balance and increase the risk of forest fires. Water areas can reduce the incidence and rate of spread of forest fires. Land cover is therefore an important factor in determining the probability of forest fires and needs to be taken into account in forest management and prevention measures.

As for the topographic factors, their influence on forest fire occurrence is small. The order of contribution is elevation, slope and slope direction, with slope direction having a relatively small effect, probably because differences in elevation affect climatic conditions (e.g., air humidity, air temperature, rainfall, etc.) and vegetation conditions (e.g., combustible species, combustible water content and load). The ability of the terrain to retain moisture is negatively correlated with slope size, with steeper slopes leading to moisture loss, which can lead to faster drying of combustibles and a reduction in combustible moisture content.

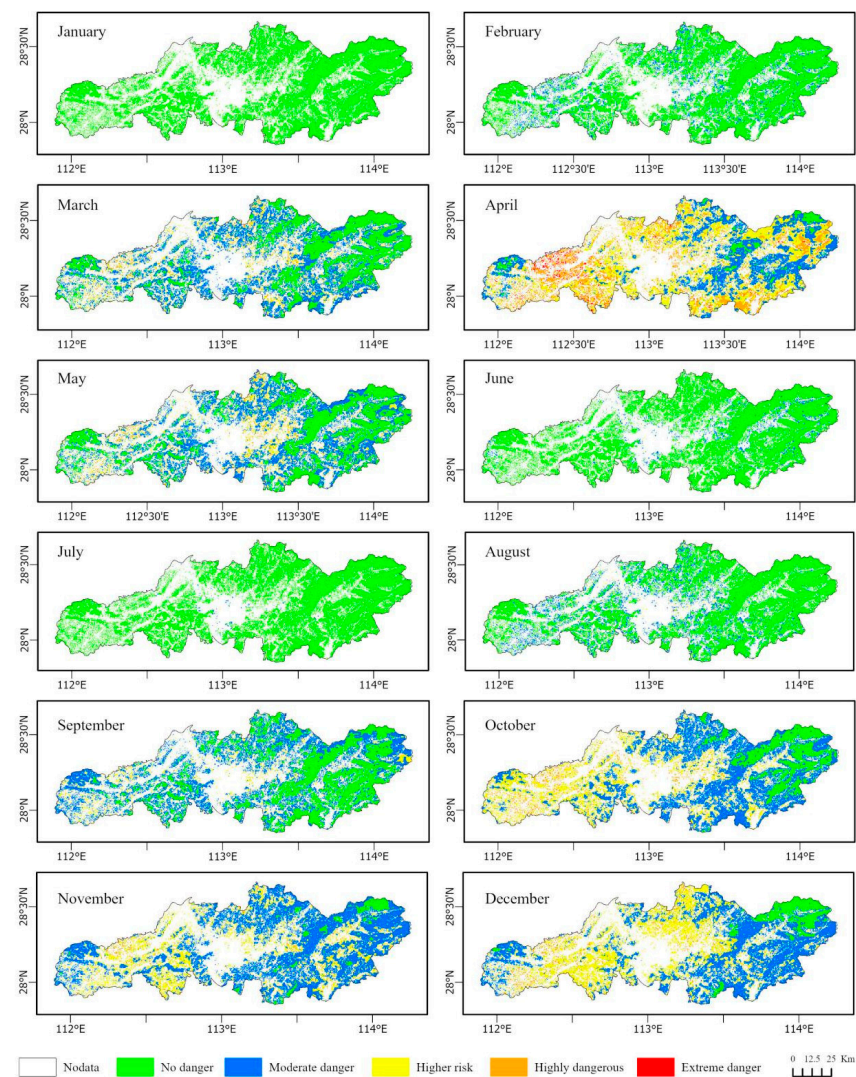
The contribution of human activity to the forest fire risk early warning model is relatively small among all the drivers, with the order of contribution being GDP, closest distance to road and rail lines, population density and closest distance to residential areas. As China's forest fire prevention education has intensified and people's awareness of forest fire prevention has gradually increased, for example, there have been no forest fires for many years in the scenic area of Mount Yuelu, which is heavily trafficked during holidays and weekends.

### 3.3. Mapping of Forest Fire Risk for Different Months

In this study, the RF model with optimal performance was selected, and important characteristic factors (evaporation from vegetation canopy, water content of vegetation canopy, cumulative rainfall, dew point temperature, date, NDVI, volume of water in the 0–7 cm soil layer, land class, surface temperature 2 m, total evaporation, latitude, longitude, etc.) were input for Changsha City month by month in 2022. The inverse distance weighting method was used to interpolate the predicted points to generate a monthly forest fire probability distribution map for Changsha City. The fire risk level in this study was divided into five levels (I–V), I: the probability range of 0.0–0.2 represents the no-risk zone, where forest fires are almost impossible; II: the probability range of 0.2–0.4 represents the medium-risk zone, where forest fires are less likely to occur; III: the probability range of 0.4–0.6 probability range represents a higher risk zone, with a higher likelihood of forest fires; IV: 0.6–0.8 probability range represents a high risk zone, with a high likelihood of forest fires and V: 0.8–1.0 probability range represents an extremely high risk zone, with an extremely high likelihood of forest fires, and is represented by red on the indicated in red on the graph.

According to Figure 12, the monthly spatial distribution of forest occurrence risk in Changsha City varies significantly. Overall, the period from October to May is the high incidence period of forest fires, of which March, April, May, September, October, November and December are the months with high forest fire risk in Changsha City. January and February have low temperatures, low rainfall, a long rainy season and high water content of combustible materials, which makes the risk relatively low. Areas with high probability of forest fires in March are mainly located in parts of central Ningxiang and eastern Changsha County. April is the month with the highest probability of forest fires in Changsha, in which a larger part of the northern part of Ningxiang is an extremely high-risk area, and there are high-risk distribution areas around the main urban area, in the south and in the northeastern part of the Liuyang area, and there are higher-risk areas

distributed throughout most of the city. This phenomenon is mainly due to the fact that the Qingming Festival on 5 April each year is a traditional Chinese holiday, which requires people to go up to the mountains to carry out sacrificial activities and set off fireworks, leading to a sharp increase in the risk of forest fires. The areas with a high probability of forest fires in May are mainly located in the eastern part of Changsha city and part of the southwestern and northern parts of Ningxiang. In September, areas with high probability of forest fires were mainly located in the western and northern parts of Ningxiang, the central and western parts of Changsha City and a small part of the eastern part of Liuyang. In October, high-risk areas were mainly distributed in the northern part of Ningxiang, and higher-risk areas were mainly distributed in parts of Ningxiang City, central Changsha County and southern Liuyang. In November, the high-risk areas were mainly located in southern and northern Ningxiang and central Liuyang City, and the higher-risk areas were mainly located in central Ningxiang, central and northern Changsha County, and around the urban area of Liuyang City. In December, except for parts of western Ningxiang and the Liuyang area, the northern, central and western parts of Changsha were in the high-risk and higher-risk zones.



**Figure 12.** Monthly forest fire risk warning map of Changsha city. Different colours represent different forest fire risk levels, with green representing no-risk areas, blue representing medium-risk areas, yellow representing higher-risk areas, orange representing high-risk areas and red representing very high-risk areas.

Although the temperatures in Changsha City in June, July and August are the highest in the year, the mapping results show that the probability of forest fires is small and the probability of forest fires is not proportional to the temperature. This result is caused by a combination of several factors. Firstly, Changsha region belongs to the subtropical monsoon climate zone. The evaporation from the top of the vegetation canopy and the water content of the vegetation canopy are usually high in June, July and August each year. Vegetation is relatively well supplied with water in the summer months, which reduces the risk of forest fires. Secondly, these three months are a period of higher rainfall, higher cumulative rainfall during the same period and lower risk due to the high water content of combustible materials during the same period of rain and heat. In addition, Changsha has higher relative humidity and higher soil surface water volume during these three months. This allows for timely replenishment of the water needed for plant growth and slows down the possibility of forest fires caused by wilting vegetation.

#### 4. Discussion

Three machine learning algorithms are used in this study, namely Adaptive Augmentation Algorithm (AdaBoost), Gradient Boosted Tree (GBDT) and Random Forest (RF). All three algorithms are common classification algorithms with a wide range of application scenarios and good predictive performance. The results show that the accuracy of all three models is above 0.89, which is a substantial improvement in accuracy relative to recent related studies [5,6,25,28,36–43]. Among them, the RF model (AUC: 0.981) in this study has the best prediction effect and further improves the accuracy by 1–2% relative to tan (AUC: 0.972), shao (AUC: 0.951) [5,41]. The GBDT model (AUC: 0.978) improves the accuracy by 3–6% relative to tan (AUC: 0.958), shao (AUC: 0.912) [5,41]. The AdaBoost model (AUC: 0.891) with relatively poor prediction accuracy also has higher accuracy than some recent studies [35]. Meanwhile, both related studies and this paper found that the accuracy of the RF model is better relative to other methods, which means that early warning of forest fires using the RF machine learning model is relatively reliable [5,6,37,41]. The AdaBoost model has high accuracy and adaptive adjustment ability in forest fire prediction and can provide relatively accurate prediction results (AUC: 0.891) [39]. However, it is sensitive to noise and has a long training time, requiring careful parameter selection. Therefore, when using AdaBoost for forest fire prediction, attention needs to be paid to noise handling and parameter tuning to achieve the best results, and AdaBoost is not an optimal choice when dealing with the problem of forest fire occurrence prediction. Gradient Boosting Decision Tree (GBDT) is an integrated algorithm based on decision trees that can process nonlinear, nonconvex, and nonsmooth data. GBDT can adaptively select features and is robust and adaptive [69]. However, GBDT requires a large amount of data for training, and if the dataset is small, it is easy to overfit and requires careful parameter tuning. In this study, the amount of data is relatively large, and the prediction accuracy of the GBDT algorithm is high (AUC: 0.978), but there is still a certain gap compared to the RF algorithm. Random Forest is an integrated algorithm based on decision trees, which is based on random ideas and reduces the risk of overfitting by constructing multiple decision trees. RF is more tolerant to noise and missing data and can effectively assess complex interactions and nonlinear relationships between explanatory variables. The RF algorithm uses a tree-based approach, and the tree structure has a great advantage over other data structures in terms of interpretability and visualization [5,6,25,41,43]. In this study, the RF algorithm has the best prediction accuracy (AUC: 0.981), indicating that the RF algorithm is more suitable for forest fire occurrence prediction models than other algorithms.

When predicting the occurrence of forest fires, related research mainly introduces four major categories of driving factors, such as meteorology (temperature, humidity, wind speed, wind direction), vegetation (NDVI, land type), topography (elevation, slope, slope direction) and socio-humanities (GDP, population density, distance from settlements, and distance from highway lines and railway lines) [5,6,23,40–42,49,52,66,69]. In this study, we innovatively introduced forest fuel factors (vegetation evapotranspiration, vegetation

canopy water content) to construct a forest fire prediction model on the basis of the previous study [2,12,14,16,70,71]. By evaluating the importance of these factors, we found that canopy evapotranspiration, canopy moisture content, cumulative rainfall and dew point temperature were the four factors that contributed most to the construction of the forest fire risk prediction model, and vegetation canopy evapotranspiration and canopy moisture content, which characterise the fuel state, ranked as the top two factors, which proved that the fuel factor has a significant impact on forest fire occurrence. By introducing the fuel matter factor, the prediction accuracy of the model reached more than 98%, which was further improved compared with the previous related studies [5,41,49]. This result suggests that more attention needs to be paid to vegetation characteristics and combustible material status, especially the water content and evaporation of combustible material, when predicting forest fire occurrence. Meanwhile, the results of this paper show that the influence of surface temperature on the occurrence of forest fires in Changsha is ranked low, with a proportion of 6.1%. In other studies, the proportion of temperature is high, and probably the effect of temperature is very important [5,25,36,41,45,72,73]. However, temperature is not the most important factor in Changsha, and the forest fire driving factors cannot be simply transplanted to Changsha for analysis. The local vegetation characteristics and meteorological conditions need to be fully considered before introducing the impact parameters.

In order to prevent and control forest fires more effectively, this study also has identified some areas that deserve further research. For example, in the process of sample data set construction, a spatio-temporal random method was used to generate random non-fire points, although the process of generating these points avoided overlapping with the original fire points as much as possible, but whether the different random generation methods would affect the accuracy of the model needs to be further explored. Whether the accuracy of acquiring relevant parameters of meteorological reanalysis data affects the accuracy of the model also needs to be further investigated. The interactions between the various factors need to be further investigated in the future.

## 5. Conclusions

In terms of model prediction, three machine learning methods, AdaBoost, GBDT and RF, were used to predict the probability of forest fire occurrence in Changsha City. From the accuracy validation results, the accuracy of these three models is better than 89.1%. Among them, RF has the highest generalisation ability in predicting forest fire occurrence in Changsha City, which has an AUC of 98.1%, an accuracy of 93.5%, a precision of 93.9%, a recall of 94.0% and an F1 of 93.5%. This indicates that the RF model can better predict the probability of forest fire occurrence in Changsha, and the results can provide a reference for future forest fire risk prediction in Changsha.

In terms of the importance of forest fire occurrence driving factors, the evaluation results showed that the vegetation characteristics had the highest importance, and the meteorological factors were the next most important. In this study, evapotranspiration from the top of the vegetation canopy (weight: 15.4%) and water content of the vegetation canopy (weight: 12.7%) were innovatively introduced. We found that they were the two factors that had the greatest influence on the construction of forest fire prediction models, which suggests that the characteristics of the vegetation itself and the state of combustibles are more important factors that deserve attention when predicting the occurrence of forest fires. Meanwhile, we found that the temperature (weight: 8.2%) among the meteorological factors has relatively little influence on the construction of the forest fire prediction model in Changsha. The discovery of this law can, to a certain extent, influence people's subjective view that air temperature is closely related to the occurrence of forest fires and provide a new way of thinking for the optimisation of the index system for predicting the occurrence of forest fires. It is conducive to improving the ability of forest fire prediction in subtropical monsoon climate zones.



By producing monthly forest fire occurrence probability prediction maps for Changsha City, the results show that the monthly difference of forest fire risk in Changsha City is obvious, and March, April, May, September, October, November and December are the seasons of high forest fire risk in Changsha City, with a large area of medium-high fire risk area, mainly distributed in the central and northern areas of Changsha City. Areas in central and northern Changsha City with a high probability of forest fires should be closely monitored during the period of high forest fire risk, so that forest fires can be prevented early and a quick response can be made if a fire occurs. At the same time, the forest fire risk map drawn can further optimise the spatial distribution of forest fire monitoring equipment and the rational allocation of fire prevention and suppression resources. Meanwhile, June, July and August are the highest temperature periods of the year, but the probability of forest fires is relatively low, and we found that the occurrence of forest fires in Changsha City is not proportional to the temperature. In the future, we hope to optimise the model by using combustible data and meteorological data with higher timeliness and accuracy and further study the important factors affecting the occurrence of forest fires in different climatic environments.

**Author Contributions:** X.W. conceived and designed the study; X.W., Z.Y., Y.Y., G.Z., Z.P. wrote the first draft, collected the data and wrote the code to process the data; G.Z., Z.Y., Y.Y., Z.P., S.T. and Z.P. provided innovative insights in editing the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China Youth Project (grant no. 32201552), Science and Technology Innovation Platform and Talent Plan Project of Hunan Province (grant no. 2017TP1022), Youth Project of Hunan Provincial Philosophy and Social Science Foundation (grant no. 21YBQ054), Changsha City Natural Science Foundation (grant no. kq2202274) and Hunan Forestry Science and Technology Research and Innovation Project (grant no. XLKY202331).

**Data Availability Statement:** The data presented in this study are available upon request from the corresponding author.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Clarke, H.; Nolan, R.H.; De Dios, V.R.; Bradstock, R.; Griebel, A.; Khanal, S.; Boer, M.M. Forest fire threatens global carbon sinks and population centres under rising atmospheric water demand. *Nat. Commun.* **2022**, *13*, 7161. [\[CrossRef\]](#) [\[PubMed\]](#)
- Xie, J.; Qi, T.; Hu, W.; Huang, H.; Chen, B.; Zhang, J. Retrieval of Live Fuel Moisture Content Based on Multi-Source Remote Sensing Data and Ensemble Deep Learning Model. *Remote Sens.* **2022**, *14*, 4378. [\[CrossRef\]](#)
- Justino, F.; Bromwich, D.H.; Schumacher, V.; Dasilva, A.; Wang, S.-H. Arctic Oscillation and Pacific-North American pattern dominated-modulation of fire danger and wildfire occurrence. *NPJ Clim. Atmos. Sci.* **2022**, *5*, 52. [\[CrossRef\]](#)
- Ma, W.; Feng, Z.; Cheng, Z.; Chen, S.; Wang, F. Identifying Forest Fire Driving Factors and Related Impacts in China Using Random Forest Algorithm. *Forests* **2020**, *11*, 507. [\[CrossRef\]](#)
- Tan, C.; Feng, Z. Mapping Forest Fire Risk Zones Using Machine Learning Algorithms in Hunan Province, China. *Sustainability* **2023**, *15*, 6292. [\[CrossRef\]](#)
- Pang, Y.; Li, Y.; Feng, Z.; Feng, Z.; Zhao, Z.; Chen, S.; Zhang, H. Forest Fire Occurrence Prediction in China Based on Machine Learning Methods. *Remote Sens.* **2022**, *14*, 5546. [\[CrossRef\]](#)
- Wang, L.; Quan, X.; He, B.; Yebra, M.; Xing, M.; Liu, X. Assessment of the dual polarimetric sentinel-1A data for forest fuel moisture content estimation. *Remote Sens.* **2019**, *11*, 1568. [\[CrossRef\]](#)
- White, D.A.; Balocchi-Contreras, F.; Silberstein, R.P.; de Arellano, P.R. The effect of wildfire on the structure and water balance of a high conservation value Hualo (*Nothofagus glauca* (Phil.) Krasser.) forest in central Chile. *For. Ecol. Manag.* **2020**, *472*, 118219. [\[CrossRef\]](#)
- Li, W.; Xu, Q.; Yi, J.; Liu, J. Predictive model of spatial scale of forest fire driving factors: A case study of Yunnan Province, China. *Sci. Rep.* **2022**, *12*, 19029. [\[CrossRef\]](#)
- Sun, J.; Qi, W.; Huang, Y.; Xu, C.; Yang, W. Facing the Wildfire Spread Risk Challenge: Where Are We Now and Where Are We Going? *Fire* **2023**, *6*, 228. [\[CrossRef\]](#)
- Rossa, C.G. The effect of fuel moisture content on the spread rate of forest fires in the absence of wind or slope. *Int. J. Wildland Fire* **2017**, *26*, 24–31. [\[CrossRef\]](#)



12. Costa-Saura, J.M.; Balaguer-Beser, Á.; Ruiz, L.A.; Pardo-Pascual, J.E.; Soriano-Sancho, J.L. Empirical Models for Spatio-Temporal Live Fuel Moisture Content Estimation in Mixed Mediterranean Vegetation Areas Using Sentinel-2 Indices and Meteorological Data. *Remote Sens.* **2021**, *13*, 3726. [\[CrossRef\]](#)
13. Konings, A.G.; Saatchi, S.S.; Frankenberg, C.; Keller, M.; Leshyk, V.; Anderegg, W.R.L.; Humphrey, V.; Matheny, A.M.; Trugman, A.; Sack, L.; et al. Detecting forest response to droughts with global observations of vegetation water content. *Glob. Chang. Biol.* **2021**, *27*, 6005–6024. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Luo, K.; Quan, X.; He, B.; Yebra, M. Effects of Live Fuel Moisture Content on Wildfire Occurrence in Fire-Prone Regions over Southwest China. *Forests* **2019**, *10*, 887. [\[CrossRef\]](#)
15. Canadell, J.G.; Meyer, C.P.; Cook, G.D.; Dowdy, A.; Briggs, P.R.; Knauer, J.; Pepler, A.; Haverd, V. Multi-decadal increase of forest burned area in Australia is linked to climate change. *Nat. Commun.* **2021**, *12*, 6921. [\[CrossRef\]](#)
16. Chuvieco, E.; González, I.; Verdú, F.; Aguado, I.; Yebra, M. Prediction of fire occurrence from live fuel moisture content measurements in a Mediterranean ecosystem. *Int. J. Wildland Fire* **2009**, *18*, 430–441. [\[CrossRef\]](#)
17. Yin, S.; Shan, Y.; Tang, S.; Douglas, G.; Yu, B.; Cui, C.; Cao, L. Study on the Limit of Moisture Content of Smoldering Humus during Sub-Surface Fires in the Boreal Forests of China. *Forests* **2023**, *14*, 252. [\[CrossRef\]](#)
18. Lehmann, C.E.R.; Anderson, T.M.; Sankaran, M.; Higgins, S.I.; Archibald, S.; Hoffmann, W.A.; Hanan, N.P.; Williams, R.J.; Fensham, R.J.; Felfili, J.; et al. Savanna Vegetation-Fire-Climate Relationships Differ Among Continents. *Science* **2014**, *343*, 548–552. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Zhang, H.; Qi, P.; Guo, G. Improvement of fire danger modelling with geographically weighted logistic model. *Int. J. Wildland Fire* **2014**, *23*, 1130–1146. [\[CrossRef\]](#)
20. Fang, L.; Yang, J.; Zu, J.; Li, G.; Zhang, J. Quantifying influences and relative importance of fire weather, topography, and vegetation on fire size and fire severity in a Chinese boreal forest landscape. *For. Ecol. Manag.* **2015**, *356*, 2–12. [\[CrossRef\]](#)
21. Fan-Hua, K.; Xiu-Zhen, L.; Hai-Wei, Y. Landscape change on burned blanks in Daxing's Mountains. *J. For. Res.* **2004**, *15*, 33–38. [\[CrossRef\]](#)
22. Akther, M.S.; Hassan, Q.K. Remote sensing-based assessment of fire danger conditions over boreal forest. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2011**, *4*, 992–999. [\[CrossRef\]](#)
23. Fry, D.L.; Stephens, S.L. Influence of humans and climate on the fire history of a ponderosa pine-mixed conifer forest in the southeastern Klamath Mountains, California. *For. Ecol. Manag.* **2006**, *223*, 428–438. [\[CrossRef\]](#)
24. Williams, A.P.; Abatzoglou, J.T.; Gershunov, A.; Guzman-Morales, J.; Bishop, D.A.; Balch, J.K.; Lettenmaier, D.P. Observed Impacts of Anthropogenic Climate Change on Wildfire in California. *Earths Futur.* **2019**, *7*, 892–910. [\[CrossRef\]](#)
25. He, W.; Shirowzhan, S.; Pettit, C.J. GIS and Machine Learning for Analysing Influencing Factors of Bushfires Using 40-Year Spatio-Temporal Bushfire Data. *ISPRS Int. J. Geo Inf.* **2022**, *11*, 336. [\[CrossRef\]](#)
26. Sun, L.; Yang, L.; Xia, X.; Wang, D.; Zhang, T. Climatological Aspects of Active Fires in Northeastern China and Their Relationship to Land Cover. *Remote Sens.* **2022**, *14*, 2316. [\[CrossRef\]](#)
27. Kumari, B.; Pandey, A.C. MODIS based forest fire hotspot analysis and its relationship with climatic variables. *Spat. Inf. Res.* **2019**, *28*, 87–99. [\[CrossRef\]](#)
28. Rodrigues, M.; De la Riva, J. An insight into machine-learning algorithms to model human-caused wildfire occurrence. *Environ. Model. Softw.* **2014**, *57*, 192–201. [\[CrossRef\]](#)
29. Nami, M.H.; Jaafari, A.; Fallah, M.; Nahiuni, S. Spatial prediction of wildfire probability in the Hyrcanian ecoregion using evidential belief function model and GIS. *Int. J. Environ. Sci. Technol.* **2017**, *15*, 373–384. [\[CrossRef\]](#)
30. Pourtaghi, Z.S.; Pourghasemi, H.R.; Aretano, R.; Semeraro, T. Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques. *Ecol. Indic.* **2016**, *64*, 72–84. [\[CrossRef\]](#)
31. Jaafari, A.; Gholami, D.M.; Zenner, E.K. A Bayesian modeling of wildfire probability in the Zagros Mountains, Iran. *Ecol. Inform.* **2017**, *39*, 32–44. [\[CrossRef\]](#)
32. Hong, H.; Panahi, M.; Shirzadi, A.; Ma, T.; Liu, J.; Zhu, A.-X.; Chen, W.; Kougiass, I.; Kazakis, N. Flood susceptibility assessment in Hengfeng area coupling adaptive neuro-fuzzy inference system with genetic algorithm and differential evolution. *Sci. Total Environ.* **2017**, *621*, 1124–1141. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Termeh, S.V.R.; Kornejady, A.; Pourghasemi, H.R.; Keesstra, S. Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms. *Sci. Total Environ.* **2018**, *615*, 438–451. [\[CrossRef\]](#)
34. Guo, F.; Zhang, L.; Jin, S.; Tigabu, M.; Su, Z.; Wang, W. Modeling Anthropogenic Fire Occurrence in the Boreal Forest of China Using Logistic Regression and Random Forests. *Forests* **2016**, *7*, 250. [\[CrossRef\]](#)
35. Rosadi, D.; Andriyani, W.; Arisanty, D.; Agustina, D. Prediction of Forest Fire using Hybrid SOM-AdaBoost Method. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2021.
36. Coughlan, R.; Di Giuseppe, F.; Vitolo, C.; Barnard, C.; Lopez, P.; Drusch, M. Using machine learning to predict fire-ignition occurrences from lightning forecasts. *Meteorol. Appl.* **2021**, *28*, e1973. [\[CrossRef\]](#)
37. He, Q.; Jiang, Z.; Wang, M.; Liu, K. Landslide and Wildfire Susceptibility Assessment in Southeast Asia Using Ensemble Machine Learning Methods. *Remote Sens.* **2021**, *13*, 1572. [\[CrossRef\]](#)
38. Reyes-Bueno, F.; Loján-Córdova, J. Assessment of Three Machine Learning Techniques with Open-Access Geographic Data for Forest Fire Susceptibility Monitoring—Evidence from Southern Ecuador. *Forests* **2022**, *13*, 474. [\[CrossRef\]](#)

39. Rubí, J.N.; de Carvalho, P.H.; Gondim, P.R. Application of machine learning models in the behavioral study of forest fires in the Brazilian Federal District region. *Eng. Appl. Artif. Intell.* **2023**, *118*, 105649. [\[CrossRef\]](#)
40. Saha, S.; Bera, B.; Shit, P.K.; Bhattacharjee, S.; Sengupta, N. Prediction of forest fire susceptibility applying machine and deep learning algorithms for conservation priorities of forest resources. *Remote Sens. Appl. Soc. Environ.* **2023**, *29*, 100917. [\[CrossRef\]](#)
41. Shao, Y.; Feng, Z.; Sun, L.; Yang, X.; Li, Y.; Xu, B.; Chen, Y. Mapping China's Forest Fire Risks with Machine Learning. *Forests* **2022**, *13*, 856. [\[CrossRef\]](#)
42. Tariq, A.; Shu, H.; Siddiqui, S.; Munir, I.; Sharifi, A.; Li, Q.; Lu, L. Spatio-temporal analysis of forest fire events in the Margalla Hills, Islamabad, Pakistan using socio-economic and environmental variable data with machine learning methods. *J. For. Res.* **2021**, *33*, 183–194. [\[CrossRef\]](#)
43. Xie, L.; Zhang, R.; Zhan, J.; Li, S.; Shama, A.; Zhan, R.; Wang, T.; Lv, J.; Bao, X.; Wu, R. Wildfire Risk Assessment in Liangshan Prefecture, China Based on An Integration Machine Learning Algorithm. *Remote Sens.* **2022**, *14*, 4592. [\[CrossRef\]](#)
44. Adab, H.; Kanniah, K.D.; Solaimani, K.; Sallehuddin, R. Modelling static fire hazard in a semi-arid region using frequency analysis. *Int. J. Wildland Fire* **2015**, *24*, 763–777. [\[CrossRef\]](#)
45. Jaafari, A.; Zenner, E.K.; Pham, B.T. Wildfire spatial pattern analysis in the Zagros Mountains, Iran: A comparative study of decision tree based classifiers. *Ecol. Inform.* **2018**, *43*, 200–211. [\[CrossRef\]](#)
46. Bui, D.T.; Bui, Q.-T.; Nguyen, Q.-P.; Pradhan, B.; Nampak, H.; Trinh, P.T. A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. *Agric. For. Meteorol.* **2017**, *233*, 32–44. [\[CrossRef\]](#)
47. Goldarag, Y.J.; Mohammadzadeh, A.; Ardakani, A.S. Fire Risk Assessment Using Neural Network and Logistic Regression. *J. Indian Soc. Remote Sens.* **2016**, *44*, 885–894. [\[CrossRef\]](#)
48. Vecín-Arias, D.; Castedo-Dorado, F.; Ordóñez, C.; Rodríguez-Pérez, J.R. Biophysical and lightning characteristics drive lightning-induced fire occurrence in the central plateau of the Iberian Peninsula. *Agric. For. Meteorol.* **2016**, *225*, 36–47. [\[CrossRef\]](#)
49. Zhao, P.; Zhang, F.; Lin, H.; Xu, S. GIS-Based Forest Fire Risk Model: A Case Study in Laoshan National Forest Park, Nanjing. *Remote Sens.* **2021**, *13*, 3704. [\[CrossRef\]](#)
50. Rodrigues, M.; Zúñiga-Antón, M.; Alcasena, F.; Gelabert, P.; Vega-Garcia, C. Integrating geospatial wildfire models to delineate landscape management zones and inform decision-making in Mediterranean areas. *Saf. Sci.* **2022**, *147*, 105616. [\[CrossRef\]](#)
51. Wang, S.; Zhang, G.; Tan, S.Q.; Wang, P.; Wu, X. Assessment of forest fire risk in Hunan province based on spatial logistic model. *J. Cent. South Univ. For. Technol.* **2020**, *40*, 88–95.
52. Yang, X.; Jin, X.; Zhou, Y. Wildfire Risk Assessment and Zoning by Integrating Maxent and GIS in Hunan Province, China. *Forests* **2021**, *12*, 1299. [\[CrossRef\]](#)
53. Luo, F.; Wang, C.; Lei, H.; Xiao, Z. Young adults' perception of forests using landscape-image-sketching technique: A case study of Changsha, Central China. *Int. J. Environ. Res. Public Health* **2023**, *20*, 2986. [\[CrossRef\]](#)
54. Hu, T.; Dong, J.; Hu, Y.; Qiu, S.; Yang, Z.; Zhao, Y.; Cheng, X.; Peng, J. Stage response of vegetation dynamics to urbanization in megacities: A case study of Changsha City, China. *Sci. Total Environ.* **2023**, *858*, 159659. [\[CrossRef\]](#)
55. Liu, Z.H.; Yang, J.; He, H.S.; Chang, Y. Spatial point analysis of fire occurrence and its influence factor in Huzhong forest area of the Great Xing'an Mountains in Heilongjiang Province, China. *Acta Ecol. Sin.* **2011**, *31*, 1669–1677.
56. Bhattarai, N.; Dahal, S.; Thapa, S.; Pradhananga, S.; Karky, B.S.; Rawat, R.S.; Windhorst, K.; Watanabe, T.; Thapa, R.B.; Avtar, R. Forest Fire in the Hindu Kush Himalayas: A Major Challenge for Climate Action. *J. For. Livelihood* **2022**, *21*, 14–31. [\[CrossRef\]](#)
57. Chen, J.; Chen, J.; Liao, A.; Cao, X.; Chen, L.; Chen, X.; He, C.; Han, G.; Peng, S.; Lu, M.; et al. Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS J. Photogramm. Remote Sens.* **2015**, *103*, 7–27. [\[CrossRef\]](#)
58. Li, Y.; Feng, Z.; Chen, S.; Zhao, Z.; Wang, F. Application of the Artificial Neural Network and Support Vector Machines in Forest Fire Prediction in the Guangxi Autonomous Region, China. *Discret. Dyn. Nat. Soc.* **2020**, *2020*, 1–14. [\[CrossRef\]](#)
59. Son, R.; Stacke, T.; Gayler, V.; Nabel, J.E.M.S.; Schnur, R.; Silva, L.A.; Mesa, C.R.; Winkler, A.; Hantson, S.; Zaehle, S.; et al. Integration of a deep-learning-based fire model into a global land surface model. *J. Adv. Model. Earth Syst.* **2023**.
60. Eddin, M.H.S.; Roscher, R.; Gall, J. Location-Aware Adaptive Normalization: A Deep Learning Approach for Wildfire Danger Forecasting. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–18. [\[CrossRef\]](#)
61. Maffei, C.; Menenti, M. Predicting forest fires burned area and rate of spread from pre-fire multispectral satellite measurements. *ISPRS J. Photogramm. Remote Sens.* **2019**, *158*, 263–278. [\[CrossRef\]](#)
62. Oliveira, S.; Oehler, F.; San-Miguel-Ayanz, J.; Camia, A.; Pereira, J.M. Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *For. Ecol. Manag.* **2012**, *275*, 117–129. [\[CrossRef\]](#)
63. Gigović, L.; Pourghasemi, H.R.; Drobniak, S.; Bai, S. Testing a New Ensemble Model Based on SVM and Random Forest in Forest Fire Susceptibility Assessment and Its Mapping in Serbia's Tara National Park. *Forests* **2019**, *10*, 408. [\[CrossRef\]](#)
64. Guo, H.; Yu, W. Study weather grade prediction model of forest-fire risk in Hunan province. *J. Central South Univ. For. Technol.* **2016**, *36*, 44–47.
65. Weiss, D.J.; Nelson, A.; Gibson, H.S.; Temperley, W.; Peedell, S.; Lieber, A.; Hancher, M.; Poyart, E.; Belchior, S.; Fullman, N.; et al. A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature* **2018**, *553*, 333–336. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Su, Z.; Hu, H.; Wang, G.; Ma, Y.; Yang, X.; Guo, F. Using GIS and Random Forests to identify fire drivers in a forest city, Yichun, China. *Geomat. Nat. Hazards Risk* **2018**, *9*, 1207–1229. [\[CrossRef\]](#)

67. Zhang, F.; Yang, X. Improving land cover classification in an urbanized coastal area by random forests: The role of variable selection. *Remote Sens. Environ.* **2020**, *251*, 112105. [[CrossRef](#)]
68. Naderpour, M.; Rizeei, H.M.; Khakzad, N.; Pradhan, B. Forest fire induced Natech risk assessment: A survey of geospatial technologies. *Reliab. Eng. Syst. Saf.* **2019**, *191*, 106558. [[CrossRef](#)]
69. Sachdeva, S.; Bhatia, T.; Verma, A.K. GIS-based evolutionary optimized Gradient Boosted Decision Trees for forest fire susceptibility mapping. *Nat. Hazards* **2018**, *92*, 1399–1418. [[CrossRef](#)]
70. Miller, L.; Zhu, L.; Yebra, M.; Rüdiger, C.; Webb, G.I. Multi-modal temporal CNNs for live fuel moisture content estimation. *Environ. Model. Softw.* **2022**, *156*, 105467. [[CrossRef](#)]
71. Sevinc, V.; Kucuk, O.; Goltas, M. A Bayesian network model for prediction and analysis of possible forest fire causes. *For. Ecol. Manag.* **2019**, *457*, 117723. [[CrossRef](#)]
72. Abram, N.J.; Henley, B.J.; Gupta, A.S.; Lippmann, T.J.R.; Clarke, H.; Dowdy, A.J.; Sharples, J.J.; Nolan, R.H.; Zhang, T.; Wooster, M.J.; et al. Connections of climate change and variability to large and extreme forest fires in southeast Australia. *Commun. Earth Environ.* **2021**, *2*, 8. [[CrossRef](#)]
73. Boubeta, M.; Lombardía, M.J.; Marey-Pérez, M.F.; Morales, D. Prediction of forest fires occurrences with area-level Poisson mixed models. *J. Environ. Manag.* **2015**, *154*, 151–158. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.