forests

Article

Prediction of Wildfire Occurrence in the Southern Forest Regions of China in the Future Scenario

Jing Li [1,2,3], Duan Huang [1,4,*], Beiping Long [1,5], Yakui Shao [6], Mengwei Xiao [1,7], Linhao Sun [8], Xusheng Li [9], Aiai Wang [10], Xuanchi Chen [11] and Weike Li [12]

[1] Key Laboratory of Mine Environmental Monitoring and Improving Around Poyang Lake of Ministry of Natural Resources, East China University of Technology, Nanchang 330013, China; lijing@csuft.edu.cn (J.L.); jxsdzjdxdd@163.com (B.L.); 2022120388@ecut.edu.cn (M.X.)
[2] Wenzhou Key Laboratory of Resource Plant Innovation and Utilization, Zhejiang Institute of Subtropical Crops, Zhejiang Academy of Agricultural Sciences, Wenzhou 325005, China
[3] School of Forestry, Central South University of Forestry and Technology, Changsha 410004, China
[4] School of Surveying and Geoinformation Engineering, East China University of Technology, Nanchang 330013, China
[5] Jiangxi Provincial Geological Bureau Geographic Information Engineering Brigade, Nanchang 330038, China
[6] Precision Forestry Key Laboratory of Beijing, Beijing Forestry University, Beijing 100083, China; syk227816_gis@bjfu.edu.cn
[7] Jiangxi Key Laboratory of Watershed Ecological Process and Information (Platform No. 2023SSY01051), East China University of Technology, Nanchang 330013, China
[8] China Key Laboratory of Forestry Intelligent Monitoring and Information Technology Research of Zhejiang Province, Zhejiang A & F University, Hangzhou 311300, China; acesunlh@126.com
[9] Tianjin Centre of Geological Survey, China Geological Survey, Tianjin 300170, China; saintlxs@foxmail.com
[10] School of Geographical Sciences, Harbin Normal University, Harbin 150028, China; wangaiai0420@163.com
[11] School of Water Conservancy and Transportation, Zhengzhou University, Zhengzhou 450001, China; 202011221010277@gs.zzu.edu.cn
[12] Ecology and Nature Conservation Institute, Chinese Academy of Forestry, Beijing 100091, China; 13020020455@163.com
* Correspondence: huangduan@ecut.edu.cn

Abstract: In the context of global climate warming, climate change is subtly reshaping the patterns of wildfires. Therefore, it is particularly urgent to conduct in-depth research on climate change, wildfires, and their management strategies. This study relies on detailed fire point data from 2001 to 2020, skillfully incorporating a spatial autocorrelation analysis to uncover the mysteries of spatial heterogeneity, while comprehensively considering the influences of multiple factors such as climate, terrain, vegetation, and socioeconomic conditions. To simulate fire conditions under future climates, we adopted the BCC-CSM2-MR climate model, presetting temperature and precipitation data for two scenarios: a sustainable low-development path and a high-conventional-development path. The core findings of the study include the following: (i) In terms of spatial heterogeneity exploration, global autocorrelation analysis reveals a striking pattern: within the southern forest region, 63 cities exhibiting a low–low correlation are tightly clustered in provinces such as Hubei, Anhui, and Zhejiang, while 48 cities with a high–high correlation are primarily distributed in Guangxi and Guangdong. Local autocorrelation analysis further refines this observation, indicating that 24 high–high correlated cities are highly concentrated in specific areas, 14 low–low correlated cities are located in Hainan, and there are only 3 sparsely distributed cities with a low–high correlation. (ii) During the model construction and validation process, this study innovatively adopted the LR-RF-SVM ensemble model, which demonstrated exceptional performance indicators: an accuracy of 91.97%, an AUC value of 97.09%, an F1 score of 92.13%, a precision of 90.75%, and a recall rate of 93.55%. These figures, significantly outperforming those of the single models SVM and RF, strongly validate the superiority of the ensemble learning approach. (iii) Regarding predictions under future climate scenarios, the research findings indicate that, compared to the current fire situation in southern forest areas, the spatial distribution of wildfires will exhibit a noticeable expansion trend. High-risk regions will not only encompass multiple cities in Hunan, Hubei, southern Anhui, all of Jiangxi, and Zhejiang but will also extend northward into southern forest areas that were previously considered low-risk,

check for updates

Citation: Li, J.; Huang, D.; Long, B.; Shao, Y.; Xiao, M.; Sun, L.; Li, X.; Wang, A.; Chen, X.; Li, W. Prediction of Wildfire Occurrence in the Southern Forest Regions of China in the Future Scenario. *Forests* **2024**, *15*, 2029. https://doi.org/10.3390/f15112029

Academic Editor: Paul-Antoine Santoni

Received: 8 October 2024
Revised: 7 November 2024
Accepted: 12 November 2024
Published: 18 November 2024

suggesting a gradual northward spread of fire risk. Notably, despite the relatively lower fire risk in some areas of Fujian Province under the SS585 scenario, overall, the probability of wildfires occurring in 2090 is slightly higher than that in 2030, further highlighting the persistent intensification of forest fire risk due to climate change.

Keywords: global climate warming; wildfires; spatial heterogeneity; ensemble model (LR-RF-SVM); future climate scenarios

1. Introduction

Forests, which are among the planet's most essential ecosystems, offer critical habitats for biodiversity and are indispensable for human economic growth, ecological stability, and climate regulation [1–4]. However, wildfires, a highly destructive natural disaster, pose a significant threat to forest conservation [5–7]. Particularly in southern China, where a complex interplay of climate, topography, vegetation structure, and human activities increases both the frequency and severity of wildfires, the ecological and economic challenges are mounting. With the intensification of global climate change, the incidence of and areas affected by wildfires have significantly risen. Notably, the proportion of major and catastrophic fires is progressively increasing, exacerbating global environmental challenges [8–10]. The exacerbation of forest fire risks due to global warming and the El Niño phenomenon has altered fire occurrence patterns and behavior [11,12].

In the past century, global warming has become a widely accepted phenomenon. According to the latest Fifth Assessment Report from the Intergovernmental Panel on Climate Change (IPCC), both the land and oceans have experienced significant average temperature increases over the past 100 years, with this warming trend expected to continue in the future [13]. This rise in temperatures has led to an extended growing season in forested areas, resulting in more vigorous vegetation growth. However, it has also made these areas drier and increased the accumulation of fuels [14]. This drier environment provides favorable conditions for wildfires, making them more likely to ignite and spread during warmer seasons. Changes in precipitation also significantly impact forest fire occurrence. On the one hand, reduced rainfall exacerbates soil and vegetation dryness, increasing fire risk. On the other hand, alterations in precipitation patterns, such as an increase in extreme rainfall events, can have complex effects on fire occurrence and spread. For example, the rapid drying of forests following extreme precipitation events can create highly flammable conditions [15,16].

Machine learning has demonstrated significant advantages in forest fire prediction. It efficiently processes large volumes of meteorological, geographical, and environmental data, continually improving prediction accuracy and providing robust support for disaster prevention and mitigation [7,17,18]. However, machine learning also faces challenges, such as a limited ability to recognize complex patterns, the need for extensive training data, and poor model interpretability [19–21]. Ensemble models significantly enhance overall prediction accuracy and robustness by combining the predictions of multiple base learners. Different models often excel at capturing various aspects of the data, and ensemble methods integrate these strengths to reduce the risk of overfitting associated with individual models and decrease prediction variance. This approach not only improves adaptability to diverse data but also effectively handles complex prediction tasks, delivering more stable and reliable results.

In response to the anticipated increase in the frequency and intensity of wildfires in southern China's forest regions due to climate change, this study aims to develop an accurate prediction model for fire risk zoning and early warning. By integrating detailed fire point data from 2001 to 2020, meteorological data, historical fire records, and GIS technology, we hypothesize that the model will effectively capture the spatial heterogeneity of wildfires, comprehensively assess the impacts of climate, terrain, vegetation, and socioeconomic conditions on wildfire occurrence, and simulate fire scenarios under various future climate conditions. The ultimate goal is to formulate targeted fire prevention strategies to support sustainable regional management and mitigate threats to ecological safety and biodiversity. This hypothesis provides a clear focus for the study and outlines the expected relationships or outcomes being investigated, enhancing the clarity and direction of the research.

2. Materials and Methods

2.1. The Study Area

As illustrated in Figure 1, the southern forest region of China includes the provinces of Hubei, Hunan, Guangdong, Fujian, Guangxi, Jiangxi, Anhui, Zhejiang, and Hainan. Spanning approximately 1.386 million km^2, this vast area is located in central and eastern China. This region is characterized by a warm and humid climate with ample rainfall, influenced by temperate and subtropical monsoon climates, providing optimal conditions for plant growth. The terrain is diverse, featuring plateaus, mountains, hills, and plains, which offer varied site conditions for forest development. The region also has a range of soil types, including red soil, yellow soil, and brick-red soil, which supply essential nutrients for forest growth. Additionally, the region benefits from an extensive water network, with numerous rivers and lakes providing adequate moisture and promoting material cycling and energy flow. Despite these favorable conditions, the southern forest region faces severe forest fire threats. Fires can have devastating impacts on the fragile ecosystem and pose significant risks to local socio-economic conditions, including potential casualties, property damage, and long-term challenges in ecological restoration.

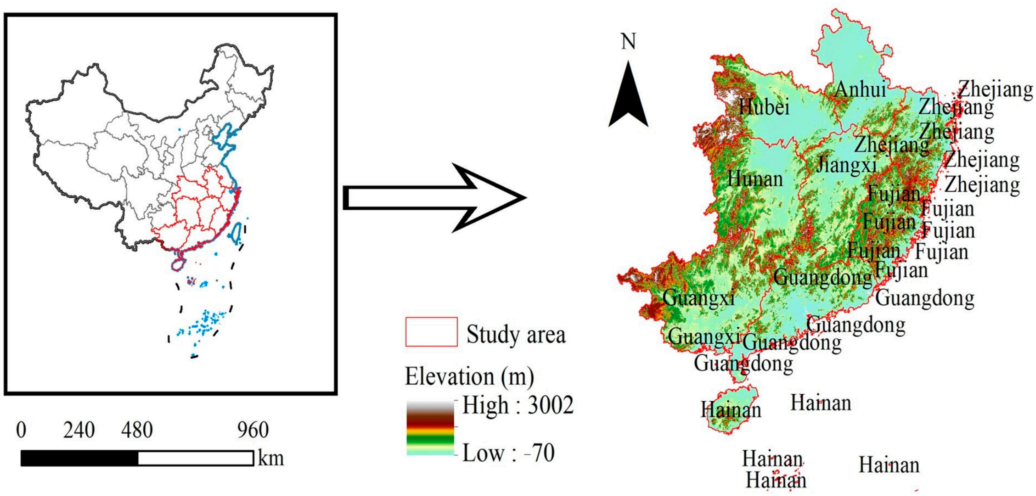

Figure 1. The map of the southern forest regions (The blue line represents the coastline, and the dashed line represents the Nine-Dashed Line).

2.2. Data and Methods

In our analysis, we utilized twenty years of MODIS fire data alongside forest type data, current and future climate data, population data, and GDP data. The Moderate-Resolution Imaging Spectroradiometer (MODIS), aboard NASA's Terra and Aqua satellites, is a pivotal tool for monitoring environmental phenomena, including fire events, providing near-daily global coverage with a high spatial resolution (https://firms.modaps.eosdis.nasa.gov/active_fire/, accessed on 10 May 2022) [22–24]. Regarding the utilization of MODIS fire data for each specific fire event, it is important to clarify that MODIS captures various data points for each fire, contingent on its duration. To ensure precision and relevance, we did not consider all recorded MODIS data for each fire incident. Instead, we focused on fire points with a confidence level exceeding 80%, as this threshold guaranteed the data were accurate. Additionally, to maintain the reliability of the data, fire points located outside forest cover areas were excluded from our analysis. The selected fire data were then divided into training and validation sets in a 70:30 ratio to facilitate model development and validation. This approach allowed us to harness the valuable insights offered by the MODIS fire data into fire dynamics and their environmental impacts, while ensuring the accuracy and reliability of our analysis.

Data on forest types, digital elevation models, and GDP were obtained from the Chinese Academy of Sciences Resource Environment and Data Center (http://www.resdc.cn/, accessed on 8 October 2023), featuring a resolution of 1000 m [25,26]. Datasets for roads and residential areas were retrieved from the National Geographic Information Resource Catalog Service System (https://www.webmap.cn, accessed on 23 October 2023). Temperature data came from He et al. [27], who based their study on meteorological station data, categorizing them into monthly mean temperatures, maximum temperatures, and minimum temperatures. A high-resolution monthly temperature dataset, with a resolution of 1 km, was created using Gaussian Process Regression (GPR). Precipitation data were sourced from Qu et al. [28], who performed an interpolation using precipitation data from over 2400 meteorological stations to produce a monthly precipitation dataset with a resolution of 1 km.

Data for various carbon emission scenarios were sourced from the WorldClim research website (https://www.worldclim.org/, accessed on 5 May 2021). The BCC-CSM2-MR climate model, which has a resolution of 2.5 min, was employed. It demonstrated a correlation coefficient of 0.86 for temperature and 0.73 for precipitation when compared to observed values from 1850 to 2005, indicating its strong simulation capabilities [29]. The BCC-CSM2-MR model's two future scenarios, SSP126 (sustainable low-development pathway) and SSP585 (high-conventional-development pathway), were used to simulate forest fire scenarios for the periods 2021–2040 and 2081–2100. The data were resampled to a spatial resolution of 1 km to accommodate different future climate scenarios.

2.2.1. Technical Workflow

The detailed technical workflow of this study is outlined in Figure 2 and involves the following steps:

(i) Analysis of spatial heterogeneity characteristics: Advanced spatial autocorrelation methods are employed to analyze forest fire data from the past 20 years. This step aims to reveal the spatial distribution patterns, clustering characteristics, and potential spatial association patterns of fire activities, laying a solid foundation for subsequent comprehensive analysis and prediction.

(ii) Integration of multisource data and spatial distribution analysis: Using MODIS fire data from 2001 to 2020, this study integrates various data dimensions, including meteorological conditions (temperature, humidity, wind speed, etc.), vegetation types and coverage, topographical features (elevation, slope, aspect, etc.), and human activity factors (population density, agricultural activities, tourism development, etc.). A comprehensive analysis of the spatial distribution of wildfires under the current climatic conditions in China is conducted, and a predictive model is developed to better understand future fire occurrence trends.

(iii) Ensemble learning model development and performance evaluation: To enhance prediction accuracy, an optimal ensemble learning model is designed and constructed. This model integrates various influencing factors to accurately predict forest fire occurrences. Its performance is thoroughly assessed using internationally recognized metrics, including recall, F1 score, accuracy, and AUC (Area Under the Curve), to ensure reliability and effectiveness in practical applications.

(iv) Forecasting forest fires based on future change scenarios: Utilizing the advanced BCC-CSM2-MR climate model, two representative greenhouse gas emission scenarios—SSP126 (low-emission scenario) and SSP585 (high-emission scenario)—are selected. The optimal ensemble learning model is employed to predict and map the distribution of forest fire risks in the southern forest region under future climatic changes across various timeframes. This approach provides a scientific foundation for forestry management and fire prevention, offering essential decision support for tackling forest fire challenges amid global climate change.

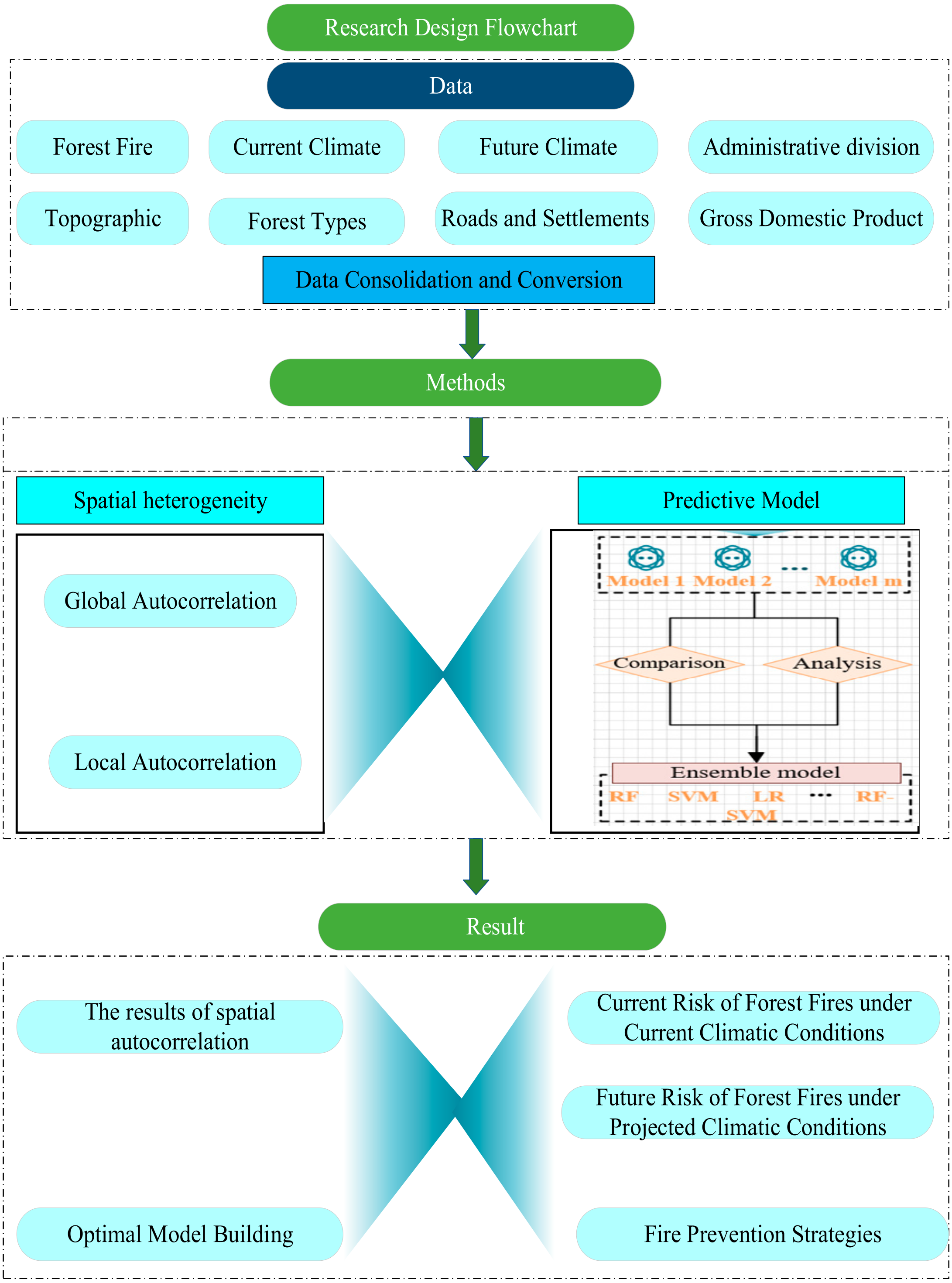

Figure 2. Technical roadmap of this study.

2.2.2. Spatial Autocorrelation

Spatial autocorrelation is a core concept in spatial statistics, used to examine the interdependencies of geographic phenomena or attributes of their spatial distribution. Specifically, it refers to the statistical correlation between observed values at different locations within a geographic space, indicating whether the value at one location is influenced by values at nearby locations [30,31]. The global Moran's index was computed for each year from 2001 to 2020 to assess this spatial correlation [30,32].

$$I = \frac{n \times \sum_i^n \sum_j^n w_{ij} \times (x_i - \bar{x})(x_j - \bar{x})}{\left(\sum_i^n \sum_j^n w_{ij}\right) \times \sum_i^n (x_i - \bar{x})^2}, \tag{1}$$

In the context of calculating the global Moran index, "I", represents forest fire point samples, where "n" denotes the total number of such samples, and "w_{ij}" represents the weights assigned to the forest fire points from i to j. The expressions $x_i - \bar{x}$ and $x_j - \bar{x}$ signify the deviations of the forest fire points i and j from the mean deviation of all forest fire points, respectively.

The standardized Z value featured in Equation (2) serves as a metric to quantify the degree of spatial autocorrelation present in the dataset. It measures the strength of the spatial correlation or clustering among the forest fire points, indicating the level of spatial autocorrelation between them.

$$Z = \frac{I - E(I)}{STD(I)}, \tag{2}$$

In this context, $E(I)$ represents the expected value of I. When the absolute value of the standardized Z score, denoted as $|Z|$, equals 2.54, it corresponds to a confidence level of 0.01. Similarly, a $|Z|$ value of 1.96 corresponds to a confidence level of 0.05. These $|Z|$ values, whether 1.96 or 2.54, indicate the presence of a significant spatial autocorrelation among forest fire occurrences. Specifically, if $|Z|$ is greater than 1.96 or 2.54, this suggests that the forest fire points indicate a pattern of spatial concentration. Conversely, if $|Z|$ is less than -1.96 or -2.54, it indicates that the forest fire points in the study area are discrete, suggesting a pattern of spatial dispersion.

Local indicators of spatial association (LISA) play a crucial role in quantifying the degree of spatial correlation between neighboring regions and assessing the statistical significance of these correlations through significance testing. Essentially, LISA serves as a local version of the global Moran's I index, providing a valuable tool for analyzing spatial data at a finer resolution [33,34].

2.2.3. Construction of Predictive Models

Logistic regression is a widely used statistical method for binary classification problems. By combining a linear regression model with the Sigmoid function, it maps input features to the $[0,\ 1]$ interval, thereby estimating the probability of an event occurring. This method is straightforward and efficient, with strong interpretability [35,36]. The formula for logistic regression is as follows:

$$LogitP = ln\ [P/(1-P)], \tag{3}$$

Logistic regression establishes a specific mathematical relationship (a monotone differentiable function) to link the actual classification results with the predicted values of a linear model. In this example, we focus on the probability P of a forest fire occurring. By calculating $LogitP$, we can understand how likely a fire is to occur relative to the likelihood of it not occurring. The value of $LogitP$ increases as P increases, indicating a positive correlation between them.

The random forest model, an efficient learning method, is based on the construction of multiple decision trees and incorporates dual randomness in both its samples and features. This design not only effectively reduces the risk of model overfitting but also significantly enhances prediction accuracy. During the prediction process, random forests aggregate the predictions from each decision tree, using methods such as voting or averaging to derive the final decision. This approach ensures robustness and reliability in the predictions [37–39].

$$h(x) = \frac{1}{T}\sum_{t=1}^{T}\{h(x,\theta_t)\}, \tag{4}$$

where T denotes the number of decision trees, θ_t represents an independent and identically distributed random vector, and x is the input vector. The final prediction is obtained by calculating the mean of the outputs from each regression subtree, denoted as $\{h(x,\theta_t)\}$. The predictive strength of the model ultimately hinges on the number of random features selected and the total number of trees, T, as these factors collectively determine the model's capacity to generalize and make accurate predictions. In essence, the interplay between the quantity of random features and decision trees is crucial in shaping the overall predictive power and effectiveness of the random forest model.

SVM (Support Vector Machine) is rooted in statistical learning theory and functions as a generalized linear classifier for binary classification within the framework of supervised learning [40,41]. Its decision boundary is determined by the maximum-margin hyperplane derived from the learning samples. When dealing with linearly inseparable problems, SVM utilizes a mechanism to transform data from a low-dimensional space into a high-dimensional space, relying on suitable kernel functions for this mapping process. Frequently employed kernel functions in SVM include the linear kernel, polynomial kernel, and sigmoid kernel, among others [42].The formula is as follows.

Linear kernel:

$$K(x_i, x_j) = < x_i, x_j >; \tag{5}$$

Polynomial kernel (poly):

$$K\left(x_i, x_j\right) = (< x_i, x_j > +1)^q; \tag{6}$$

Radial basis function (RBF):

$$K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}. \tag{7}$$

Figure 3 outlines the detailed process. In the training phase, predictions are generated for each training sample (i.e., the i-th sample) using several base models, such as random forest (RF) and Support Vector Machine (SVM). The predictions are then combined, with the prediction from the j-th base model assigned as the j-th feature for the i-th sample in a new training set. This transformation creates a new feature space where each sample's features consist of the outputs from the various base models. Subsequently, a logistic regression model is trained on this newly constructed training set.

During the prediction phase, predictions for each sample in the test set are also made using all base models. These predictions serve as new features that are input into the logistic regression model to yield the final prediction results. This methodology represents an ensemble learning strategy that merges predictions from multiple base models, thereby improving the overall model's predictive performance and stability.

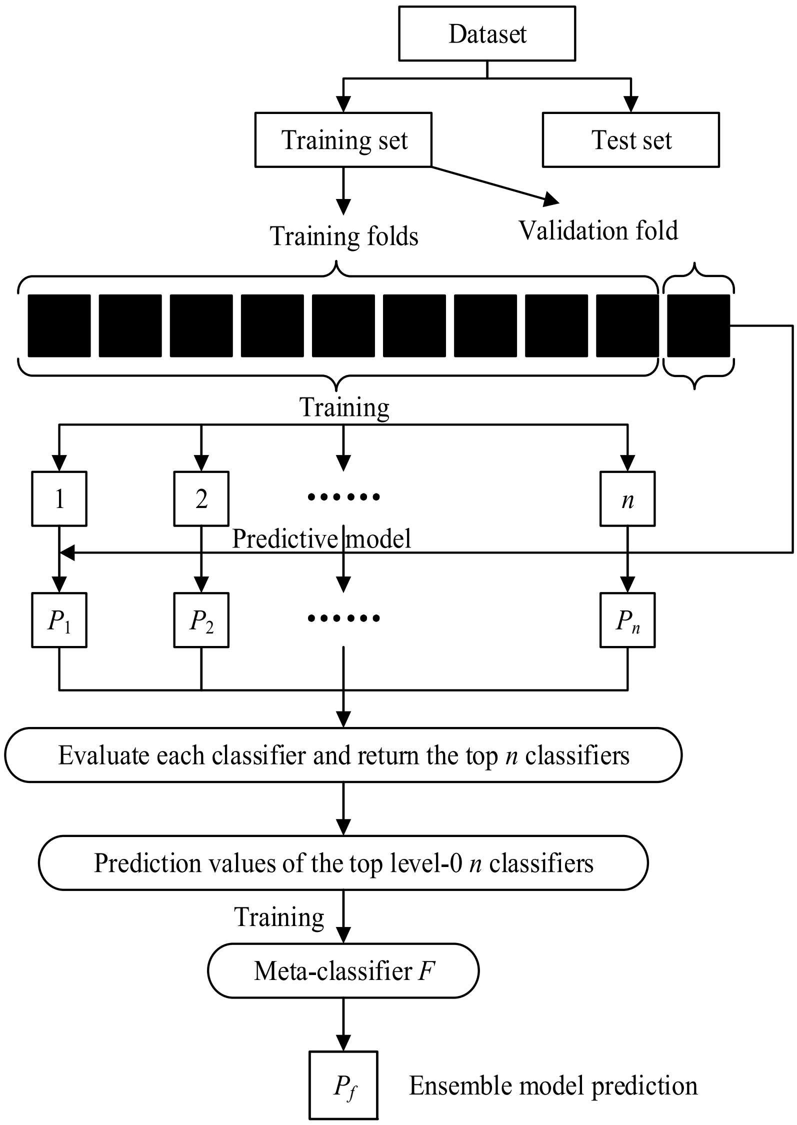

Figure 3. Schematic diagram of the model used in this study.

2.2.4. Evaluation of Model Performance

In forest fire risk prediction, commonly used evaluation metrics include accuracy, precision, recall, $F1$ score, and AUC (Area Under the Curve). These metrics quantify the performance of classification models from different perspectives, helping to comprehensively assess and understand the strengths and weaknesses of the model [43–45]. The formulas are presented as follows [46]:

$$Accuracy = (TP + TN)/(TP + FP + TN + FN), \quad (8)$$

$$Recall = TP/(TP + FN), \quad (9)$$

$$Precision = TP/(TP + FP) \quad (10)$$

$$F1 = 2\times(Precision \times Recall)/(Precision + Recall) \quad (11)$$

Accuracy measures the proportion of correct predictions made by the model, reflecting its overall average performance across all predictions. However, in imbalanced datasets like forest fire risk prediction, accuracy may not be a reliable metric. This is because a model that correctly predicts most non-fire samples might still achieve high accuracy, even if it fails to accurately predict fire events. Precision refers to the proportion of predicted fire cases that are actually fires. This metric is crucial for assessing the reliability of the model in predicting fires. Recall indicates the proportion of true fire incidents that the model successfully identifies, highlighting its sensitivity to fire occurrences. The *F*1 score, which represents the harmonic mean of precision and recall, effectively balances the trade-off between these metrics. In the context of predicting forest fire risks, the *F*1 score serves as a valuable performance indicator that takes into account both the model's reliability and its sensitivity. Finally, AUC (Area Under the Curve) represents the area under the Receiver Operating Characteristic (ROC) curve, which assesses the model's overall performance across different thresholds. A higher AUC value signifies that the model effectively differentiates between fire and non-fire samples across different thresholds. In this context, *TP* (True Positives) refers to the correctly identified fire instances, *TN* (True Negatives) indicates the accurately identified non-fire instances, *FP* (False Positives) represents the instances incorrectly classified as fires, and *FN* (False Negatives) denotes the cases mistakenly identified as non-fires.

3. Results

3.1. Distribution Map of Forest Fire Occurrences Based on Current Climate Conditions

Figure 4 depicts the forecasted risk of wildfires based on current climatic conditions, highlighting key areas where fire incidents are expected to be notably concentrated throughout the year. These high-risk regions are predominantly found in specific cities, including Hezhou and Baise in the Guangxi Zhuang Autonomous Region, as well as Heyuan, Huizhou, and Shaoguan in Guangdong Province. Additional high-risk areas include Hengyang and Shaoyang in Hunan Province, Ganzhou and Ji'an in Jiangxi Province, and Fuzhou, Sanming, and Nanping in Fujian Province. These regions are characterized by complex terrain, variable climate patterns, and a high coverage of flammable vegetation. Additionally, the frequent human activities in these areas further exacerbate the risk of natural disasters such as fires. Consequently, it is imperative for these high-risk cities to implement enhanced fire prevention and control measures throughout the year. This involves enhancing the management of ignition sources, increasing public awareness about fire prevention, and refining fire warning and emergency response systems. Implementing these strategies can substantially mitigate the impact of fire incidents.

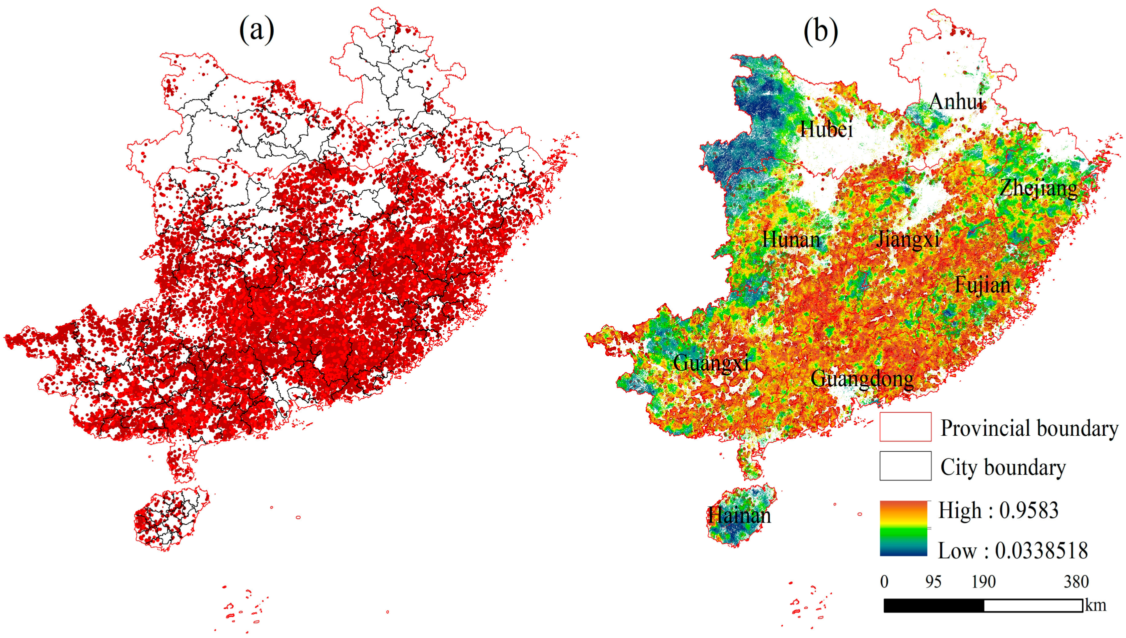

Figure 4. (**a**) Fire occurrences (2001–2020); (**b**) forest fire risk mapping using ensemble learning models under current climate conditions.

3.2. Evaluation of the Spatial Patterns Associated with Wildfires

Figure 5 depicts a global spatial autocorrelation analysis of the southern forest region, revealing the distribution traits and spatial configurations of various city types. A total of 63 cities were classified as the low–low (L-L) type, predominantly concentrated in the Hubei, Anhui, and Zhejiang provinces. This clustering suggests a comparatively low level of similarity or interconnection among the cities within these provinces.

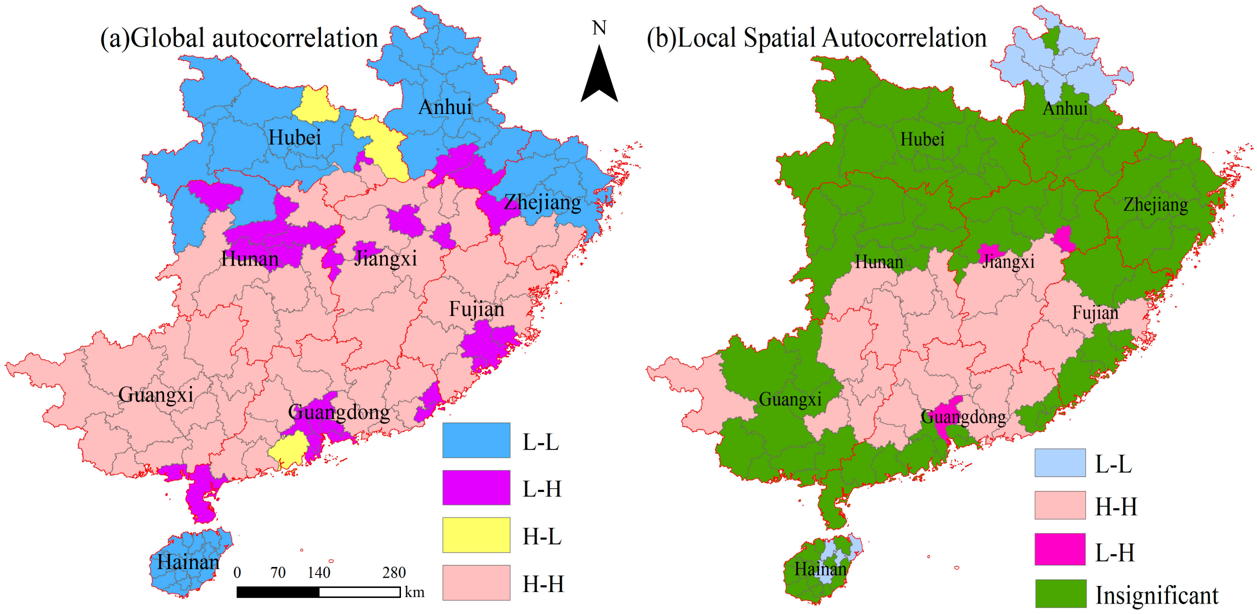

Figure 5. (**a**) The global Moran's I index for forest fire risk and (**b**) the aggregation chart of local indicators of spatial association (LISA) for forest fire risk levels.

On the other hand, 27 cities were identified as the low–high (L-H) type, scattered across Yiyang, Xiangtan, and Changsha in Hunan Province; Quanzhou, Xiamen, and Putian in Fujian Province; as well as Dongguan and Zhanjiang in Guangdong Province. These cities demonstrate spatial heterogeneity, displaying marked contrasts in certain characteristics when compared to their neighboring cities.

Conversely, 48 cities were categorized as the high–high (H-H) type, mainly situated in Hezhou, Baise, and Hechi within the Guangxi Zhuang Autonomous Region. Other significant clusters include Qingyuan, Heyuan, Huizhou, and Shaoguan in Guangdong Province; Fuzhou, Sanming, and Nanping in Fujian Province; and Ganzhou and Fuzhou in Jiangxi Province, and Shaoyang and Hengyang in Hunan Province. These cities exhibit a pronounced clustering effect, characterized by relatively high attributes and strong inter-city correlations.

3.3. Prediction Performance Evaluation

As shown in Figure 6, the study first assessed the predictive performance of the individual models, SVM and RF, separately. The SVM model demonstrated a reasonable level of accuracy, achieving an accuracy of 0.73, a precision of 0.70, a recall of 0.80, an F1 score of 0.74, and an AUC value of 0.80. These indicators suggest that while the SVM model has some strengths in its performance, there is still room for improvement. In contrast, the RF model showed a significant improvement in predictive performance, with an accuracy of 0.84, a precision of 0.82, a recall of 0.88, an F1 core of 0.85, and an AUC value of 0.92, demonstrating the RF model's stronger ability to handle complex data and classification tasks. However, by constructing the LR-RF-SVM ensemble learning model, we achieved a further leap in predictive performance. The ensemble model had an accuracy of 0.92, an AUC value of 0.97, an F1 score of 0.92, a precision of 0.91, and a recall of 0.94, all of which far exceeded the results obtained by the single models. These results fully demonstrate the significant advantage of ensemble learning in integrating the prediction results of multiple base models and improving overall performance. This research not only verifies the effectiveness of ensemble learning methods but also provides strong support and a reference for subsequent related research and applications. (Note: The accuracy results are rounded to two decimal places.)

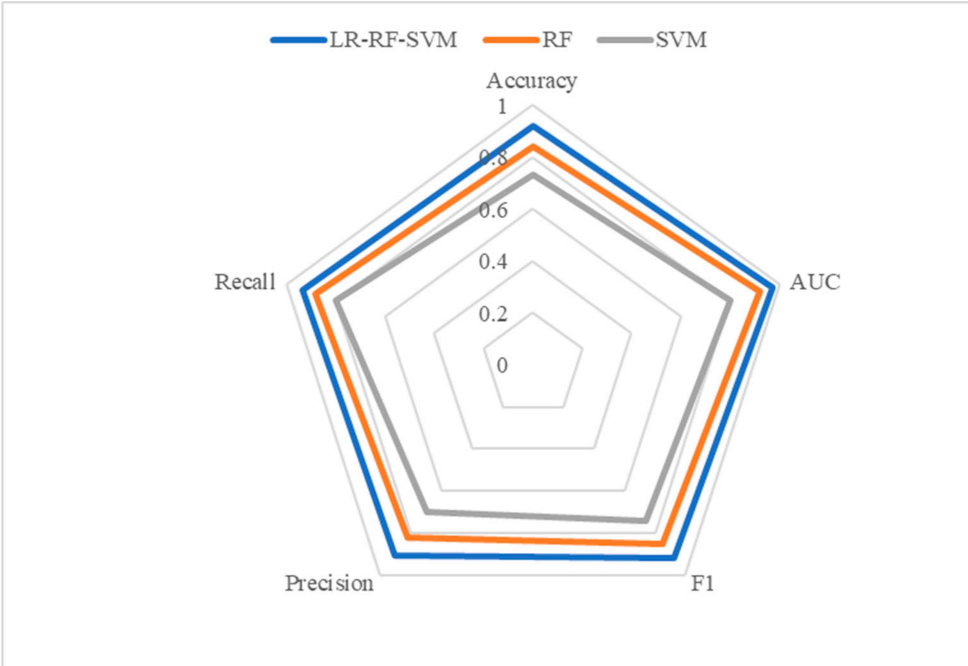

Figure 6. Evaluation charts for machine learning and ensemble learning.

3.4. Prediction and Zoning of Forest Fire Occurrence in Future Scenarios

Under the severe challenges posed by future climate scenarios, Figure 7 illustrates the unprecedented changes in the spatial distribution of wildfire occurrences across southern forest regions. High-occurrence areas persistently encompass numerous cities in Hunan, Hubei, southern Anhui, Jiangxi, and Zhejiang, and a notable trend emerges: wildfire occurrences are gradually spreading northward, posing significant threats to previously low-occurrence areas. In the SS585 scenario, some parts of Fujian Province experience a relative decrease in wildfire occurrences amidst a general upward trend, highlighting the complexity and intensification of forest fire threats driven by climate change. This necessitates a deeper understanding of climate change's impacts and an urgent reevaluation and adjustment of existing forest fire prevention strategies.

Figure 8, through detailed data comparisons and scenario simulations, further emphasizes the profound impact of climate change on the geographical distribution of forest fire occurrences, revealing the relative rate of change under different climate scenarios. This figure serves as a warning that, as climate conditions continue to evolve, areas once deemed low-occurrence may rapidly transition to high-occurrence zones. Consequently, forest fire prevention efforts must be increased to address these comprehensive and ever-changing challenges, ensuring the safety of forest resources and human life and property.

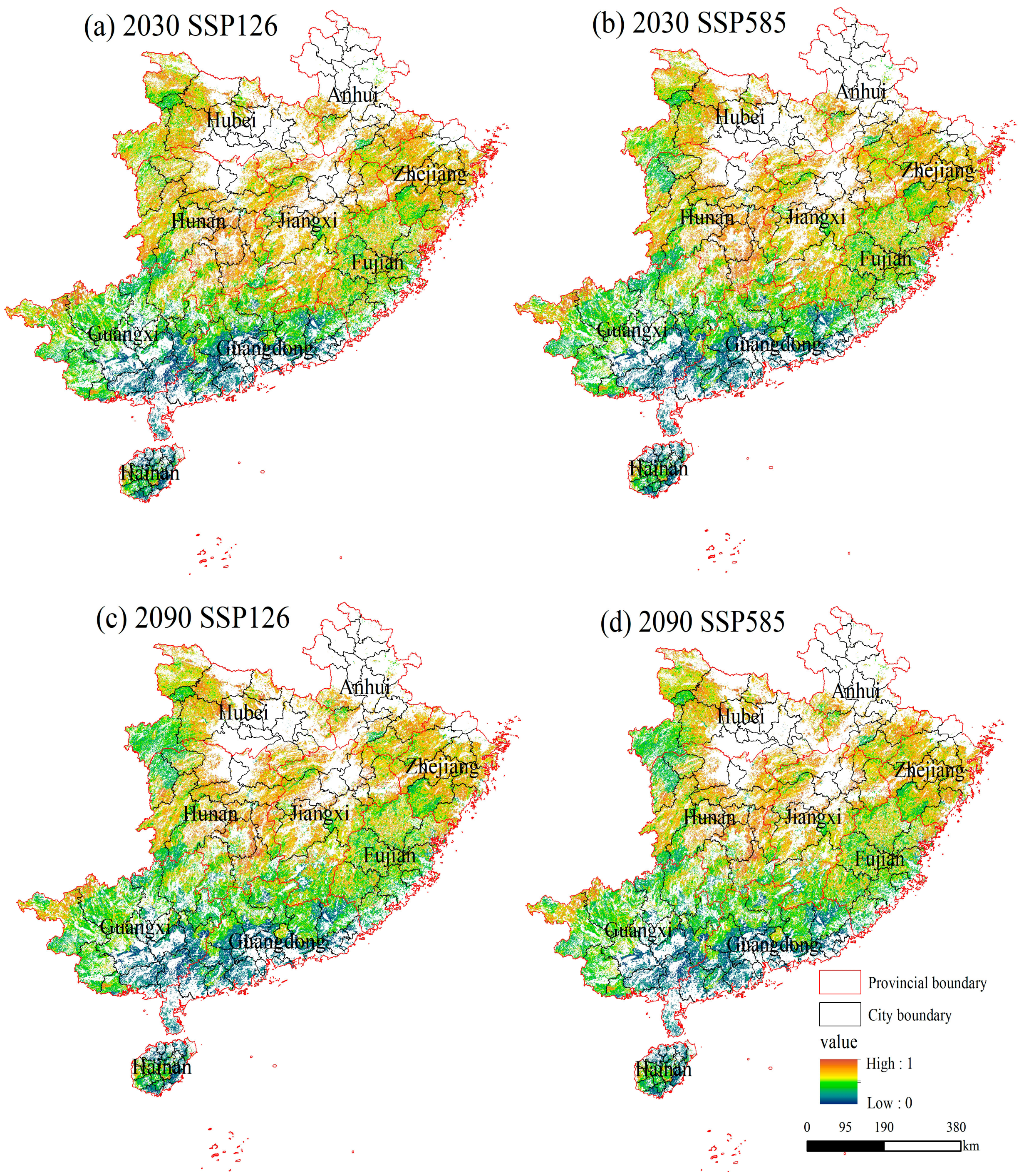

Figure 7. Predictions of forest fire occurrences in the southern forest region of China, utilizing the BCC-CSM2-MR scenarios for the years 2030 to 2090 (the lighter the color, the lower the probability of occurrence; the darker the color, the higher the probability).

In summary, the quantitative analysis presented confirms that the spatial distribution of forest fires will indeed exhibit a noticeable expansion trend in terms of occurrence. This trend is manifested not only in the persistence and expansion of high-occurrence areas but also in the transformation of low-occurrence areas into high-occurrence ones. To effectively tackle this trend, it is imperative to enhance research on the impacts of climate change on forest fires, optimize and adjust forest fire prevention strategies, and increase investment in fire prevention measures.

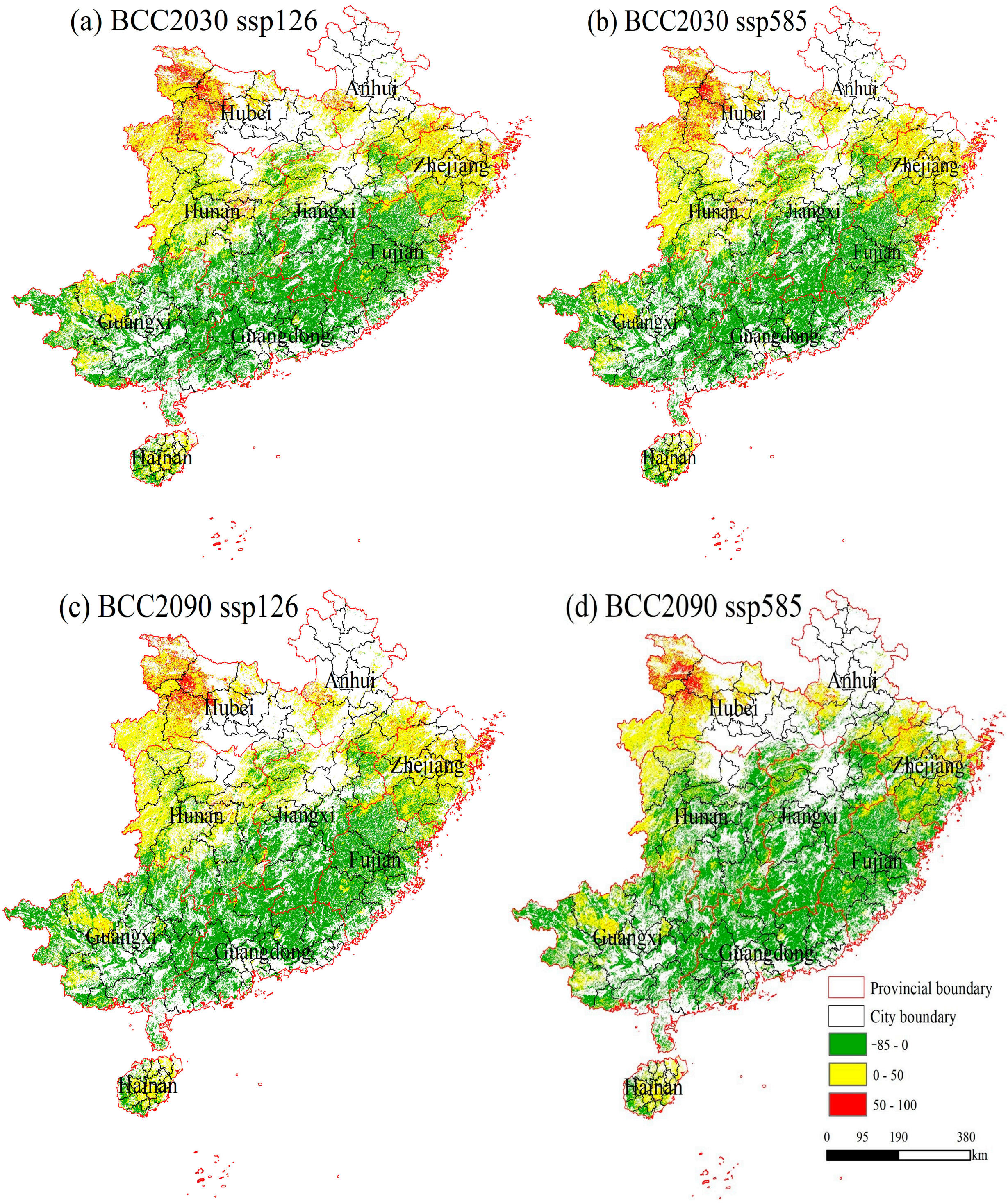

Figure 8. Evaluation of relative changes in forest fire occurrences based on current and future climate scenarios (green represents negative values).

To effectively respond to this challenge, a series of integrated and forward-looking measures must be adopted. First and foremost, strengthening monitoring and early warning systems is crucial. Leveraging modern technological tools, such as remote sensing satellites, drone monitoring, and big data analysis, can help build efficient and precise fire monitoring and early warning systems, improving the accuracy of early detection and the timeliness of warnings [47,48].

Second, optimizing forest fire prevention infrastructure is an urgent priority. This includes establishing firebreaks, constructing firewater pools, providing advanced firefighting gear, and enhancing the training and capabilities of forest fire brigades, all aimed at improving the efficiency and effectiveness of fire suppression efforts. Additionally, increasing public awareness of fire prevention is essential. Through education, training, and drills, public understanding of and involvement in forest fire prevention can be strengthened, fostering a culture of collective responsibility for fire safety [49].

Lastly, establishing cross-regional forest fire prevention mechanisms is key to achieving resource-sharing and coordinated responses to large-scale fire events. Strengthening regional cooperation and coordination can help form an interconnected forest fire prevention network, enabling swift, effective firefighting and post-disaster recovery.

4. Discussion

The integration of machine learning, particularly ensemble learning, with detailed climate change scenario analysis represents a significant advancement in the field of forest fire prediction. Our study demonstrates how this combined approach can effectively tackle the complexities and uncertainties associated with forest fire risks under climate change. By leveraging the strengths of multiple models within an ensemble framework, we achieved enhanced prediction stability and accuracy, echoing findings from similar studies that highlight the benefits of ensemble methods in environmental forecasting [29,50,51].

A notable strength of our research lies in the comprehensive consideration of climate change's impacts on forest fire risk through detailed scenario analyses. This approach provided the model with rich, dynamic input data, enabling it to capture the nuances of climate–fire interactions. As a result, we not only identified key drivers of wildfires but also significantly improved prediction accuracy compared to traditional methods. This aligns with recent research that emphasizes the importance of incorporating climate change projections into fire risk assessments to ensure the relevance and effectiveness of management strategies [31,52]. However, it is crucial to acknowledge that while our study contributes to the understanding of forest fire risks, it is not without limitations. Spatial autocorrelation analysis, for instance, revealed the clustering and heterogeneity of cities in the southern forest region, highlighting potential interactions and influences. Yet, as noted in the literature, spatial autocorrelation may not fully explain the specific mechanisms underlying these relationships, necessitating further exploration through complementary research methods and field investigations [31,52].

In comparison to other recent studies on forest fire prediction, our work stands out in its innovative use of ensemble learning combined with detailed climate scenarios. While some studies have focused on the application of individual machine learning models [53], or the integration of remote sensing technologies [54], our approach goes a step further by combining multiple models within an ensemble framework and incorporating detailed climate change projections. This holistic approach not only improves prediction accuracy but also provides a more robust framework for decision-making in the context of climate change.

The significance of our findings is further underscored by the urgent need to adapt and optimize fire management strategies in the face of climate change. Our research highlights the importance of designing dense vegetation and open forest ecological belts in high-risk areas, establishing comprehensive fire alarm systems, and repairing and improving existing fire prevention infrastructure. These measures are crucial for ensuring timely responses to fires and for minimizing damage, as evidenced by their effectiveness in other regions [49,54,55].

Moreover, our study emphasizes the need for innovative solutions in fire detection and response. Advanced detection systems, such as those integrating IoT technology and WSNs, have shown promise in providing early fire warnings and accurately locating fires [54,55]. These technologies complement traditional methods and enhance our ability to respond rapidly and effectively to fire events.

Our research contributes to the growing literature on forest fire prediction under climate change by demonstrating the potential of combining ensemble learning with detailed climate scenario analyses. While our findings are promising, they also highlight the need for continued research and innovation in this field. By adopting a comprehensive and forward-thinking approach, we can effectively reduce fire risks, protect valuable forest resources, and ensure a greener, safer, and more sustainable planet for future generations.

5. Conclusions

This research has made significant strides in understanding the intricate relationship between climate change, wildfires, and their management strategies.

(i) By leveraging detailed fire point data spanning two decades and incorporating spatial autocorrelation analysis, we uncovered notable patterns of spatial heterogeneity in forest fire occurrences.

(ii) Our findings reveal distinct clusters of cities with varying levels of fire risk within the southern forest region. Furthermore, the innovative use of the LR-RF-SVM ensemble model proved highly effective, surpassing the performance of individual models in predicting wildfires. This underscores the advantages of integrating multiple machine learning techniques to enhance prediction accuracy.

(iii) Looking to the future, our predictions based on two climate change scenarios indicate a concerning trend of expanding forest fire risk, particularly in previously low-risk areas. This highlights the urgent need for proactive management strategies to mitigate the impacts of climate change on forest fire occurrence.

Overall, this study provides valuable insights for decision-makers and researchers working to address the challenges posed by climate change and wildfires.

Author Contributions: J.L. was responsible for the research design, data analysis and the manuscript. D.H. conducted the experiments and drafted the initial manuscript. B.L. and M.X. contributed to data collection and the literature review, while Y.S. handled data processing and the interpretation of results. L.S. assisted with the experimental design and data analysis. X.L. provided methodological support and technical guidance, and X.C. participated in the discussion of the results and manuscript review. A.W. and W.L. conducted the literature research and assisted in writing the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Open Fund of the Key Laboratory of Mine Environmental Monitoring and Improvement around Poyang Lake, Ministry of Natural Resources (MEMI-2023-02, MEMI-2021-2022-16), the Jiangxi Provincial Natural Science Foundation (20224BAB213038), and the Ph.D. Project of East China University of Technology (DHBK2019179). Additionally, funding was provided by the Postgraduate Innovation Fund of Jiangxi Province (YC2023-S586).

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Acknowledgments: We anticipate being grateful to our colleagues and peers for their insightful discussions and feedback, which we expect will significantly enrich our work. Additionally, we look forward to thanking the participants of our study for their time and willingness to contribute to this research in the future.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ma, J.; Li, J.; Wu, W.; Liu, J. Global forest fragmentation change from 2000 to 2020. *Nat. Commun.* **2023**, *14*, 3752. [CrossRef] [PubMed]
2. Qiu, Z.; Feng, Z.; Song, Y.; Li, M.; Zhang, P. Carbon sequestration potential of forest vegetation in China from 2003 to 2050: Predicting forest vegetation growth based on climate and the environment. *J. Clean. Prod.* **2020**, *252*, 119715. [CrossRef]
3. Agnoletti, M.; Piras, F.; Venturi, M.; Santoro, A. Cultural values and forest dynamics: The Italian forests in the last 150 years. *For. Ecol. Manag.* **2022**, *503*, 119655. [CrossRef]
4. Zhang, H.; Feng, Z.; Shen, C.; Li, Y.; Feng, Z.; Zeng, W.; Huang, G. Relationship between the geographical environment and the forest carbon sink capacity in China based on an individual-tree growth-rate model. *Ecol. Indic.* **2022**, *138*, 108814. [CrossRef]

5. Agbeshie, A.A.; Abugre, S.; Atta-Darkwa, T.; Awuah, R. A review of the effects of forest fire on soil properties. *J. For. Res.* **2022**, *33*, 1419–1441. [CrossRef]
6. Ying, L.; Han, J.; Du, Y.; Shen, Z. Forest fire characteristics in China: Spatial patterns and determinants with thresholds. *For. Ecol. Manag.* **2018**, *424*, 345–354. [CrossRef]
7. Mohajane, M.; Costache, R.; Karimi, F.; Bao Pham, Q.; Essahlaoui, A.; Nguyen, H.; Laneve, G.; Oudija, F. Application of remote sensing and machine learning algorithms for forest fire mapping in a Mediterranean area. *Ecol. Indic.* **2021**, *129*, 107869. [CrossRef]
8. Shao, Y.; Fan, G.; Feng, Z.; Sun, L.; Yang, X.; Ma, T.; Li, X.; Fu, H.; Wang, A. Prediction of forest fire occurrence in China under climate change scenarios. *J. For. Res.* **2023**, *34*, 1217–1228. [CrossRef]
9. Liu, Z.; Yang, J.; Chang, Y.; Weisberg, P.J.; He, H.S. Spatial patterns and drivers of fire occurrence and its future trend under climate change in a boreal forest of Northeast China. *Glob. Change Biol.* **2012**, *18*, 2041–2056. [CrossRef]
10. Jones, M.W.; Abatzoglou, J.T.; Veraverbeke, S.; Andela, N.; Lasslop, G.; Forkel, M.; Smith, A.J.P.; Burton, C.; Betts, R.A.; van der Werf, G.R.; et al. Global and Regional Trends and Drivers of Fire Under Climate Change. *Rev. Geophys.* **2022**, *60*, e2020RG000726. [CrossRef]
11. Farfan, M.; Dominguez, C.; Espinoza, A.; Jaramillo, A.; Alcantara, C.; Maldonado, V.; Tovar, I.; Flamenco, A. Forest fire probability under ENSO conditions in a semi-arid region: A case study in Guanajuato. *Environ. Monit. Assess.* **2021**, *193*, 684. [CrossRef] [PubMed]
12. Bai, M.; Wang, X.; Yao, Q.; Fang, K. ENSO modulates interaction between forest insect and fire disturbances in China. *Nat. Hazard. Res.* **2022**, *2*, 138–146. [CrossRef]
13. Allen, M.; Dube, O.; Solecki, W.; Aragón-Durand, F.; Cramer, W.; Humphreys, S.; Kainuma, M. *Special Report: Global Warming of 1.5 °C*; Intergovernmental Panel on Climate Change (IPCC): Geneva, Switzerland, 2018; p. 677.
14. Brown, T.P.; Duff, T.J.; Inbar, A.; Lane, P.N.J.; Sheridan, G.J. Forest reorganisation effects on fuel moisture content can exceed changes due to climate warming in wet temperate forests. *Glob. Chang. Biol.* **2024**, *30*, e17023. [CrossRef] [PubMed]
15. Manuel Galván Ortiz, L.; Antonio Ávalos Lozano, J. Regional Climate Change as a Driver in Wildfires. In *Climate Change and Environmental Sustainability—Annual Volume 2024*; Sustainable Development; IntechOpen: London, UK, 2024.
16. Zivanovic, S.; Gocic, M. The Influence of Climatic Conditions on the Occurrence of Large Wildfires: A Case Study of Stara Planina Nature Park in 2007 and 2019. *J. Appl. Life Sci. Environ.* **2023**, *56*, 195–209. [CrossRef]
17. Kalantar, B.; Ueda, N.; Idrees, M.O.; Janizadeh, S.; Ahmadi, K.; Shabani, F. Forest Fire Susceptibility Prediction Based on Machine Learning Models with Resampling Algorithms on Remote Sensing Data. *Remote Sens.* **2020**, *12*, 3682. [CrossRef]
18. Tien Bui, D.; Le, H.V.; Hoang, N.-D. GIS-based spatial prediction of tropical forest fire danger using a new hybrid machine learning method. *Ecol. Inf.* **2018**, *48*, 104–116. [CrossRef]
19. Alkhatib, R.; Sahwan, W.; Alkhatieb, A.; Schütt, B. A Brief Review of Machine Learning Algorithms in Wildfires Science. *Appl. Sci.* **2023**, *13*, 8275. [CrossRef]
20. Mahesh, B. Machine Learning Algorithms—A Review. *Int. J. Sci. Res. (IJSR)* **2020**, *9*, 381–386. [CrossRef]
21. Chaudhary, S.; Yadav, S.; Kushwaha, S.; Shahi, S.R.P. A brief review of machine learning and its applications. *Samriddhi A J. Phys. Sci. Eng. Technol.* **2020**, *12*, 218–223.
22. Justice, C.O.; Giglio, L.; Korontzi, S.; Owens, J.; Morisette, J.T.; Roy, D.; Descloitres, J.; Alleaume, S.; Petitcolin, F.; Kaufman, Y. The MODIS fire products. *Remote Sens. Environ.* **2002**, *83*, 244–262. [CrossRef]
23. Hawbaker, T.J.; Radeloff, V.C.; Syphard, A.D.; Zhu, Z.; Stewart, S.I. Detection rates of the MODIS active fire product in the United States. *Remote Sens. Environ.* **2008**, *112*, 2656–2664. [CrossRef]
24. Hantson, S.; Padilla, M.; Corti, D.; Chuvieco, E. Strengths and weaknesses of MODIS hotspots to characterize global fire occurrence. *Remote Sens. Environ.* **2013**, *131*, 152–159. [CrossRef]
25. Yi, L.; Xiong, L.-Y.; Yang, X. Method of pixelizing GDP data based on the GIS. *J. Gansu Sci.* **2006**, *18*, 54–58.
26. Liu, H.; Jiang, D.; Yang, X.; Luo, C. Spatialization approach to 1 km grid GDP supported by remote sensing. *Geo-Inf. Sci.* **2005**, *7*, 120–123.
27. He, Q.; Wang, M.; Liu, K.; Li, K.; Jiang, Z. GPRChinaTemp1km: A high-resolution monthly air temperature data set for China (1951–2020) based on machine learning. *Earth Syst. Sci. Data* **2022**, *14*, 3273–3292. [CrossRef]
28. Qu, L.; Zhu, Q.; Zhu, C.; Zhang, J. Monthly precipitation data set with 1 km resolution in China from 1960 to 2020. *Sci. Data Bank* **2022**. [CrossRef]
29. 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. [CrossRef]
30. Chen, Y. An analytical process of spatial autocorrelation functions based on Moran's index. *PLoS ONE* **2021**, *16*, e0249589. [CrossRef]
31. Li, L.; Tang, H.; Lei, J.; Song, X. Spatial autocorrelation in land use type and ecosystem service value in Hainan Tropical Rain Forest National Park. *Ecol. Indic.* **2022**, *137*, 108727. [CrossRef]
32. Fu, W.J.; Jiang, P.K.; Zhou, G.M.; Zhao, K.L. Using Moran's I and GIS to study the spatial pattern of forest litter carbon density in a subtropical region of southeastern China. *Biogeosciences* **2014**, *11*, 2401–2409. [CrossRef]
33. Bivand, R.S.; Wong, D.W.S. Comparing implementations of global and local indicators of spatial association. *TEST* **2018**, *27*, 716–748. [CrossRef]

34. Song, W.; Wang, C.; Chen, W.; Zhang, X.; Li, H.; Li, J. Unlocking the spatial heterogeneous relationship between Per Capita GDP and nearby air quality using bivariate local indicator of spatial association. *Resour. Conserv. Recycl.* **2020**, *160*, 104880. [CrossRef]
35. Song, X.; Liu, X.; Liu, F.; Wang, C. Comparison of machine learning and logistic regression models in predicting acute kidney injury: A systematic review and meta-analysis. *Int. J. Med. Inform.* **2021**, *151*, 104484. [CrossRef] [PubMed]
36. Schuster, N.A.; Twisk, J.W.R.; Ter Riet, G.; Heymans, M.W.; Rijnhart, J.J.M. Noncollapsibility and its role in quantifying confounding bias in logistic regression. *BMC Med. Res. Methodol.* **2021**, *21*, 136. [CrossRef] [PubMed]
37. Sun, D.; Wen, H.; Wang, D.; Xu, J. A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. *Geomorphology* **2020**, *362*, 107201. [CrossRef]
38. Jevšenak, J.; Skudnik, M. A random forest model for basal area increment predictions from national forest inventory data. *For. Ecol. Manag.* **2021**, *479*, 118601. [CrossRef]
39. Wojtowicz, A.; Piekarczyk, J.; Czernecki, B.; Ratajkiewicz, H. A random forest model for the classification of wheat and rye leaf rust symptoms based on pure spectra at leaf scale. *J. Photochem. Photobiol. B* **2021**, *223*, 112278. [CrossRef]
40. Tanveer, M.; Rajani, T.; Rastogi, R.; Shao, Y.-H.; Ganaie, M. Comprehensive review on twin support vector machines. *Ann. Oper. Res.* **2022**, *339*, 1223–1268. [CrossRef]
41. Bansal, M.; Goyal, A.; Choudhary, A. A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning. *Decis. Anal. J.* **2022**, *3*, 100071. [CrossRef]
42. Bhavsar, H.; Panchal, M.H. A review on support vector machine for data classification. *Int. J. Adv. Res. Comput. Eng. Technol. (IJARCET)* **2012**, *1*, 185–189.
43. Wardhani, N.W.S.; Rochayani, M.Y.; Iriany, A.; Sulistyono, A.D.; Lestantyo, P. Cross-validation Metrics for Evaluating Classification Performance on Imbalanced Data. In Proceedings of the 2019 International Conference on Computer, Control, Informatics and its Applications (IC3INA), Tangerang, Indonesia, 23–24 October 2019; pp. 14–18.
44. Powers, D.M. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv* **2020**, arXiv:2010.16061.
45. Imran, A.A.; Amin, M.N.; Johora, F.T. Classification of Chronic Kidney Disease using Logistic Regression, Feedforward Neural Network and Wide & Deep Learning. In Proceedings of the 2018 International Conference on Innovation in Engineering and Technology (ICIET), Dhaka, Bangladesh, 27–28 December 2018; pp. 1–6.
46. 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]
47. Carta, F.; Zidda, C.; Putzu, M.; Loru, D.; Anedda, M.; Giusto, D. Advancements in Forest Fire Prevention: A Comprehensive Survey. *Sensors* **2023**, *23*, 6635. [CrossRef] [PubMed]
48. Giannakidou, S.; Radoglou-Grammatikis, P.; Lagkas, T.; Argyriou, V.; Goudos, S.; Markakis, E.K.; Sarigiannidis, P. Leveraging the power of internet of things and artificial intelligence in forest fire prevention, detection, and restoration: A comprehensive survey. *Internet Things* **2024**, *26*, 101171. [CrossRef]
49. Shao, Y.; Wang, Z.; Feng, Z.; Sun, L.; Yang, X.; Zheng, J.; Ma, T. Assessment of China's forest fire occurrence with deep learning, geographic information and multisource data. *J. For. Res.* **2022**, *34*, 963–976. [CrossRef]
50. Preeti, T.; Kanakaraddi, S.; Beelagi, A.; Malagi, S.; Sudi, A. Forest Fire Prediction Using Machine Learning Techniques. In Proceedings of the 2021 International Conference on Intelligent Technologies (CONIT), Hubli, India, 25–27 June 2021; pp. 1–6.
51. Amiri, M.; Pourghasemi, H.R. Predicting areas affected by forest fire based on a machine learning algorithm. In *Computers in Earth and Environmental Sciences*; Elsevier: Amsterdam, The Netherlands, 2022; pp. 351–362.
52. Hu, X.; Ma, C.; Huang, P.; Guo, X. Ecological vulnerability assessment based on AHP-PSR method and analysis of its single parameter sensitivity and spatial autocorrelation for ecological protection—A case of Weifang City, China. *Ecol. Indic.* **2021**, *125*, 107464. [CrossRef]
53. George, J.S.; Thomas, J.E.; Thomas, M.G.; Thomas, R.G. Fire Safety Alert System in Forest and Rural Areas. In Proceedings of the 2nd International Conference on IoT Based Control Networks & Intelligent Systems—ICICNIS 2021, Kerala, India, 28–29 June 2021.
54. Naing, Y.; Lai Thein, T.L. Forest Fire Detection and Warning System for Disaster Prevention. In Proceedings of the 2023 IEEE Conference on Computer Applications (ICCA), Yangon, Myanmar, 27–28 February 2023; pp. 135–139.
55. Abdessemed, F.; Bouam, S.; Arar, C. A Review on Forest Fire Detection and Monitoring Systems. In Proceedings of the 2023 International Conference on Electrical Engineering and Advanced Technology (ICEEAT), Batna, Algeria, 5–7 November 2023; pp. 1–7.