MDPI

Article

Predicting Forest Fire Area Growth Rate Using an Ensemble Algorithm

Long Zhang [ID], Changjiang Shi and Fuquan Zhang *

College of Information Science and Technology & Artificial Intelligence, Nanjing Forestry University, Nanjing 210037, China
* Correspondence: zfq@njfu.edu.cn

Abstract: Due to its unique geographical and climatic conditions, the Liangshan Prefecture region is highly prone to large fires. There is an urgent need to study the growth rate of fire-burned areas to fill the research gap in this region. To address this issue, this study uses the Grey Wolf Optimizer (GWO) algorithm to optimize the hyperparameters in the eXtreme Gradient Boosting (XGBoost) model, constructing a GWO-XGBoost model. Finally, the optimized ensemble model (GWO-XGBoost) is used to create a fire growth rate warning map for the Liangshan Prefecture in Sichuan Province, China, filling the research gap in forest fire studies in this area. This study comprehensively selects factors such as monthly climate, monthly vegetation, terrain, and socio–economic aspects and incorporates monthly reanalysis data from forest fire assessment systems in Canada, the United States, and Australia as features to construct the forest fire dataset. After collinearity tests to filter redundant features and Pearson correlation analysis to explore features related to the burned area growth rate, the Synthetic Minority Oversampling Technique (SMOTE) is used to oversample the positive class samples. The GWO algorithm is used to optimize the hyperparameters in the XGBoost model, constructing the GWO-XGBoost model, which is then compared with XGBoost, Random Forest (RF), and Logistic Regression (LR) models. Model evaluation results showed that the GWO-XGBoost model, with an AUC value of 0.8927, is the best-performing model. Using the SHapley Additive exPlanations (SHAP) value analysis method to quantify the contribution of each influencing factor indicates that the Ignition Component (IC) value from the United States National Fire Danger Rating System contributes the most, followed by the average monthly temperature and the population density. The growth rate warning map results indicate that the southern part of the study area is the key fire prevention area.

Keywords: fire growth rate; forest fire prediction; GWO-XGBoost; machine learning

1. Introduction

Forests are indispensable ecosystems on Earth, playing crucial roles in maintaining soil and water, facilitating carbon cycling, and preserving biodiversity: all essential for sustaining the global ecological balance and human environment. However, forest fires, as frequent natural disasters, cause devastating damage to forest ecosystems [1]. These fires result in massive loss of biomass, destroy ecosystem functions, and exacerbate global climate change through carbon emissions [2]. With global warming and the increase in extreme weather events, the frequency and intensity of forest fires are rising significantly, making them a critical challenge in global environmental protection and climate change research. Annually, forest fires affect hundreds of thousands of hectares globally [3]. In 2020, the massive bushfires in Australia not only caused unprecedented ecological damage but also severe economic and social impacts, with 11.46 million hectares burned and approximately three billion wildlife fatalities. On 30 March 2020, a severe forest fire in Xichang, Liangshan, Sichuan Province, China, resulted in the deaths of 19 firefighters and

three injuries and burned 3047 hectares [4]. In August 2023, large-scale wildfires in Maui, Hawaii, USA, resulted in 97 deaths and economic losses of USD 5.52 billion.

Due to the diversity of geographical environments and climatic conditions, the causes of forest fires are complex, and they lack a unified standard for prevention and control [5]. Hence, constructing localized forest fire warning systems specific for regional characteristics is crucial. Driven by this need, many countries have established comprehensive forest fire warning systems using ground stations, satellite remote sensing, and aerial reconnaissance to monitor fires in real time. Examples include the National Fire Danger Rating System [6] (NFDRS) in the USA and the Canadian Forest Fire Danger Rating System [7] (CFFDRS) in Canada. With advances in remote sensing technology and computing power, the capacity for data analysis and feature extraction in fire research has improved significantly. High-resolution satellite imagery enables researchers to accurately capture key parameters such as forest topography, vegetation, climate, and human activities [8]. These complex multidimensional datasets are well-suited for machine learning models, making advanced machine learning algorithms a hot research topic in forest fire studies.

The severity of forest fires is primarily determined by the area burned, making it essential to explore key factors and trends in burn area growth for regional fire management [9]. Fire growth rate studies are typically divided into physical and empirical models. Physical models, are based on an in-depth understanding of fire behavior and simulate burning, propagation, and heat release processes [10] and rely on precise data on weather, vegetation type, fuel moisture, and topography, thus requiring significant resources and computational power. In contrast, empirical models establish patterns through statistical analyses of historical fire data and commonly use statistical or machine learning methods like linear regression, decision trees, or random forests [11]. With the development of remote sensing technology, accessing high-quality, detailed fire-related data has become easier. Researchers have successfully used these models to predict forest fire growth rates in specific regions. For example, Juang et al. [12] utilized daily growth data of forest fires in the western USA from 2001 to 2020 to investigate the reasons behind the exponential increase in Annual Forest Area Burned (AFAB) with aridity in the western United States. The study collected a large amount of empirical data, and the results showed that the exponential growth of AFAB is primarily attributed to the exponential growth of individual forest fire areas. Furthermore, the rapid expansion of large fires is the dominant factor contributing to the exponential growth of AFAB. This study has significant implications for understanding and predicting future forest fire trends. Markuzon et al. [13] utilized multi-source heterogeneous data, including remote sensing, meteorological, and land cover information, in conjunction with machine learning algorithms such as random forest decision trees, Bayesian networks, and k-nearest neighbors to predict whether nearly 3000 wildfires in the southwestern United States would escalate into "large fires" within the next one to two days. The study results indicated that the predictive performance of the three machine learning algorithms was similar, with an accuracy rate of approximately 75%. However, there are issues with low data accuracy and unbalanced data distribution in the study, which have impact the predictive capabilities of the model. In the study of forest fire burn areas, compared with the integrated model, the single prediction model does not perform very well.

As each algorithm has limitations affecting model accuracy, researchers have found that ensemble models outperform single models, leading to their widespread use in related fields [14,15]. Bhadoria et al. [16] combined Support Vector Machine (SVM) and random forest regression models to develop a Random Vector Forest Regression (RVFR) model that was applied for predicting forest fire areas in India. They also introduced a predictive density function and vectorization, which enhanced the model's adaptability and robustness. The results demonstrated that the RVFR model achieved higher predictive accuracy (94%) and better variance (1.0) compared to other traditional single models. Mohajane et al. [17] standardized forest fire impact factors using the Frequency Ratio (FR) method and combined several machine learning models (multilayer perceptron, LR, classi-

fication and regression trees, SVM, and RF) to develop a forest fire prediction model for northern Morocco and showed the RF-FR model had the best performance (AUC = 0.989). Hao et al. [18] quantified the contribution of fire risk and related indices using RF, Gradient Boosting Decision Tree (GBDT), and XGBoost models, followed by regression analysis with Back-Propagation Neural Networks (BPNNs) and Geographically Weighted Regression (GWR), and showed that integrating RF and BPNN offered the best performance (R^2 = 0.97). Obviously, compared to single models, ensemble models effectively integrate the strengths of various models, enhancing the predictive performance and research quality.

Currently, ensemble models show good predictive capabilities in this field. Shmuel et al. [19] utilized monthly global forest fire satellite data from 2015 to extract various factors that may influence forest fires, including meteorology, topography, fuel, and population, as feature variables. They employed multiple machine learning models, including random forest, XGBoost, and multilayer perceptron, to construct classification models for predicting forest fire occurrence probability and regression models for predicting burned area size. The best model (XGBoost) achieved an AUC of 0.97 for fire occurrence prediction and a Mean Absolute Error (MAE) of 3.13 km^2 for burned area prediction. They also used 2016 global wildfire data to extract factors influencing wildfire spread, including meteorology, topography, and fuel as features. They attempted XGBoost, random forest, and MLP models to construct regression models for predicting daily burned area and classification models for predicting whether the burned area would increase. In both tasks, XGBoost continued to perform the best, significantly outperforming logistic regression [20]. It is evident that XGBoost consistently demonstrates excellent performance in tasks related to forest fires. Therefore, in this study, we choose XGBoost 1.7 as our optimization target and apply it to predict the growth rate of forest fire areas.

XGBoost has been used for the assessment of forest fire risk. Xie et al. [21] constructed a forest fire risk assessment model with strong generalization ability and robustness and applied it for fire risk prediction and mapping across the entire Liangshan Prefecture. They collected fire occurrence records from MODIS data and selected 10 factors representing the topography, meteorology, vegetation, and human activities as influencing factors. The FR method is used to assign objective weights to each factor in their method. They also employed Bayesian Optimization (BO) algorithms to automatically optimize the hyperparameters of various machine learning models, including SVM, RF, and XGBoost, and compared their models' performances. The results indicated that among the three proposed models, the FR-BO-XGBoost model performed the best, with an AUC value of 0.887. Similarly, Li et al. [22] combined four different machine learning methods—RF, XGBoost, Light Gradient Boosting Machine (LightGBM), and MLP—to propose an ensemble-learning-based forest fire risk prediction model, which they applied to fire risk prediction in Yunnan Province, China. They performed hyperparameter tuning on the base models using Bayesian optimization and employed SHAP analysis to evaluate the importance of various influencing factors. The results showed that the ensemble model achieved an accuracy of 0.906, while individual models ranged between 0.86 and 0.89. However, it is worth noting that their approaches can be considered a re-ensemble of ensemble models, which increases model complexity, thus leading to higher training time, cost, and demands on data quality. Compared to their studies focused on fire risk [19–22], our research focuses on the growth rate of forest fire in specific regions. This emphasis on growth rate offers a more dynamic perspective and contributes to a better understanding and forecasting of fire growth trends within those regions. This is critical for targeted fire management and resource allocation strategies.

The application of deep learning models in the field of wildfires is also becoming increasingly widespread. According to a detailed study by Ghali et al. [23], the use of deep learning in wildfire detection and mapping primarily includes fire detection and mapping, severity estimation, and spread prediction. Deep learning models demonstrate outstanding performance in these tasks related to wildfires. They also analyzed commonly used remote sensing data, noting that the most common type of data is image data, which are suited

for deep learning. Compared with deep learning models, the proposed GWO-XGBoost model, with its lower data and computational requirements and easier parameter tuning, is better suited for practical applications in scenarios where data acquisition is difficult or computational resources are limited.

Fire area growth rate studies typically select large regions as research subjects, as small areas rarely experience multiple fires in a short period. Only by expanding the study area can sufficient data on forest fires be collected. Research on fire growth rates is still nascent in China, but its research value cannot be overlooked. Predictions of fire growth rates on a broad scale provide management with recommendations for fire resource allocation from a macro perspective, facilitating more effective fire management and response strategies. This study focuses on predicting the growth rate of fire areas in Liangshan, Sichuan Province. During the data extraction phase, after reviewing the existing literature and conducting preliminary experiments, we have found that, compared to the simple Fire Weather Index, it is necessary to collect multi-dimensional datasets to ensure better model performance. The incorporation of additional features provides a more comprehensive understanding of the complex relationships between weather conditions, vegetation, and wildfire risk, leading to improved predictive accuracy. Incorporating additional features allows for a more comprehensive understanding of the complex relationships between terrain, weather conditions, vegetation, human activities, and fire risk, thereby improving prediction accuracy. In addition to common fire-related factors, a series of reanalysis data from national fire assessment systems are incorporated; we selected 22 potential influencing factors. After colinearity testing and correlation analysis, 13 key indicators were chosen as input features for the machine learning model. The parameters of the XGBoost model were optimized using the GWO, and the GWO-XGBoost model was compared with RF, XGBoost, and LR models. The performance assessment showed that the GWO-XGBoost model excelled in several performance metrics, demonstrating good classification fit. To further interpret the model's predictions, SHAP value analysis was used to quantify the contributions of various factors to the model output. The analysis revealed that the Ignition Component (IC), Mean Monthly Temperature (MT), and Population Density (PD) were the most significant indicators affecting the trend of fire area growth. Finally, by fitting the data from March of historical fire years in Liangshan, a fire area growth rate warning chart was drawn, and together with a risk difference map of the Luzhou region, the study's rationality was validated. By thoroughly analyzing these key indicators, this research provides data-driven insights for local forest fire management, promoting data-based decision-making processes and providing a scientific basis for fire prevention and management efforts in the Liangshan area.

2. Materials and Methods

In this section, we present the study area and data sources and analyze the factors influencing the growth rate of forest fire areas, including terrain factors, vegetation factors, meteorological factors, human activity factors, and fire risk assessment factors. We then provide a detailed introduction to the proposed GWO-XGBoost model, including data preprocessing methods and the construction of the GWO-XGBoost model.

2.1. Study Area

Liangshan Yi Autonomous Prefecture, located in the southwestern part of Sichuan Province, is an area with beautiful natural scenery and rich ecological resources. The prefecture covers an area of 60,423 square kilometers and had a population of 5.3103 million at the end of 2019. Geographically, Liangshan is situated on the northeastern edge of the Hengduan Mountains in southwestern Sichuan, between the Sichuan Basin and the central Yunnan Plateau [24]. The terrain is high in the northwest and low in the southeast, resulting in varied topography and landforms. The climate is classified as subtropical monsoon, with distinct wet and dry seasons, creating a unique ecological environment. The vegetation mainly consists of Yunnan pine and alpine pine, with a forest coverage rate

of 43%, making it an important forestry and pastoral area in Sichuan Province. However, these geographical features and climatic conditions also make the region prone to forest fires, especially during the dry season from November to April each year, when the risk of forest fires is high [21,25]. The geographical location of the research area is shown in Figure 1.
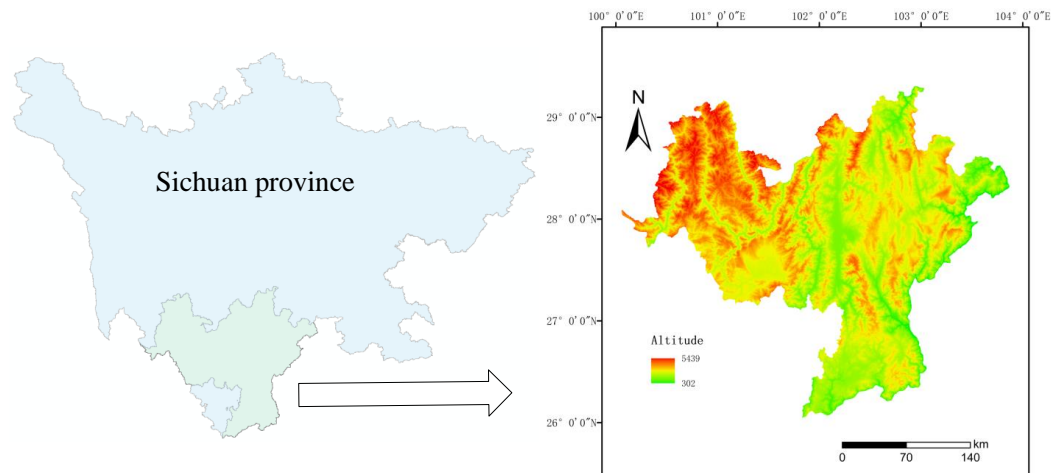
Figure 1. Study area.

2.2. Data Sources

The forest-fire-related data used in this study can be categorized into independent-variable data and dependent-variable data. The independent-variable data include common fire influencing factors: elevation, slope, aspect, Normalized Difference Vegetation Index ($NDVI$), temperature, relative humidity, precipitation, wind speed, roads, and population. To further enhance the quality of dataset features and better explain the key factors of forest fire growth rate, fire risk assessment indicators were extracted from the CFF-DRS [26], the NFDRS [27] of the United States, and the Australian Fire Danger Rating System (AFDRS) [28].

From the CFFDRS, we extracted the Build-Up Index (BUI), danger-risk, Drought Code (DC), Duff Moisture Code (DMC), Daily Severity Rating (DSR), Fine Fuel Moisture Code (FFMC), Fire Weather Index (FWI), and Initial Spread Index (ISI). From the AFDRS, we collected the Fire Danger Index (FDI) and Keetch–Byram Drought Index (KBDI). From the NFDRS, we extracted the Burning Index (BI), Energy Release Component (ERC), Ignition Component (IC), and Spread Component (SC).

The study area was divided into 0.2° × 0.2° (latitude and longitude) grids using ArcMap 10.2 software [20], resulting in a total of 140 grids. We then collected forest fire data for each grid from 2006 to 2019, a span of 14 years, and summarized them by month. The variations in the burned area for the same months across different years within the corresponding grid were calculated. If the burned area in the following year was larger than that of the current year, it was considered an increase in the burned area, and the change value was labeled as 1; otherwise, the change value was 0, indicating no increase in the area. Generally, the larger the burned area, the higher the reliability of the fire points detected by remote sensing images. Therefore, to obtain more reliable data on forest fire burned areas, only fire points with a burned area greater than 20 hectares were selected for recording [29]. There were only 365 samples with a change value of 1. To obtain a sufficient dataset [30], 730 samples with a change value of 0 were randomly selected, resulting in a total of 1095 data points for the dataset on changes in burned areas. The specific factors are shown in Table 1.

Table 1. Multicollinearity analysis for relevant factors.

Data Type	Category	Data Range	
		Minimum Value	Maximum Value
Terrain Factors	Altitude	302	5439
Vegetation Factors	*NDVI*	0	0.9
Meteorological Factors	MT (K)	247.21	294.5
	MH (Monthly Humidity, kg·kg^{-1})	0	0.017
	MP (Monthly Precipitation, kg·m^{-2}·s^{-1})	0	4.43×10^{-4}
	MSR (Monthly Solar Radiation, W·m^{-2})	161.99	1362.15
	MW (Monthly Wind Speed, m·s^{-1})	0.25	2.61
Human Activity Factors	PD (people/km^2)	0	609.38
Fire Risk Assessment Factors	BUI	0	178.7
	Danger-risk	1	5.31
	DC	0	547.73
	DMC	0	163.56
	DSR	0	26.38
	FFMC	15.57	94.39
	FWI	0	48.11
	ISI	0	17.06
	FDI	0	13
	KBDI	0	58
	BI	0	53
	ERC	0	56
	IC	0	38.83
	SC	0	9.52

2.2.1. Fire Point Extraction

The fire point data involved in this study were obtained from NASA's Fire Information for Resource Management System (FIRMS). FIRMS is an important part of NASA's Earth Observing System Data and Information System (EOSDIS) and focuses on providing real-time information on active fire points and thermal anomalies globally (https://www.earthdata.nasa.gov/learn/find-data/near-real-time/firms/active-fire-data, accessed on 5 March 2024). This system utilizes data collected by the Moderate Resolution Imaging Spectroradiometer (MODIS) and the Visible Infrared Imaging Radiometer Suite (VIIRS) onboard NASA satellites. VIIRS, with its 375 m spatial resolution [31], offers finer and more sensitive monitoring capabilities for small-scale fire incidents. FIRMS integrates data from both MODIS and VIIRS, providing real-time, high-value information resources for global fire management and related research; it particularly excels at the identification and analysis of smaller fire sources.

In this study, fire incidents in the Liangshan region from 2006 to 2022 were extracted. Data from 2006 to 2022 were extracted from the MODIS sensor, and data from 2012 to 2022 were extracted from the VIIRS sensor. To avoid duplicate fire points, ArcMap 10.2 software was used to remove redundant fire points.

The satellite sensor data include "scan" and "track" indicators, which represent the pixel sizes of the scan and track, respectively. Based on the resolutions of 1000 m for the MODIS sensor and 375 m for the VIIRS sensor, combined with specific pixel sizes, the burned area at the time of fire point observation was calculated [32]. Most fire prediction studies focus on whether a fire incident occurs at a fire point, whereas this study focuses on the burned area after a fire incident. To improve the accuracy of the area measurement, only fire incidents with a burned area greater than 20 hectares were selected.

2.2.2. Vegetation Factors

Surface vegetation is the main source of fuel for fires and plays a crucial role in forest fire research. Different vegetation types and coverage directly affect the characteristics

and quantity of surface fuels, significantly influencing the occurrence and spread of fires. The $NDVI$ is a remote sensing measurement index that is widely used in forest fire research to measure vegetation density and health.

The $NDVI$ data used in this study were obtained from the "Monthly 1KM Resolution Vegetation Index Spatial Distribution Dataset of China", provided by the Resource and Environment Science and Data Center of the Chinese Academy of Sciences (https://www.resdc.cn/, accessed on 9 March 2024). This dataset is generated from 1 KM resolution vegetation index data on SPOT/VEGETATION satellite images, which are processed using the maximum value composite method to create monthly vegetation index data, with negative values treated as zero. $NDVI$ is calculated as the ratio of the difference in the sums of the reflectances in the near-infrared (NIR) and red (RED) bands [33], as shown in Equation (1).

$$NDVI = (NIR - RED)/(NIR + RED) \tag{1}$$

$NDVI$ values range from -1 to 1. Values closer to 1 indicate denser vegetation, values closer to -1 generally indicate water bodies, and values near 0 may represent bare soil or sparsely vegetated surfaces. This index effectively reveals vegetation coverage, providing crucial information for understanding potential wildfire risks and spread paths.

2.2.3. Topographic Factors

Topographic factors are important indicators affecting fires [34]. Due to the limitations of large-scale research, this study selects the average altitude of the study area as a topographic factor. Altitude directly influences climatic conditions, particularly temperature and humidity, which indirectly affect the dryness and combustibility of vegetation. Generally, the higher the altitude, the lower the temperature, the higher the humidity, the lower the vegetation's combustibility, and the lower the probability and burning area of fires.

This study extracts the average altitude from the Digital Elevation Model (DEM). DEM data come from the ASTER GDEM 30M product of the Geospatial Data Cloud Platform, which is generated based on ASTER satellite data and has a high resolution of 30 m (https://www.gscloud.cn/search, accessed on 6 March 2024).

2.2.4. Meteorological Factors

Meteorological factors are key in triggering wildfires [35]. Long-term climate patterns influence local fire behavior and patterns. This study selects the monthly average temperature, monthly average humidity, monthly average precipitation, and monthly average solar radiation as meteorological factors in the study area. An increase in the monthly average temperature typically causes vegetation and other combustibles to lose moisture, making them drier and more easily ignitable. Higher temperatures make it easier for fires to occur and spread, potentially leading to larger forest fires. A decrease in humidity means less moisture in the air, making vegetation and other combustibles more flammable. Under low-humidity conditions, even small ignition sources can quickly develop into large-scale fires; thus, the monthly average humidity and changes in forest fire burning areas are interrelated. An increase in precipitation increases the moisture content of the ground and vegetation, thereby reducing the probability of fire occurrence and spread. Higher monthly average precipitation usually results in smaller fire burn areas. Increased solar radiation accelerates the drying process of the ground and vegetation, increasing the flammability of combustibles. High solar radiation not only raises the vegetation temperature but may also further reduce vegetation humidity by accelerating moisture evaporation, thereby increasing fire risk and the potential burn area.

The monthly average meteorological data come from the ERA5 near-surface meteorological reanalysis dataset provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) (https://cds.climate.copernicus.eu/, accessed on 15 March 2024). These data have been altitude and monthly scale bias-corrected by the Global Precipitation

Climatology Centre (GPCC) and the National Centre for Atmospheric Science (NCAS), ensuring high reliability.

2.2.5. Fire Risk Assessment Factors

Fire risk assessment is a complex process involving multiple key indicators essential for comprehensively understanding the probability of fire occurrence and its potential spread in a given area. Currently, countries such as the United States, Canada, and Australia have developed advanced fire warning systems. These include the Canadian Forest Fire Danger Rating System (CFFDRS), the United States National Fire Danger Rating System (NFDRS), and the Australian Fire Danger Rating System (AFDRS). These systems provide critical scientific bases for fire prevention and management by integrating various meteorological factors and environmental conditions [36]. The European Forest Fire Information System (EFFIS) integrates the fire danger indices from the above-mentioned three models and utilizes ERA5 reanalysis data from the European Centre for Medium-Range Weather Forecasts (ECMWF) to calculate fire risk indices. The fire assessment data used in this study mainly originate from the ERA5 global atmospheric reanalysis dataset produced by ECMWF (https://cds.climate.copernicus.eu/, accessed on 11 March 2024).

The Build-Up Index (BUI) reflects the accumulation of combustible materials in an area and is closely related to the duration and intensity of a fire. The danger-risk index directly assesses the likelihood of fire occurrence. The Drought Code (DC) and Duff Moisture Code (DMC) indicators evaluate soil moisture conditions, where dry soil is more likely to contribute to fire spread. The Daily Severity Rating (DSR) relates to the potential damage caused by a fire. The Fine Fuel Moisture Code (FFMC) focuses on the flammability of surface fine fuels, which is crucial for initial fire ignition and spread. The Fire Weather Index (FWI) is a composite indicator that assesses fire risk under current meteorological conditions. The Initial Spread Index (ISI) evaluates the potential speed of fire spread under given conditions. The Fire Danger Index (FDI) combines multiple factors to provide a comprehensive assessment of fire risk. The Keetch–Byram Drought Index (KBDI) addresses the impact of long-term drought on soil dryness. The Burning Index (BI) and Energy Release Component (ERC) evaluate the potential severity and energy release of a fire. The Ignition Component (IC) is related to the ease of ignition, while the Spread Component (SC) predicts the potential speed and extent of fire spread.

2.2.6. Human Activity Factors

Human activities are critical factors influencing fire occurrence [20]. Around forest areas, the frequency of human activities directly impacts the rate of human-caused fires, such as due to dropped cigarette butts and outdoor cooking, which can easily ignite fires. Additionally, the presence of roads not only facilitates the extension of human activities into forest areas but also becomes a significant cause of forest fires. People tend to engage in various activities near roads, making these areas high-risk zones for human-caused fires. People in vehicles on roads might accidentally drop cigarette butts, or accidents might cause open flames, further increasing the risk of forest fires.

This study selects population density as an indicator to measure the impact of human activities on the changing trend of fire burn areas. The higher the population density in a region, the more frequent human activities are likely to be. Frequent human activities may lead to an increase in human-caused fires in the area, resulting in larger fire burn areas. These data come from the Gridded Population of the World, Version 4 (GPWv4) dataset produced by the Center for International Earth Science Information Network (CIESIN) at Columbia University's Earth Institute (https://sedac.ciesin.columbia.edu/, accessed on 16 March 2024). This dataset provides global population density estimates for the years 2000, 2005, 2010, 2015, and 2020, with a spatial resolution of 1km×1km in raster format. By matching the fire occurrence times with population density data in adjacent periods, this study obtained population data closely related to specific fire events.

2.3. Research Method

This study establishes a binary classification model to explore the relationship between the fire growth rate and fire influencing factors in the study area. The choice to use a classification method rather than regression models for handling wildfire data was primarily based on the following considerations: Firstly, regression models have stringent requirements regarding the distribution, characteristics, and quantity of the dataset. In wildfire data, the recorded burn area often contains certain inaccuracies, and due to the unpredictability of wildfires, occasional large-scale fires may result in burn areas that far exceed the total burn area of multiple smaller-scale fires. These highly skewed data make it difficult for regression models to provide effective predictions. Additionally, there is a problem of inconsistent dimensions in the wildfire dataset, necessitating normalization when using regression analysis. Although common methods like applying logarithmic transformations for exponential normalization can address dimensional issues, such transformations may alter the intrinsic regularities in the numerical distribution of the original data. For instance, logarithmic transformations are less sensitive to high-value data compared to low-value data, which may prevent the model from effectively capturing the characteristics of large-scale wildfires. Therefore, considering these data characteristics and the potential issues arising from data processing, we opted for a classification method rather than attempting to precisely predict the exact burn area. This approach better accommodates the data's characteristics and analytical needs while avoiding potential information loss due to data preprocessing.

To facilitate the extraction of fire area data in the study area, Liangshan Prefecture in Sichuan Province was selected as the research area. To improve the accuracy of fire point extraction, fire points with an area greater than 20 hectares were screened. In total, 828 fire points were extracted from the region between January 2006 and December 2019, each with time and burn area information. The study area was divided into 140 grids with a resolution of $0.2° \times 0.2°$, and the total burn area of all fire points in each grid was counted monthly. Finally, the study observed whether the total burn area of fire points in the same grid increased in the following year compared to the current year for the same month. If the area increased compared to the current year, the variable "change" was assigned a value of 1; otherwise, it was 0. In total, 365 grid points with a "change" value of 1 were obtained, and based on temporal and spatial randomness, 1000 grid points with a "change" value of 0 were selected. These two types of data were then oversampled using the SMOTE algorithm and were divided into training and test sets. The Grey Wolf Optimizer (GWO) algorithm was used to optimize the parameters of the XGBoost model, establishing the GWO-XGBoost model, which was then compared with RF, XGBoost, and LR models. Finally, the model results were visualized, and a fire growth rate map of the study area was plotted.

2.3.1. Data Preprocessing

In this study, we use the influencing factors of forest fire burned areas as features and the trend of burned area change as labels to construct a prediction model for the growth rate of forest fire burned areas. The data processing and analysis are primarily conducted using ArcMap, MATLAB R2023a, and PyCharm following these general steps:

(1) Determination of influencing factors for forest fire burned areas: Given the large grid area and time scale, we select 22 factors influencing forest fire burned area, including monthly average temperature, monthly average humidity, monthly average precipitation, monthly average wind speed, monthly average solar radiation, altitude, normalized vegetation index, population density, buildup index, fire danger index, soil moisture, surface soil moisture, fire severity rating, fine fuel moisture code, fire weather index, initial spread index, fire danger index, Keetch–Byram drought index, burning index, energy release component, ignition component, and spread component. We handle missing values and outliers in the extracted data.

(2) Collinearity check for influencing factors: Given the high dimensionality of the extracted data, especially fire assessment-related indicators, there may be high collinearity.

We perform correlation analysis on the 22 features, as shown in Figure 2, where the horizontal axis represents the log-transformed values of the Variance Inflation Factor (VIF) for each feature. Figure 2 shows significant collinearity among fire-assessment-related indicators, as they are all evaluation values of fire-related indices. Collinearity is unavoidable in this context. Based on VIF values, we sequentially delete features from high to low VIF values, prioritizing the deletion of fire-assessment-related indicators when VIF values are similar. Finally, we obtain features that satisfy collinearity checks, as shown in Figure 3, where all VIF values are below 10, meeting the requirements of the forest fire dataset.
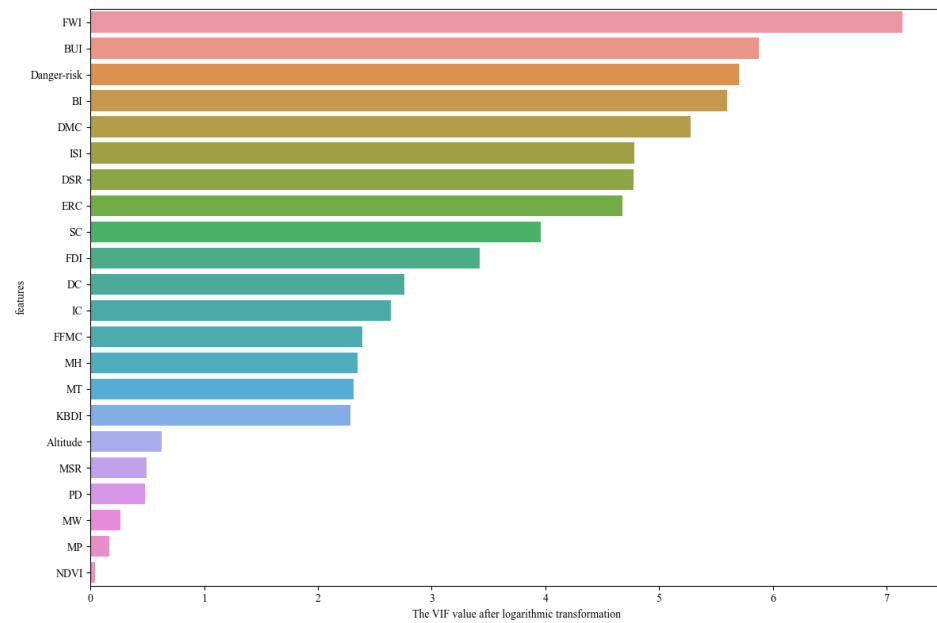
Figure 2. Log-transformed VIF values of original features.
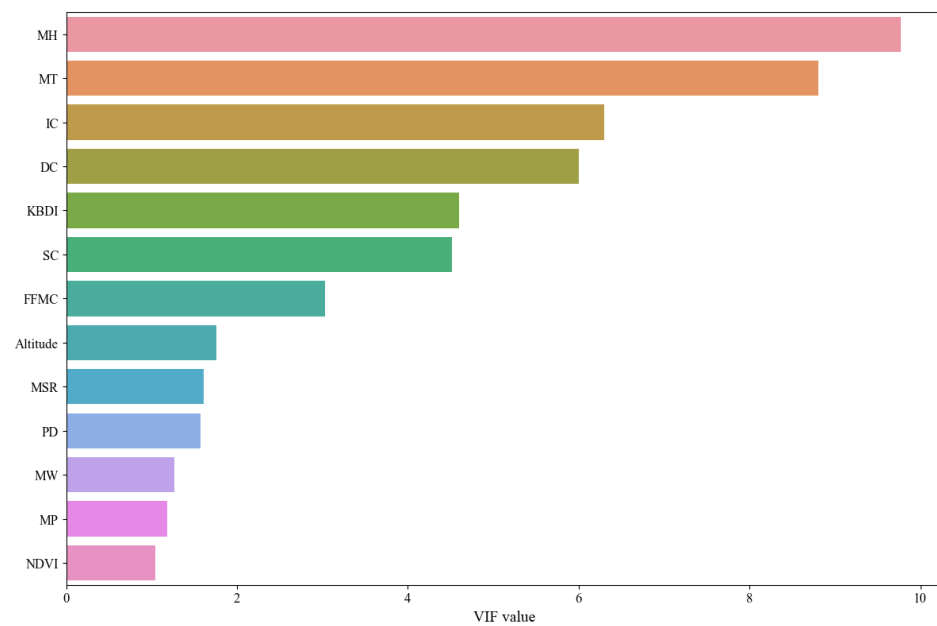
Figure 3. VIF Values of features after selection.

(3) Pearson correlation analysis for influencing factors: We conduct a correlation analysis on the factors that pass the collinearity check [37]. The results are shown in Figure 4, where clear correlations exist among fire-assessment-related indicators, as they

are all used to assess fire risk after reanalysis. Correspondingly, their correlations with the label indicating area growth change are also higher than for other factors. For instance, in Figure 4, the correlation between monthly average humidity and monthly average temperature is 0.9, indicating that humidity varies with temperature, possibly due to the unique climate of the study area. Liangshan Prefecture primarily has a subtropical humid climate, where high temperatures and humidity may coexist during the summer.
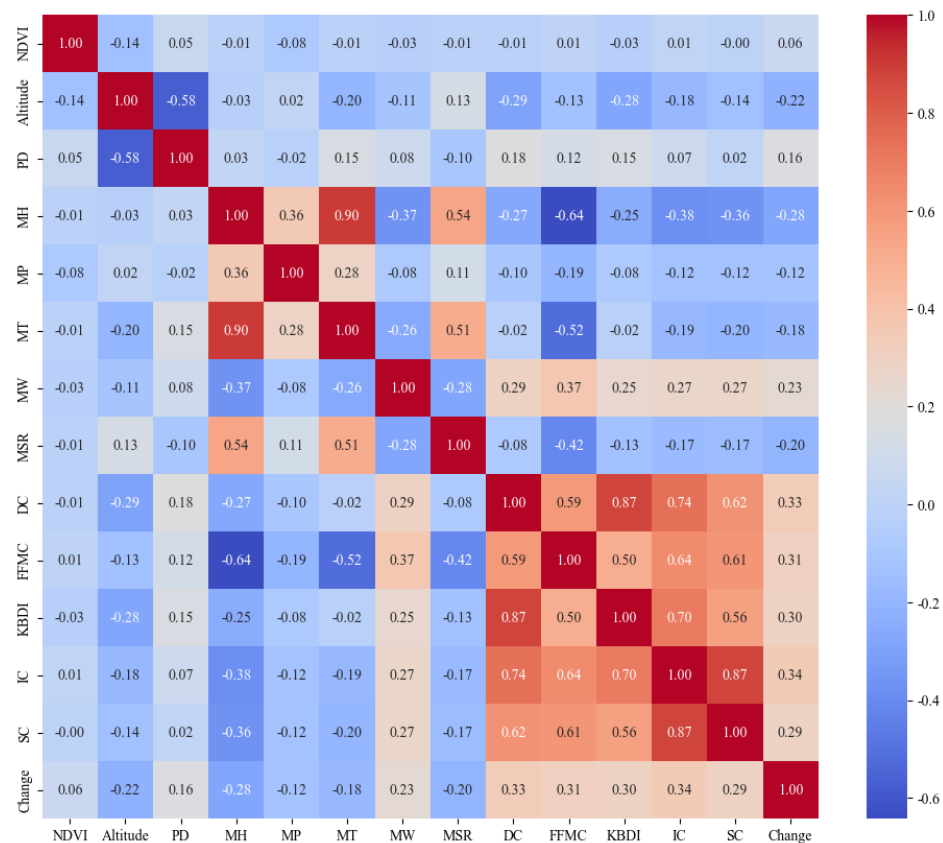
Figure 4. Heatmap of correlation analysis.

(4) Oversampling using the SMOTE: Given a 1:2 ratio of positive to negative samples, we use the SMOTE from the `imblearn` library to oversample positive class samples to achieve better model fitting [38]. By setting `sampling_strategy=auto`, the number of minority class samples increases to match the majority class, resulting in a 1:1 ratio of positive to negative samples.

(5) Model construction and data preparation: Using the `train_test_split` method from the `sklearn` library, we randomly split 30% of the dataset to be the test set and 70% as the training set, setting `random_state=42` to ensure all training models use the same dataset. We extract relevant data from 140 grids in the study area that meet the training set requirements. The data for March of historically fire-prone years are selected for final fitting and prediction and are used to draw the fire burned area growth rate warning map for the study area.

2.3.2. GWO-XGBoost Model

The XGBoost model [39] includes numerous hyperparameters that significantly impact its performance. Given the abundance of hyperparameters and the model's inherent robustness, optimizing these parameters is typically done through empirical judgment, grid search, or random search methods. In this study, the Grey Wolf Optimization (GWO) algorithm [40] is employed to optimize the hyperparameters of the XGBoost model.

Initially, a random population of grey wolves is generated, with each wolf representing a set of XGBoost hyperparameters. The accuracy obtained from three-fold cross-validation is used as the fitness function to evaluate the performance of each parameter set. During the iteration process of the algorithm, the behaviors of the wolf pack (including encircling, tracking, and attacking prey) are simulated to update the position of each wolf, with the prey representing the optimal hyperparameter combination. Through this method, the wolf pack searches within the solution space and continuously adjusts its position to approach the optimal solution. The algorithm stops once the convergence condition is met. The final result is the XGBoost model optimized by the GWO algorithm, which is then evaluated on the validation set for performance. This study primarily uses PyCharm 2023.2.1 and the sklearn library in Python to build the models. Figure 5 is a flowchart of the improved algorithm.
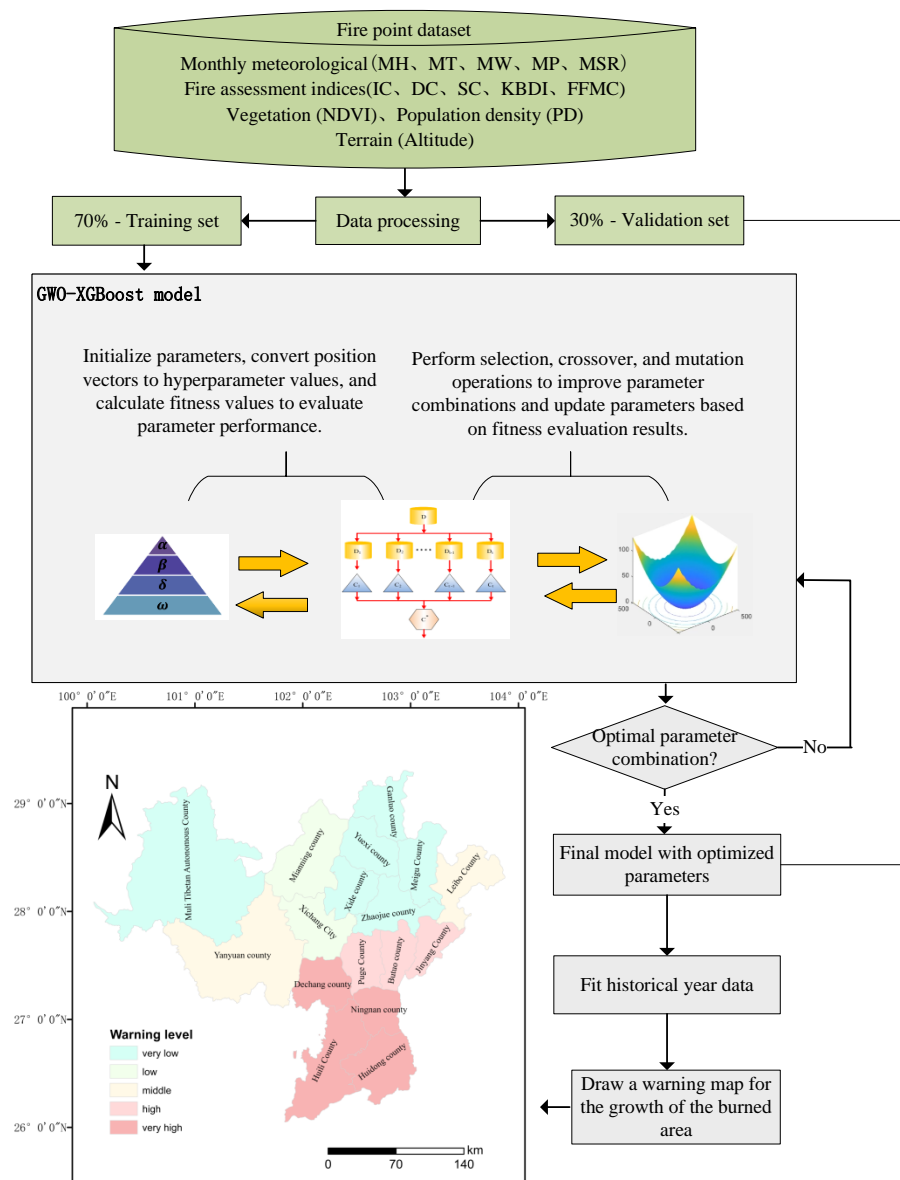

Figure 5. Flowchart of the GWO-XGBoost algorithm.

The specific steps are as follows:

(1) Define parameter boundaries: Establish feasible search ranges for each hyperparameter by setting minimum and maximum values, which constrain the search space of the algorithm. These ranges should be broad enough to include possible optimal parameter values but not too wide in order to avoid an excessively large search space.

(2) Initialize parameters: Initialize a set of candidate hyperparameter combinations based on the defined parameter boundaries.

(3) Decode parameters: The GWO algorithm optimizes the problem by simulating the social behavior of grey wolves. The decode function converts the solutions (position vectors) in the GWO algorithm to actual hyperparameter values of the XGBoost model, mapping points in the search space to specific values in the hyperparameter space.

(4) Calculate fitness: For each candidate parameter combination, evaluate its performance using the fitness_function. This typically involves training the XGBoost model with the candidate parameters and assessing its performance on the validation set. Performance metrics (such as accuracy) are used as fitness values to guide the subsequent search process.

(5) Selection, crossover, and mutation: Mimicking the natural selection mechanism, the algorithm selects better-performing solutions as parents for the next generation. Crossover and mutation operations introduce diversity into the solutions to explore new parameter combinations. The crossover operation allows solutions to share information, while mutation introduces new features.

(6) Iterative optimization and selection of the best solution: The algorithm repeatedly executes the above steps until the stopping condition is met, ultimately selecting the parameter combination with the highest fitness value as the optimal solution.

3. Results and Discussion

In this section, we present the performance of the proposed GWO-XGBoost model and conduct an in-depth analysis of the trends in forest fire burned area changes in the Liangshan Prefecture. Our model reveals several key factors that significantly influence the variations in burned area. Specifically, the Ignition Component (IC), Mean Monthly Temperature (MT), and Population Density (PD) exhibit a strong correlation with the increase in forest fire burned area. Furthermore, we have gained new insights into fire risk and prevention strategies in the study area. The results indicate that the southern region of Liangshan Prefecture is a critical area for fire prevention and control, with a high probability of an increasing rate of burned area. Additionally, areas with high ignition component values are also key areas for fire prevention, as a high IC is a crucial factor influencing the growth of forest fires in the Liangshan region.

3.1. Model Performance Analysis

In previous studies using XGBoost prediction models, the intrinsic generalization ability of XGBoost often led to the neglect of parameter selection, with parameters mostly adjusted based on simple empirical values. However, parameter settings still have a significant impact on improving the performance of the XGBoost model. In this study, the default parameters of XGBoost were chosen as the baseline model, and the conventional tuning range was used as the optimization boundary [41]. Seven key parameters were selected, with their default values, tuning ranges, and optimized values shown in Table 2.

**Table 2.** Comparison of initial and optimized parameters for the XGBoost model.

Parameter	Description	Default Value	Optimization Range	Optimized Value
max_depth	Maximum tree depth	3	(3, 10)	7
learning_rate	Learning rate	0.1	(0.01, 0.3)	0.058
n_estimators	Number of trees	100	(100, 500)	115
gamma	Minimum loss reduction	0	(0, 0.7)	0.231
min_child_weight	Minimum sum of instance weight (Hessian)	1	(1, 7)	1.632
subsample	Subsample ratio of the training instances	1	(0.6, 1.0)	0.765
colsample_bytree	Subsample ratio of columns when constructing each tree	1	(0.6, 1.0)	0.967

Additionally, to gain a deeper understanding of the effect of model optimization, the model was run 50 times, and we recorded the accuracy and AUC values each time.

The kernel density plots comparing the results before and after optimization are shown in Figures 6 and 7, which visually demonstrate the differences in the two indicators. The results indicate that the GWO-optimized XGBoost model (GWO-XGBoost) shows a significant improvement in accuracy, as accuracy was used as the fitness function for the GWO algorithm. The optimized model also exhibits superior performance in AUC values, indicating enhanced generalization capability.
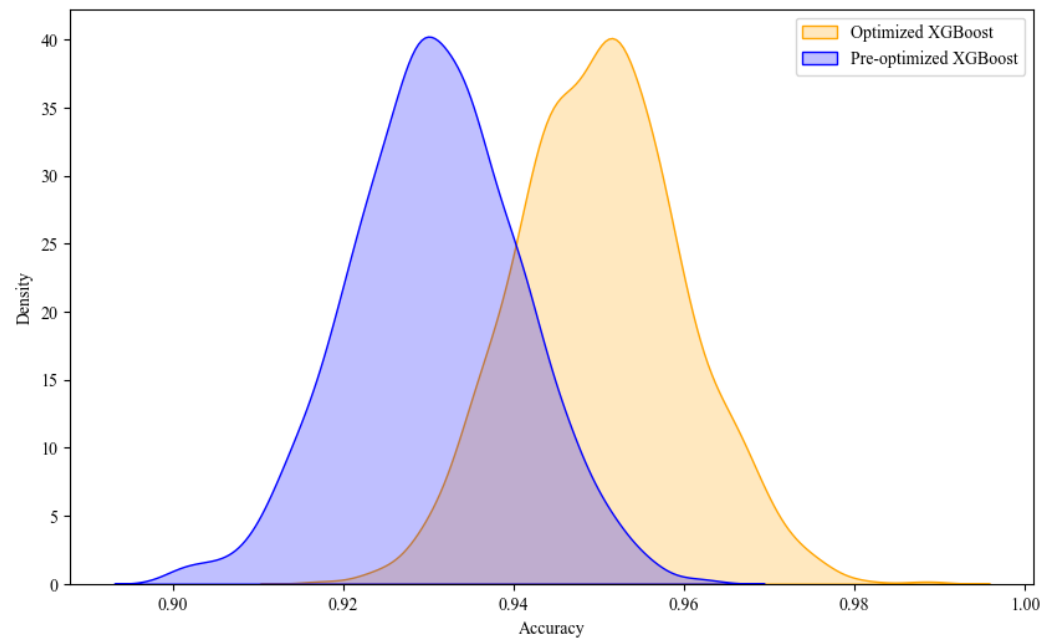
Figure 6. Kernel density plot of accuracy before and after optimization of the XGBoost model.
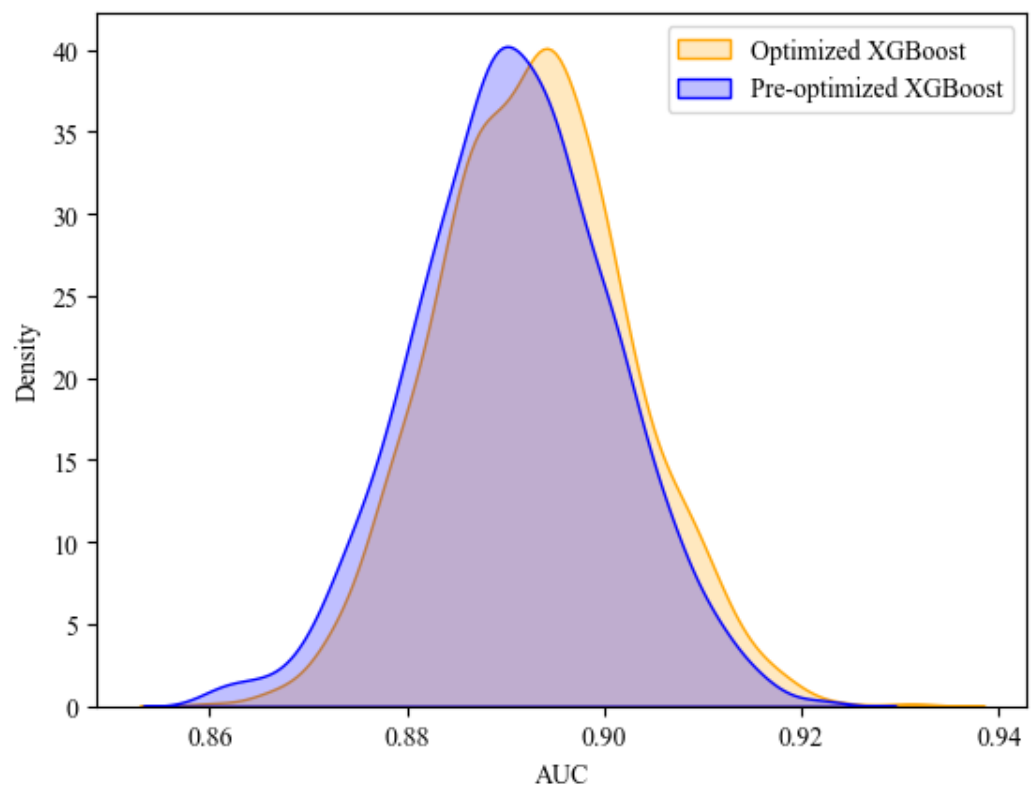
Figure 7. Kernel density plot of AUC before and after optimization of the XGBoost model.

Table 3 summarizes the mean results of these 50 runs, which are presented in numerical form. It is obvious that the GWO-optimized XGBoost model exhibits better overall performance. This study illustrates that even with the strong generalization ability of the XGBoost model, appropriate parameter optimization can further enhance model performance, which has a direct and positive impact on its practical application. By comprehensively evaluating the model's performance before and after parameter optimization, this study provides a better understanding of the importance of parameter adjustment in the model optimization process, offering a reference and guidance for future research and applications.

Table 3. Average accuracy and average AUC values before and after optimization of the XG-Boost model.

Model	Average Accuracy	Average AUC Value
XGBoost	0.7986	0.8901
GWO-XGBoost	0.8154	0.8929

The study also compared the GWO-XGBoost model with RF and LR models, which are widely used in forest fire prediction research. Performance evaluation metrics included precision, accuracy, recall, F-value, and AUC, and we focused primarily on the models' performance on the test set. Using the "train_test_split" function from sklearn, the dataset was split into training and test sets in a 7:3 ratio, with "random_state = 42" specified to ensure reproducibility, ensuring all models were trained and evaluated under the same data conditions. The evaluation metrics derived from the confusion matrix are summarized in Table 4, and Figure 8 shows the comparison of ROC curves.
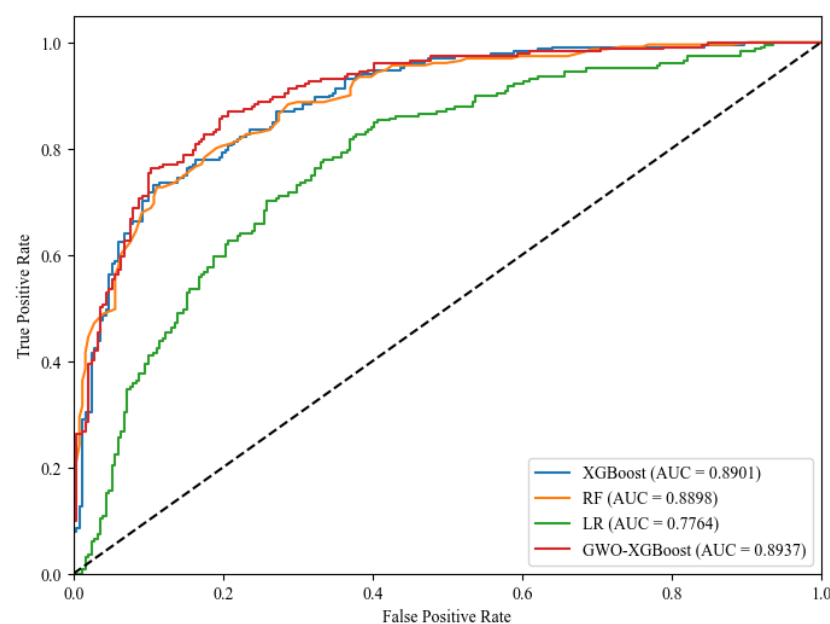
Figure 8. ROC curves of four models.

As seen in Table 4, the GWO-XGBoost model outperforms other models across all metrics. While the XGBoost model generally outperforms RF, the RF model scores higher in the recall metric, indicating it can better identify actual forest fire events, though it may also predict more non-fire events as fires. The LR model performs the worst; despite its lower computational cost and ease of understanding and implementation, it underperforms when dealing with high-dimensional and complex data. This indicates that more complex and refined models are needed to capture intricate relationships in forest fire prediction data.

Table 4. Comparative evaluation metrics of four models.

Model	Precision	Recall	F-Value	Accuracy
LR	0.6975	0.7186	0.7079	0.7158
RF	0.7406	0.8528	0.7928	0.7863
XGBoost	0.7659	0.8355	0.7992	0.7988
GWO-XGBoost	0.7752	0.8658	0.8180	0.8154

In the ROC curve comparison, the GWO-XGBoost model has the highest AUC value, indicating its superior classification performance and ability to effectively distinguish between positive and negative samples. The GWO algorithm plays a crucial role in enhancing the performance of the XGBoost model. Through parameter adjustment, the model can more accurately predict the trends in forest fire area changes, which is vital for early warning and prevention efforts.

3.2. Feature Importance Analysis Based on SHAP Method

In 2017, Lundberg et al. [42] introduced the SHAP (SHapley Additive exPlanations) method, which is based on the concept of Shapley values from game theory, to explain the behavior of machine learning model predictions. The SHAP method evaluates feature importance by accurately calculating the Shapley value of each feature's contribution to the model prediction. The SHAP value calculation formula is as follows:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \tag{2}$$

In Equation (2), ϕ_i represents the contribution of feature i to the model output, S is a subset of features, $|S|$ is the number of features in S, N is the total number of features, $f(S \cup \{i\})$ is the prediction with feature i included, and $f(S)$ is the prediction without feature i.

Additionally, SHAP defines a linear function g based on binary features, where the core idea is to decompose the model's prediction into the sum of each feature's independent contributions, represented by their SHAP values. This function is based on the following additive feature attribution method:

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z'_i \tag{3}$$

In Equation (3), ϕ_0 is the model output baseline (the prediction when no feature values are present), ϕ_i is the SHAP value of feature i, z'_i is a binary value indicating whether feature i is present (1 if present; 0 if not), and M is the total number of features in the model. The core value of this method lies in its fairness, as it mathematically distributes each feature's influence on the model's prediction fairly, ensuring precise and highly interpretable evaluations. This is particularly useful in binary classification problems, as it not only quantifies each feature's contribution to positive (e.g., increased burn area growth rate) or negative predictions but also reveals which features play a more significant role in predicting the two possible outcomes.

In this study, SHAP analysis was performed on the GWO-optimized XGBoost model. Using the TreeExplainer function from the shap library, the contribution of each feature to the prediction of forest fire area change trends was quantitatively analyzed. Figure 9 is a swarm plot, where the x-axis SHAP values represent the feature's contribution to the model prediction, and the color indicates the relative value of the feature. Additionally, Figure 10 illustrates the feature importance ranking based on the average absolute SHAP values, providing an intuitive comparison of the different features' impacts on the model predictions.
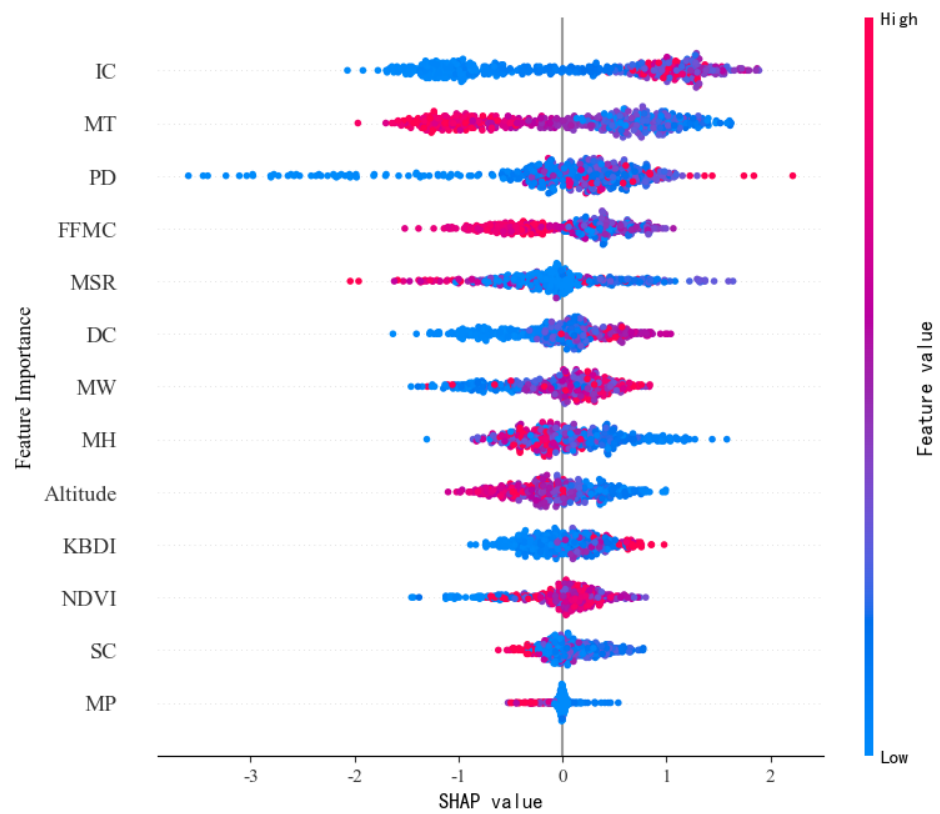
Figure 9. Feature beeswarm plot based on SHAP analysis.
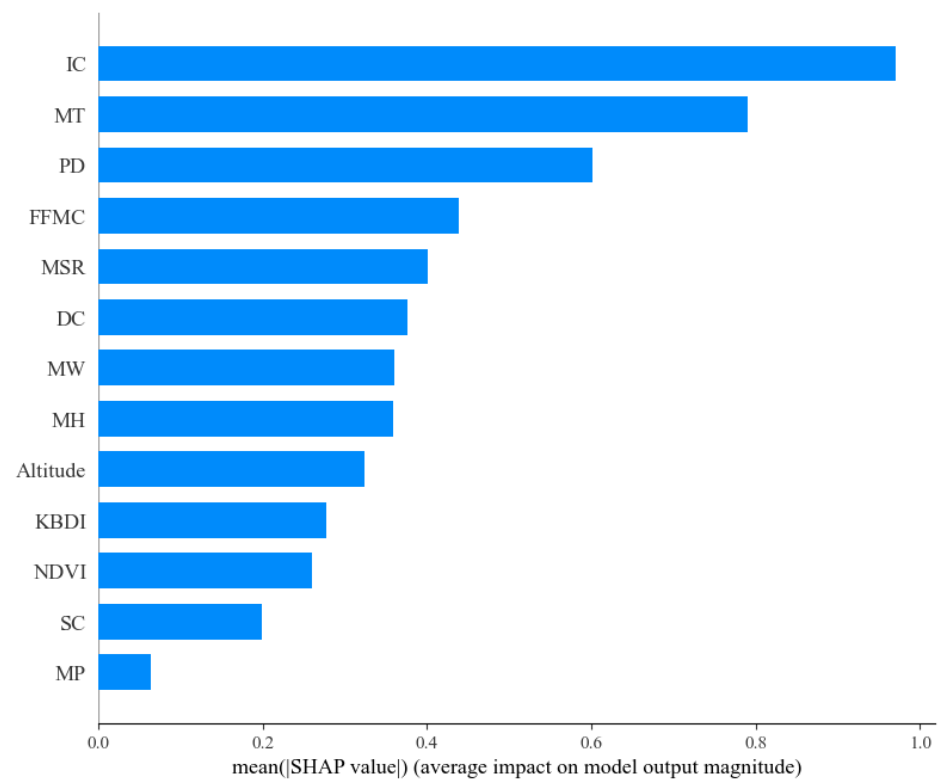

Figure 10. Feature importance chart based on SHAP analysis.

Figure 10 shows that the feature with the most significant impact on the trend of burn area change is IC, followed by MT and PD, which are the ignition component values

from the NFDRS, monthly average temperature, and population density, respectively. The ignition component value represents the probability that an ember will start a fire that requires human intervention to suppress and ranges from 0 to 100. When the ignition component value is 100, it indicates that every ember will start a fire that needs intervention, while a value of 0 indicates that embers will not start fires that need suppression. Thus, the IC value reflects the potential threat of fire and contributes most to the changes in burn areas in Liangshan Prefecture. Local fire management departments can incorporate this indicator to improve future fire prevention plans.

Figure 9 shows that the IC indicator has both positive and negative impacts on the burn area change trend. The red dots representing high IC values are concentrated on the right side of the positive value range, indicating that when the ignition component value is high, its impact on the burn area change trend is positive; that is, the larger the IC value, the more likely the burn area will increase. The blue low-value points are distributed on both sides, indicating that low IC values have a more complex impact on the label, while the red dots are more concentrated, showing a significant positive correlation between high IC values and the label. Local fire prevention departments should focus on areas with high IC values because the larger the IC value, the more likely the burn area will increase.

The MT indicator represents the monthly average temperature of the area. Figure 9 shows that most low-temperature values have a positive impact on the trend of burn area changes, indicating that in areas with lower monthly temperatures, the burn area tends to increase with rising temperatures. PD represents the population density of the area, with its distribution skewing towards the positive value range on the right side, indicating that population density has a positive impact on the trend of burn area changes; that is, the higher the population density, the more likely the burn area will increase.

In the study by Xie et al., precipitation is identified as the most important triggering factor for wildfires in Liangshan [21], as their research focuses on using integrated machine learning models for wildfire risk assessment. In contrast, factors such as IC, MT, and PD, which are more significant in this study, reveal key drivers of long-term dynamic changes in fire areas. This indicates that the roles of both natural and human factors become more critical in a long-term dynamic perspective, offering new insights into the interactions of various factors influencing wildfire dynamics. Li et al.'s research also recognizes the significant impacts of temperature and population density on forest fire risk [22], which aligns with our findings.

Figure 10 demonstrates that in the study of large-scale forest fire area change trends, reanalysis indicators extracted from fire risk assessment systems play a crucial role in prediction models. These indicators contribute more to the model output than other conventional forest fire factors. This emphasizes the importance of integrating highly relevant and influential specific risk assessment indicators when constructing prediction models. These indicators not only provide deeper insights into the impact on the burn area but also enhance the model's ability to predict dynamic changes in fires. Therefore, effective data integration and feature selection are vital for predicting forest fire areas, and integrating such unconventional forest fire factors in future fire warning models can improve model accuracy.

3.3. Fire Prevention and Control Zoning in the Study Area

The coordinates of the centers of 140 grids in the study area were used to extract 13 indicators that met the collinearity test criteria at corresponding time points, as detailed in Section 2.3. These indicators were used to construct the dataset for fitting. After fitting the dataset, the trained GWO-XGBoost model was used to plot the fire burn area growth rate warning map for March in historical years for the study area. The results were classified using the natural breaks method in ArcMap software, as shown in Figures 11 and 12. Figure 12 shows the statistics after the natural breaks classification of the original data, while Figure 11 shows further classification after calculating the average probability of forest fire area growth within each county.
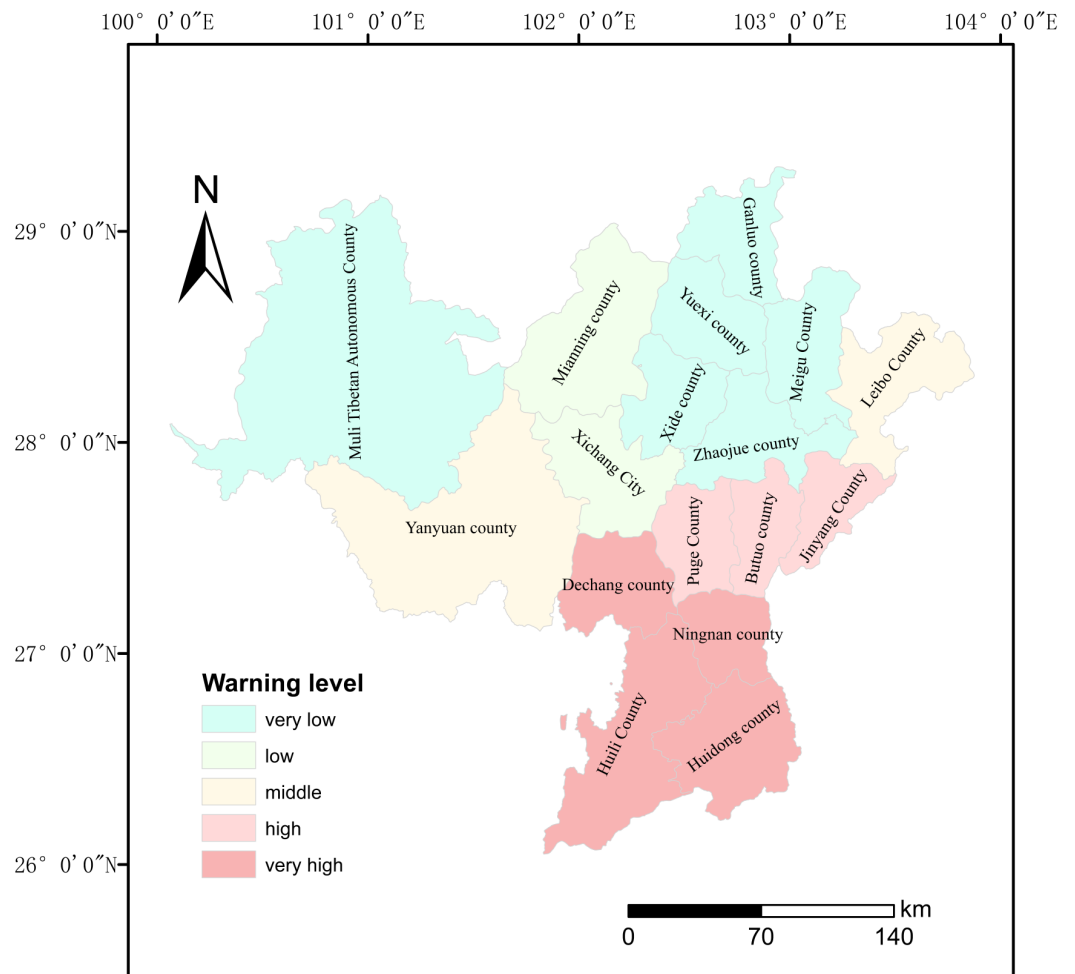
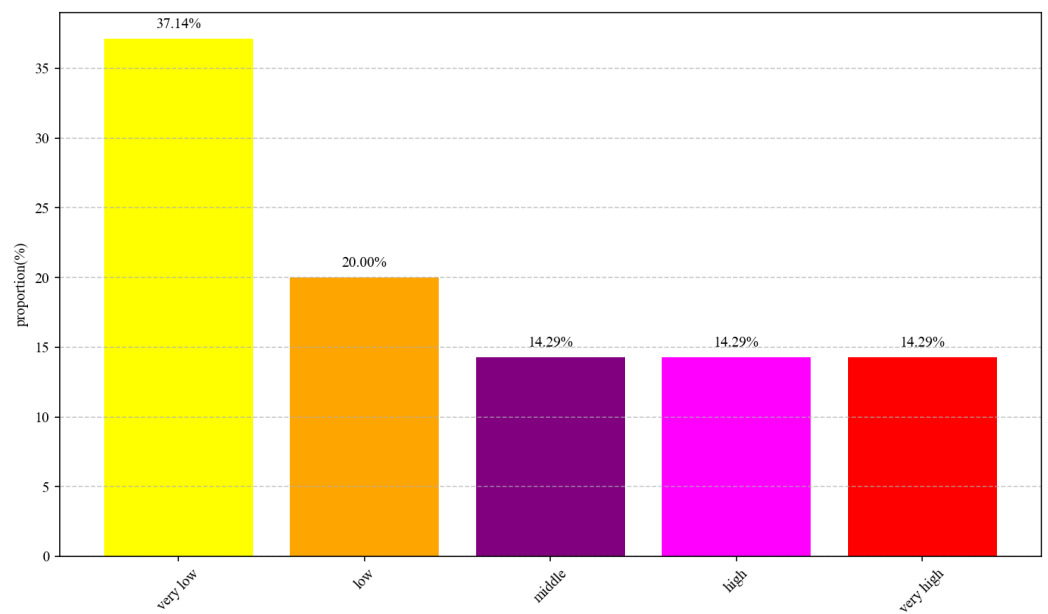**Figure 11.** Fire area growth rate warning map for the study area.

Figure 12. Proportion of fire area growth rate warning levels in the study area.

Figure 11 shows that the southern regions of Liangshan Prefecture, specifically Huili County, Huidong County, Ningnan County, and Dechang County, should be prioritized for

forest fire prevention work. These areas have a high probability of fire burn area growth. Figure 12 illustrates the proportions of fire area growth warning levels, where the high-level threshold is set at 0.41, and the very high-level threshold is set at 0.66. The combined area of these two levels accounts for 31.4% of the total region, indicating the severe fire prevention and control situation in Liangshan Prefecture. Due to the complex terrain and climate conditions in Liangshan Prefecture, relevant departments should focus on fire prevention work in the southern region in March each year with proactive allocation of firefighting resources and fire drill exercises.

4. Conclusions

Forest fire risk management is a crucial task. In recent years, many researchers have used factors such as terrain, climate, vegetation, and socio–cultural aspects to create accurate fire risk maps for specific areas using advanced machine learning techniques. These studies typically focus on predicting the immediate probability of fire occurrence at specific grid points (often high resolution). Creating fire risk maps is highly valuable for disaster prevention and mitigation. However, this approach has certain limitations, mainly because it is challenging to accurately predict the specific starting locations and timing of fires. Currently, fire detection relies primarily on satellite imagery, but due to the Earth's rotation, satellites can only periodically monitor specific locations through remote sensing, making real-time fire monitoring impossible. For dynamic factors like climate, most studies use monthly averages to substitute for the actual conditions when fires occur. Related research has shown that the closer these dynamic climate data are to the real-time conditions during fire occurrences, the better the model's performance.

Given the periodic nature of remote sensing satellite detection, this study adopts a novel approach, exploring whether the burn area at a specific grid point in the same month of the next year increases on an annual cycle. Although this method is similar to traditional fire risk mapping, its uniqueness lies in the broader temporal dimension consideration and its focus on the interannual changes in cumulative burn areas within a region. This approach provides a new perspective for understanding and predicting fire risk, aiding with more effective forest fire prevention and management. It is crucial to emphasize that, while our model is trained on monthly aggregated data, the training process, optimization method, and model structure are still applicable to daily or weekly data. Such predictive capability is essential for effective forest fire management and provides both short-term and long-term benefits. This dual capacity enhances both the rapid response to potential fire outbreaks and the strategic planning for future fire prevention and mitigation.

This study focuses on predicting the burn area growth rate in Liangshan Prefecture, Sichuan Province. During data extraction, in addition to common fire-related factors, a series of reanalysis data from a national-level fire assessment system were introduced, and we selected 22 potential influencing factors in total. After collinearity testing and correlation analysis, 13 key indicators were screened out as input features for the machine learning model. The GWO algorithm was used to optimize the parameters of the XGBoost model, establishing the GWO-XGBoost model, and we compared its performance with those of RF, XGBoost, and LR models. Model evaluation results indicate that the GWO-XGBoost model outperforms the other models in multiple performance metrics, demonstrating good classification fitting ability. To further explain the model's prediction results, the SHAP value analysis method was applied to quantify the contribution of each influencing factor to the model output. The analysis results show that IC, MT, and PD are the most significant indicators affecting the burn area growth trend. Finally, by fitting the March data of historical fire years in Liangshan Prefecture, a fire burn area growth warning map was plotted and validated with the fire risk difference map of the Luzhou area. The results indicate that the study of burn area growth rate prediction in this study is reasonable. Through in-depth analysis of these key indicators, this study provides data-driven insights for local forest fire management, promotes data-based decision-making, and provides a scientific basis for fire prevention and management in Liangshan Prefecture.

References

1. 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]
2. Abatzoglou, J.T.; Battisti, D.S.; Williams, A.P.; Hansen, W.D.; Harvey, B.J.; Kolden, C.A. Projected increases in western US forest fire despite growing fuel constraints. *Commun. Earth Environ.* **2021**, *2*, 227. [CrossRef]
3. Zhang, P.; Yan, P.; Liu, H. A Quantitative Analysis of Chinese and International Studies on Forest Fire Prediction from 2002 to 2019. *J. Wildland Fire Sci.* **2023**, *41*, 53–59.
4. Wu, Y.; Shu, L.; Wang, M.; Zhang, H.; Si, L. A Review of Forest Fires Worldwind in Recent Years. *J. Temp. For. Res.* **2022**, *5*, 49–54.
5. Zheng, B.; Ciais, P.; Chevallier, F.; Chuvieco, E.; Chen, Y.; Yang, H. Increasing forest fire emissions despite the decline in global burned area. *Sci. Adv.* **2021**, *7*, eabh2646. [CrossRef]
6. Walding, N.G.; Williams, H.T.; McGarvie, S.; Belcher, C.M. A comparison of the US National Fire Danger Rating System (NFDRS) with recorded fire occurrence and final fire size. *Int. J. Wildland Fire* **2018**, *27*, 99–113. [CrossRef]
7. Wang, X.; Wotton, B.M.; Cantin, A.S.; Parisien, M.-A.; Anderson, K.; Moore, B.; Flannigan, M.D. cffdrs: An R package for the Canadian forest fire danger rating system. *Ecol. Process.* **2017**, *6*, 5. [CrossRef]
8. Wei, X.; Bai, K.; Chang, N.B.; Gao, W. Multi-source hierarchical data fusion for high-resolution AOD mapping in a forest fire event. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102366. [CrossRef]
9. Liu, S.; Zheng, Y.; Dalponte, M.; Tong, X. A novel fire index-based burned area change detection approach using Landsat-8 OLI data. *Eur. J. Remote Sens.* **2020**, *53*, 104–112. [CrossRef]
10. Simeoni, A.; Salinesi, P.; Morandini, F. Physical modelling of forest fire spreading through heterogeneous fuel beds. *Int. J. Wildland Fire* **2011**, *20*, 625–632. [CrossRef]
11. Ford, A.E.; Harrison, S.P.; Kountouris, Y.; Millington, J.D.; Mistry, J.; Perkins, O.; Rabin, S.S.; Rein, G.; Schreckenberg, K.; Smith, C.; et al. Modelling human-fire interactions: Combining alternative perspectives and approaches. *Front. Environ. Sci.* **2021**, *9*, 649835. [CrossRef]
12. Juang, C.S.; Williams, A.P.; Abatzoglou, J.; Balch, J.; Hurteau, M.; Moritz, M. Rapid growth of large forest fires drives the exponential response of annual forest-fire area to aridity in the western United States. *Geophys. Res. Lett.* **2022**, *49*, e2021GL097131. [CrossRef] [PubMed]
13. Markuzon, N.; Kolitz, S. Data driven approach to estimating fire danger from satellite images and weather information. In Proceedings of the 2009 IEEE Applied Imagery Pattern Recognition Workshop (AIPR 2009), Washington, DC, USA, 14–16 October 2009; pp. 1–7.
14. Huang, J.C.; Ko, K.M.; Shu, M.H.; Hsu, B.M. Application and comparison of several machine learning algorithms and their integration models in regression problems. *Neural Comput. Appl.* **2020**, *32*, 5461–5469. [CrossRef]
15. Gao, D.; Ou, L.; Liu, Y.; Yang, Q.; Wang, H. DeepSpoof: Deep Reinforcement Learning-Based Spoofing Attack in Cross-Technology Multimedia Communication. *IEEE Trans. Multimed.* **2024**, 1–13. [CrossRef]
16. Bhadoria, R.S.; Pandey, M.K.; Kundu, P. RVFR: Random vector forest regression model for integrated & enhanced approach in forest fires predictions. *Ecol. Inform.* **2021**, *66*, 101471.
17. Mohajane, M.; Costache, R.; Karimi, F.; Pham, Q.B.; 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]
18. Hao, Y.; Li, M.; Wang, J.; Li, X.; Chen, J. A High-Resolution Spatial Distribution-Based Integration Machine Learning Algorithm for Urban Fire Risk Assessment: A Case Study in Chengdu, China. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 404. [CrossRef]
19. Shmuel, A.; Heifetz, E. Global Wildfire Susceptibility Mapping Based on Machine Learning Models. *Forests* **2022**, *13*, 1050. [CrossRef]
20. Shmuel, A.; Heifetz, E. A Machine-Learning Approach to Predicting Daily Wildfire Expansion Rate. *Fire* **2023**, *6*, 319. [CrossRef]
21. 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]
22. Li, Y.; Li, G.; Wang, K.; Wang, Z.; Chen, Y. Forest Fire Risk Prediction Based on Stacking Ensemble Learning for Yunnan Province of China. *Fire* **2024**, *7*, 13. [CrossRef]
23. Ghali, R.; Akhloufi, M.A. Deep Learning Approaches for Wildland Fires Using Satellite Remote Sensing Data: Detection, Mapping, and Prediction. *Fire* **2023**, *6*, 192. [CrossRef]

24. Wang, J.; Seyler, B.C.; Ticktin, T.; Zeng, Y.; Ayu, K. An ethnobotanical survey of wild edible plants used by the Yi people of Liangshan Prefecture, Sichuan Province, China. *J. Ethnobiol. Ethnomed.* **2020**, *16*, 10. [CrossRef] [PubMed]

25. Gao, D.; Liu, Y.; Hu, B.; Wang, L.; Chen, W.; Chen, Y.; He, T. Time Synchronization Based on Cross-Technology Communication for IoT Networks. *IEEE Internet Things J.* **2023**, *10*, 19753–19764. [CrossRef]

26. McFayden, C.B.; George, C.; Johnston, L.M.; Wotton, M.; Johnston, D.; Sloane, M.; Johnston, J.M. A case-study of wildland fire management knowledge exchange: The barriers and facilitators in the development and integration of the Canadian Forest Fire Danger Rating System in Ontario, Canada. *Int. J. Wildland Fire* **2022**, *31*, 835–846. [CrossRef]

27. Fujioka, F.M.; Weise, D.R.; Chen, S.C.; Kim, S.H.; Kafatos, M.C. Reaction intensity partitioning: A new perspective of the National Fire Danger Rating System Energy Release Component. *Int. J. Wildland Fire* **2021**, *30*, 351–364. [CrossRef]

28. Hollis, J.J.; Matthews, S.; Fox-Hughes, P.; Grootemaat, S.; Heemstra, S.; Kenny, B.J.; Sauvage, S. Introduction to the Australian Fire Danger Rating System. *Int. J. Wildland Fire* **2024**, *33*, WF23140. [CrossRef]

29. Bargali, H.; Gupta, S.; Malik, D.; Matta, G. Estimation of fire frequency in Nainital District of Utarakhand state by using satellite images. *J. Remote Sens. GIS* **2017**, *6*, 4. [CrossRef]

30. Gupta, S.; Gupta, A. Dealing with noise problem in machine learning data-sets: A systematic review. *Procedia Comput. Sci.* **2019**, *161*, 466–474. [CrossRef]

31. Fu, Y.; Li, R.; Wang, X.; Bergeron, Y.; Valeria, O.; Chavardès, R.D.; Wang, Y.; Hu, J. Fire detection and fire radiative power in forests and low-biomass lands in Northeast Asia: MODIS versus VIIRS Fire Products. *Remote Sens.* **2020**, *12*, 2870. [CrossRef]

32. Lizundia-Loiola, J.; Franquesa, M.; Khairoun, A.; Chuvieco, E. Global burned area mapping from Sentinel-3 Synergy and VIIRS active fires. *Remote Sens. Environ.* **2022**, *282*, 113298. [CrossRef]

33. Huang, S.; Tang, L.; Hupy, J.P.; Wang, Y.; Shao, G. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *J. For. Res.* **2021**, *32*, 1–6. [CrossRef]

34. Meigs, G.W.; Dunn, C.J.; Parks, S.A.; Krawchuk, M.A. Influence of topography and fuels on fire refugia probability under varying fire weather conditions in forests of the Pacific Northwest, USA. *Can. J. For. Res.* **2020**, *50*, 636–647. [CrossRef]

35. 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] [PubMed]

36. Zhang, Y.; Geng, P.; Sivaparthipan, C.; Muthu, B.A. Big data and artificial intelligence based early risk warning system of fire hazard for smart cities. *Sustain. Energy Technol. Assess.* **2021**, *45*, 100986. [CrossRef]

37. Gao, D.; Wang, H.; Guo, X.; Wang, L.; Gui, G.; Wang, W.; Yin, Z.; Wang, S.; Liu, Y.; He, T. Federated Learning Based on CTC for Heterogeneous Internet of Things. *IEEE Internet Things J.* **2023**, *10*, 22673–22685. [CrossRef]

38. Blagus, R.; Lusa, L. SMOTE for high-dimensional class-imbalanced data. *BMC Bioinform.* **2013**, *14*, 106. [CrossRef]

39. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.

40. Hatta, N.; Zain, A.M.; Sallehuddin, R.; Shayfull, Z.; Yusoff, Y. Recent studies on optimisation method of Grey Wolf Optimiser (GWO): A review (2014–2017). *Artif. Intell. Rev.* **2019**, *52*, 2651–2683. [CrossRef]

41. Li, J.; An, X.; Li, Q.; Wang, C.; Yu, H.; Zhou, X.; Geng, Y.-A. Application of XGBoost algorithm in the optimization of pollutant concentration. *Atmos. Res.* **2022**, *276*, 106238. [CrossRef]

42. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.