forests

Article

A Wildfire Detection Algorithm Based on the Dynamic Brightness Temperature Threshold

Yunhong Ding [1,2], Mingyang Wang [1,*], Yujia Fu [1], Lin Zhang [1] and Xianjie Wang [3]

1 College of Information and Computer Engineering, Northeast Forestry University, Harbin 150040, China
2 School of Computer Science and Information Engineering, Harbin Normal University, Harbin 150025, China
3 School of Physics, Harbin Institute of Technology, Harbin 150001, China
* Correspondence: wangmingyang@nefu.edu.cn

Abstract: Satellite remote sensing plays an important role in wildfire detection. Methods using the brightness and temperature difference of remote sensing images to determine if a wildfire has occurred are one of the main research directions of forest fire monitoring. However, common wildfire detection algorithms are mainly based on a fixed brightness temperature threshold to distinguish wildfire pixels and non-wildfire pixels, which reduces the applicability of the algorithm in different space–time regions. This paper presents an adaptive wildfire detection algorithm, DBTDW, based on a dynamic brightness temperature threshold. First, a regression dataset, MODIS_DT_Fire, was constructed based on moderate resolution imaging spectroradiometry (MODIS) to determine the wildfire brightness temperature threshold. Then, based on the meteorological information, normalized difference vegetation index (NDVI) information, and elevation information provided by the dataset, the DBTDW algorithm was used to calculate and obtain the minimum brightness temperature threshold of the burning area by using the Planck algorithm and Otsu algorithm. Finally, six regression models were trained to establish the correlation between factors and the dynamic brightness temperature threshold of wildfire. The root-mean-square error (RMSE) and mean absolute error (MAE) were used to evaluate the regression performance. The results show that under the XGBoost model, the DBTDW algorithm has the best prediction effect on the dynamic brightness temperature threshold of wildfire (leave-one-out method: RMSE/MAE = 0.0730). Compared with the method based on a fixed brightness temperature threshold, the method proposed in this paper to adaptively determine the brightness temperature threshold of wildfire has higher universality, which will help improve the effectiveness of satellite remote fire detection.

Keywords: dynamic brightness temperature threshold; adaptive wildfire detection; remote sensing; MODIS_DT_Fire

Citation: Ding, Y.; Wang, M.; Fu, Y.; Zhang, L.; Wang, X. A Wildfire Detection Algorithm Based on the Dynamic Brightness Temperature Threshold. *Forests* **2023**, *14*, 477. https://doi.org/10.3390/f14030477

Received: 1 January 2023
Revised: 15 February 2023
Accepted: 22 February 2023
Published: 27 February 2023

1. Introduction

Due to climate change [1], human activity, and other factors [2], global forests, grasslands, and other ecosystems are facing a serious threat from wildfires [3–5]. The large amount of soot emissions caused by wildfires affect the global carbon cycle and slow down the realization of carbon neutralization [6–9]. In order to prevent and detect the occurrence of forest fires, scholars have carried out research work based on remote technologies [10–14]. Moderate-resolution imaging spectroradiometry (MODIS) is a commonly used remote sensing monitoring platform that can meet the requirements of forest fire monitoring in terms of the spatial and temporal resolution of Earth observation data [15–18]. On account of MODIS, the fire detection work carried out in this study is mainly based on infrared bands sensitive to ground temperature, such as bands 21, 22, 31, and 32, to detect abnormally high temperature points on the ground according to the fixed brightness temperature threshold. However, due to the influence of buildings [19], vegetation coverage [20], and latitude and longitude [21], the threshold of the critical brightness temperature of wildfire combustion is usually different in different regions and seasons [22]. This phenomenon makes it difficult for a method based on the fixed brightness temperature threshold to achieve effective detection of wildfires in different space–time regions. How to adaptively set the brightness temperature threshold of wildfire in different space–time regions is currently a hot issue in forest wildfire monitoring.

Based on the characteristic difference between wildfires and other typical features in the brightness temperature bands, scholars have developed a variety of wildfire detection methods such as the fixed threshold method, context algorithm, and multidimensional inter-class variance method. Li et al. [23] found that under the burning state of wildfire, the brightness temperature of MODIS channel three rose faster than that of channel four; thus they proposed the fixed threshold method. This method is simple and feasible, but the fire threshold value is obtained according to empirical theory. It is highly dependent on the region, and there are many inaccurate points. Giolio et al. [24] proposed a context-based fire point detection algorithm, which divided the fire point recognition process into two steps: the first step is to divide the pixels in the remote sensing image into six types: data missing, cloud, water, non-ignition, ignition, and uncertainty. The second step is to identify the fire point among the band information of 0.86 μm, 4 μm, and 11 μm. In addition to the above two methods, the method of variance between classes (OTSU) [25], which is often used to separate foreground and background in image processing, is also used to identify the fire point. OTSU is divided into two steps. The first step is to determine the feature type of each pixel point. The second step is to identify the real fire point according to the variance between the suspicious fire point and background pixel.

However, these algorithms are easily affected by regional and seasonal environments and are prone to issues of missed detection and false detection in the application of large space–time and long timespan fire detection. On the other hand, due to the limitations of the fire detection algorithms, the timeliness and effectiveness of remote sensing satellite images in fire detection are greatly limited.

The purpose of this paper is to build a wildfire detection algorithm that can dynamically judge the brightness temperature threshold according to different geographical environments and climate characteristics. In fact, based on the data provided by remote sensing in different bands, the idea of a dynamic brightness temperature threshold has been used to detect sea ice regions and clouds [26–29]. However, few studies have introduced a dynamic threshold into forest wildfire detection. At the same time, in recent years, the rapid development of machine learning technology has brought new research prospects for remote sensing data processing [30–33]. Massive remote sensing data provide unique opportunities for the study of machine learning algorithms to identify wildfires [34,35]. However, existing machine learning algorithms for wildfire detection mainly focus on the classification of wildfire and non-wildfire pixels or pictures, such as related studies on the USTC_SmokeRS dataset [36], WildFires dataset [37], and other unpublished datasets [38,39]. It was indicated through a large number of literature studies that there is no public dataset for the wildfire temperature threshold in remote sensing fire exploration, which urges us to collect and process a large number of MODIS remote sensing images to construct a dataset for studying the wildfire dynamic brightness temperature threshold, which is named MODIS_DT_Fire. The dataset contains hundreds of remote sensing images from MODIS of burning wildfires. The establishment of MODIS_DT_Fire can make up for the lack of a public dataset in the study of the dynamic brightness temperature threshold of wildfire.

Many scholars use machine learning methods to identify remote-sensing images from the perspective of computer and image processing. Scholars in the field of computer vision (CV) synthesize MODIS images' band information of 1, 4, and 3; 21, 2, and 1; or 21, 22, and 31 into true or false color images so that remote sensing images can be identified and classified via conventional methods in the CV field. This pure CV angle causes separation between the existing fire detection algorithm and the traditional remote sensing method. CV does not fully consider the vegetation density, altitude, weather conditions, or other information which can greatly impact fire detection. At the same time, remote sensing images captured by MODIS at night lack data on the visible band, and the shape and size of the burning point are different, which presents difficulties for image recognition. Based on this, this paper proposes a new wildfire detection model based on the dynamic threshold of the brightness temperature (see Figure 1). The method makes full use of multi-source information such as remote sensing images, meteorology, altitude, and NDVI to identify wildfires. Compared to the existing algorithms based on the fixed threshold, context, and multidimensional inter-class variance, the proposed wildfire detection model has a dynamic wildfire threshold detection capability for fires in different regions and seasons, which will greatly alleviate the problem of missed and false detection of fire points over large space–time and long timespans.

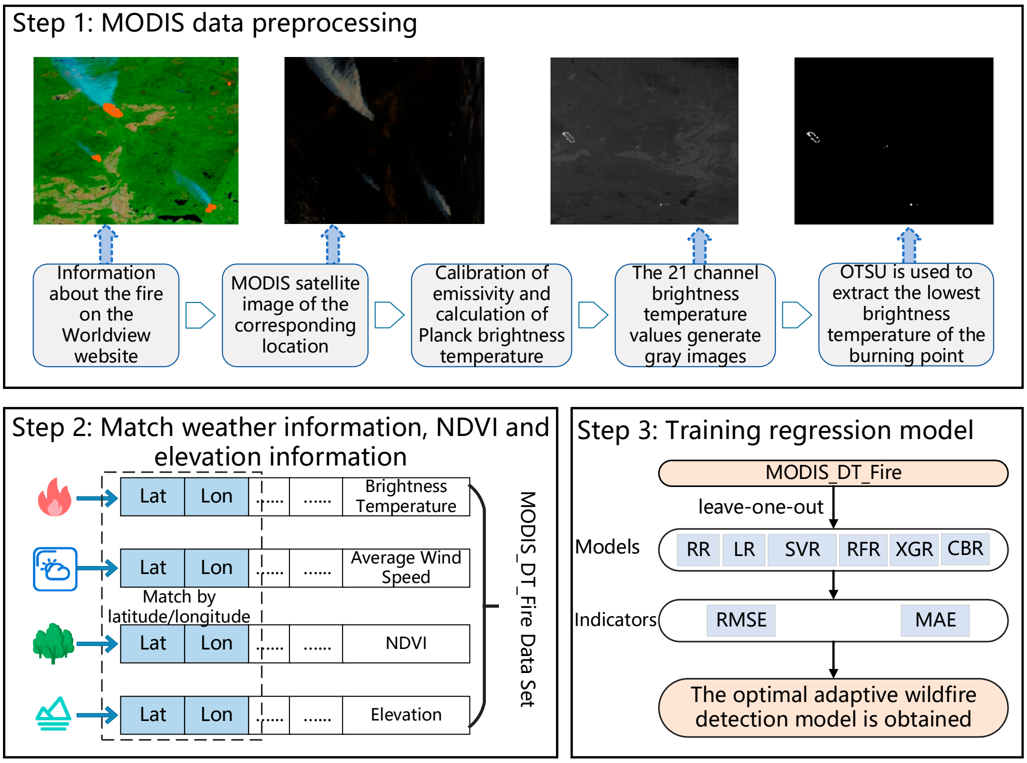

Figure 1. Overall flowchart of adaptive wildfire detection algorithm for dynamic brightness temperature threshold. Step 1 is the extraction process of the minimum brightness temperature of ignition point combustion; Step 2 is the extraction and matching process of meteorological, NDVI, altitude, and other features; Step 3 is the process of regression evaluation of the dataset generated in Step 2. Lat: latitude; Lon: longitude; NDVI: normalized differential vegetation index; RR: ridge regression; LR: least absolute shrinkage and selection operator regression; SVR: support vector machine regression; RFR: random forest regression; XGR: eXtreme gradient boosting regression; CBR: categorical gradient boosting; RMSE: root-mean-square error; MAE: mean absolute error.

2. Materials and Methods

2.1. Data Collection

2.1.1. Fire Points and Remote Sensing Images

In the MODIS_DT_Fire dataset, wildfire information is collected from the Worldview website (https://worldview.earthdata.nasa.gov/ accessed on 25 May 2021). These wildfire events span from 2017 to 2021. On the Worldview website, we extracted the time and latitude of wildfires and recorded them. The corresponding MODIS images are characterized according to fire information from the Earthdata website (https://www.earthdata.nasa.gov/ accessed on 21 June 2021).

2.1.2. Meteorological Information

Meteorological information comes from weather stations all over the world (see Figure 2). These data are international exchange data. According to the 1995 World Meteorological Organization (WMO) proposal with 40 resolutions (https://community.wmo.int/resolution-40 accessed on 25 October 2021), under the international framework provided by the World Meteorological Organization (WMO), member states must coordinate the collection and exchange of information about global atmospheric conditions and promote the free sharing of a batch of meteorological data to relevant researchers and the public by the means outlined in the convention. All data are publicly available on the website of the National Environmental Information Center (NCEI) (https://www.ncei.noaa.gov/ accessed on 25 October 2021), which was established by the United States National Oceanic and Atmospheric Administration (NOAA).

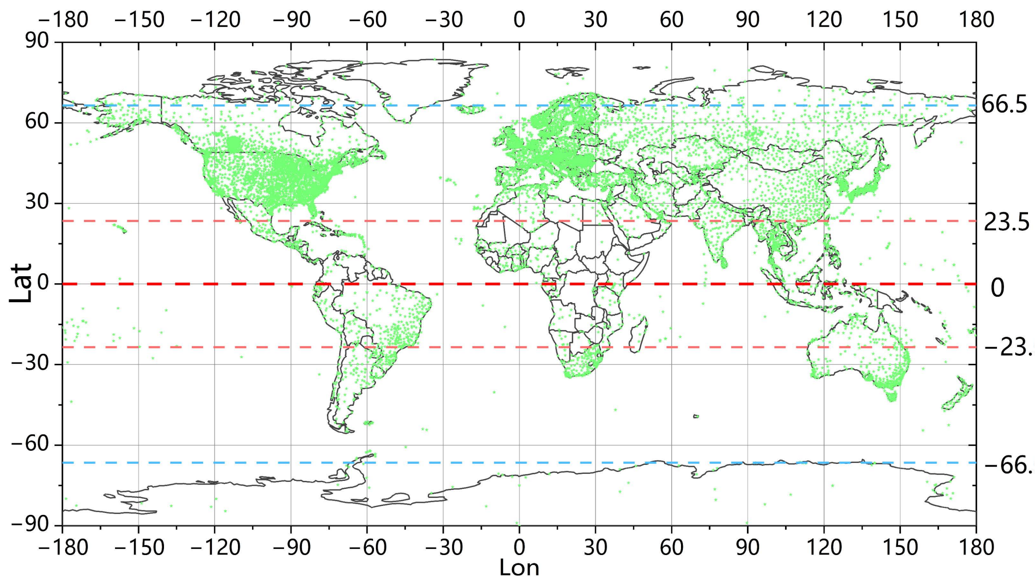

Figure 2. Global distribution of meteorological stations. The green dots represent meteorological stations. The denser the green dots, the more weather stations there are in the area.

2.1.3. Normalized Vegetation Index

NDVI data come from NASA's Earth observation (NEO) website (https://neo.sci.gsfc.nasa.gov/ accessed on 12 January 2022). NEO is part of the Earth Observing System (EOS) satellites Project Science Office located at the NASA Goddard Space Flight Center. The NDVI value is calculated using MODIS/Terra Vegetation Index L3 global 0.05C CMG V006 (MOD13C1 and MOD13C2) products. The world is divided into 3600×1800 squares with a data accuracy of 0.1×0.1 degrees. The timespan is 1 month, ranging from -0.1 to 0.9. There is no unit. The higher values (0.4 to 0.9) indicate land covered with green, leafy vegetation, and the lower values (0 to 0.4) indicate land with little or no vegetation. Figure 3 shows the distribution of the global NDVI index in May 2021.

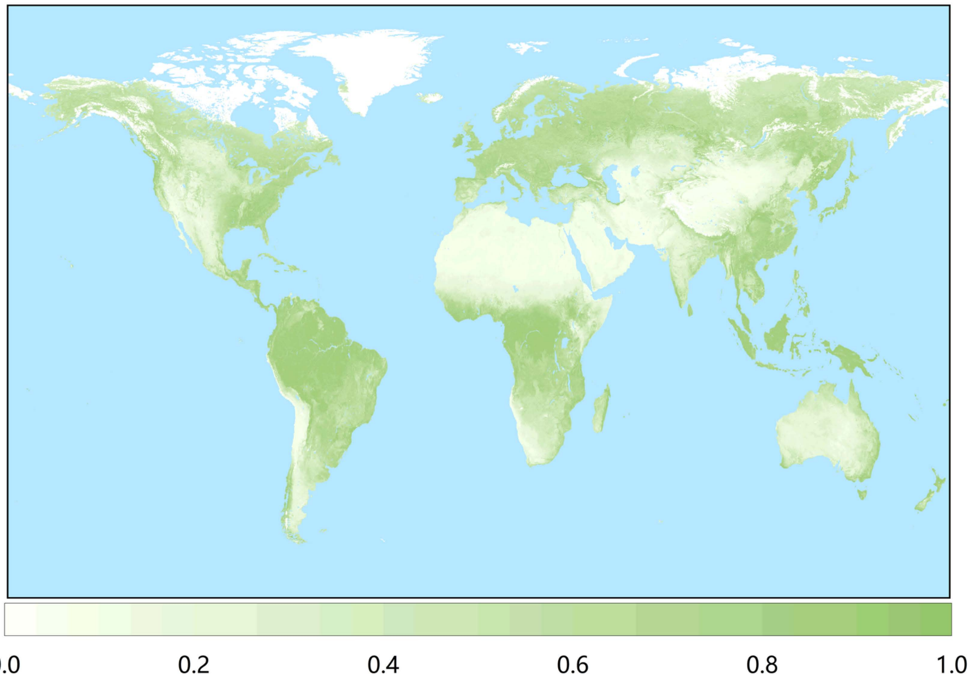

Figure 3. Global Normalized Vegetation Index (NDVI) (May 2021). The closer the color is to 1.0, the larger the area covered by green vegetation; if the color is closer to 0, this means that the area has little or no vegetation cover.

2.1.4. Elevation Information

Elevation information comes from the FAO SOILS PORTAL website (https://www.fao.org/soils-portal/en/ accessed on 15 February 2022) in the Harmonized World Soil Database v1.2 products (HWSD). HWSD was constructed by the Food and Agriculture Organization of the United Nations (FAO) and the International Institute of Applied Systems (IIASA) in Vienna, and has a wide range of applications in hydrology, agriculture, and forest fires [27,40–43]. The product provides 5 min of latitude/longitude grid cells for global altitude distribution, as shown in Figure 4.

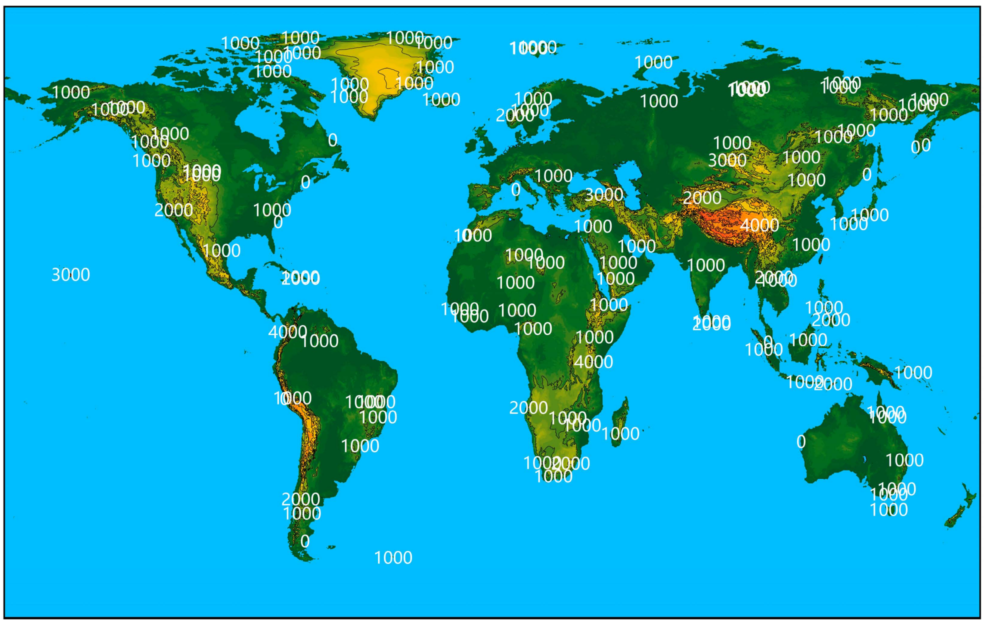

Figure 4. Distribution of global elevation. In the figure, the black boundary is the contour line of the region, and the white text is the elevation mark of the contour line.

2.2. Exporting Fire Point Information

2.2.1. Capturing the Image of the Fire Point

ENVI 5.3 is used to open the collected HDF remote-sensing images of wildfires. For multi-spectral images taken during the day, the data of each band are complete. When the temperature of the ignition point is high, the 7, 2, and 1 channels are used to synthesize RGB false-color images, in which the ignition point is red and the smoke is white. When the ignition point temperature is low, the synthesis of 22 (21), 2, and 1 tees is used, in which the ignition point is black. For the multi-spectral image shot at night, due to the absence of the first 19 bands, channels 21, 22, and 31 are adopted for RGB synthesis, and the ignition points are dark blue, which is close to the gray background and difficult to distinguish [39]. The contrast between the fire and the background can be adjusted using complementary and inverted colors to make the fire easier to identify (see Figure 5).

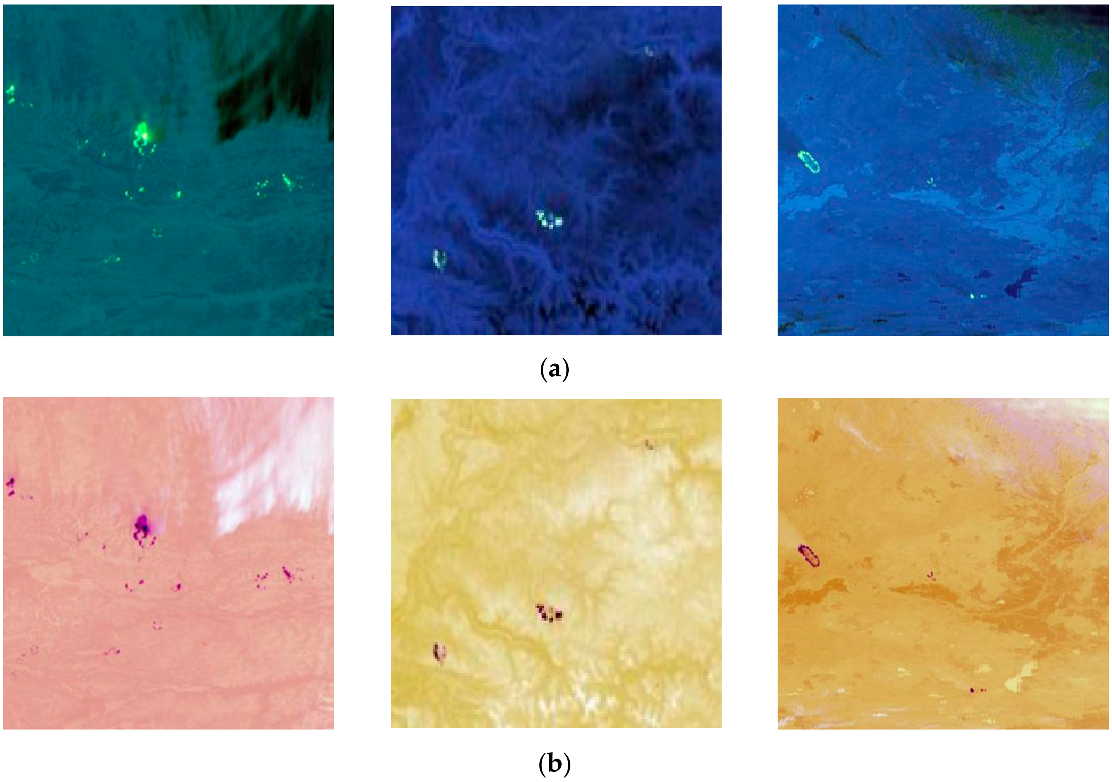

Figure 5. False-color image generated from the MODIS data channel 21/22/31. In Figure 5, (**a**) is a false color image synthesized with 21/22/31 channel primary colors. The bright green area is the fire point, and the blue-purple area is the background area without fire. In the same figure, (**b**) is a complementary picture synthesized by using channel 21/22/31. The purple area is the fire point and the brown area is the background area without fire.

2.2.2. Exporting Each Band DN of Fire Point (Digital Number)

The ROI interception tool in ENVI was used to intercept the fire point region and part of the background region, and the DN value of the multi-spectral image in this region was exported into a comma-separated CSV file format. The exported data contain the latitude, longitude, file X, and file Y of each pixel and the DN values of 36 bands.

2.3. Fire Point Inversion

2.3.1. Calibration

According to the storage characteristics of MODIS L1B data of NASA, the emissivity of band 21 is calculated according to Formula (1):

$$Radiance = Radiance - scaleB \times (SI - Radiance - offsetB), \quad (1)$$

where $Radiance$ is the rate of radiation, $Radiance - scaleB$ is the calibration gain, and $Radiance - offsetB$ is the calibration offset.

2.3.2. Planck Algorithm

The brightness temperature is an equivalent temperature parameter to describe general ground objects that are in a certain band range. When the general ground objects and the absolute blackbody have the same radiation brightness, the temperature of the absolute blackbody represents the temperature of the ground objects. This temperature is called the brightness temperature of the ground objects. The emissivity of band 21 is converted into luminance temperature by Planck's formula. Planck's formula is as follows:

$$T_i = \frac{C_2}{\lambda_i \ln(1 + \frac{C1}{{\lambda_i}^5 R_i})}, \tag{2}$$

where T_i is the temperature of the first channel (K), λ_i is the central wavelength of the channel i, and C_1 and C_2 are constants. $C_1 = 1.19104356 \times 10^{-6} W \cdot m^2$ and $C_2 = 1.4387685 \times 10^4 \mu m \cdot K$.

2.3.3. Gray Level of Brightness Temperature

Take the minimum value in file X and file Y in the export file as the drawing origin. The brightness temperature is linearly mapped to the interval [0, 255] as the grayscale value of the image, and the 21-channel brightness temperature grayscale map is drawn (Figure 6).

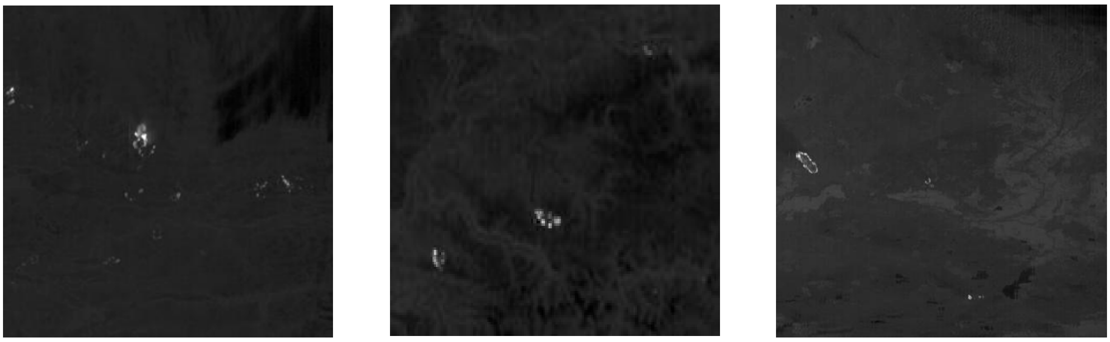

Figure 6. Grayscale of brightness temperature in channels; the bright white area represents the fire points, and the dark black area represents the unburned background area.

2.4. Calculating the Brightness Temperature Threshold

2.4.1. Otsu Algorithm

The Otsu algorithm, a classic algorithm in image binarization, was proposed by Japanese scholar Otsu in 1979 [25]. After image binarization segmentation according to the threshold value obtained by the OTSU method, the interclass variance of foreground and background image is the largest. Therefore, this method is also called the maximum interclass variance method (OTSU). OTSU is simple to calculate and is not affected by image brightness and contrast, so it is considered the best algorithm for threshold selection in image segmentation. In the OTSU algorithm, the range of grayscale values of the grayscale image is $[0, n]$. It takes a point m within $[0, n]$ at random, divides the grayscale histogram into two parts, namely C_0 and C_1, and calculates the probabilities w_0 and w_1 generated by each group, the intra-group mean μ_0 and μ_1 of each group, and the grayscale mean μ of the overall image. The variance between the two groups is

$$\sigma^2(m) = w_0(\mu_0 - \mu)^2 + w_1(\mu_1 - \mu)^2, \tag{3}$$

Varying m from 0 to n, the value of m corresponding to the time when Equation (3) is the maximum value in the requested threshold.

In the experiment, the OTSU method is used to separate the ignition region and non-ignition region. In order to avoid the influence of an abnormally low temperature on some pixels caused by smoke and cloud occlusion, and to expand the influence of high temperature pixels such as fire points on the segmentation results, a high-pass filter (Pixel $\geq$ 60) was used to filter the image before binarization segmentation. The result after binarization is shown in Figure 7. The grayscale image is divided into two parts: low temperature (unburned) and high temperature (burning).

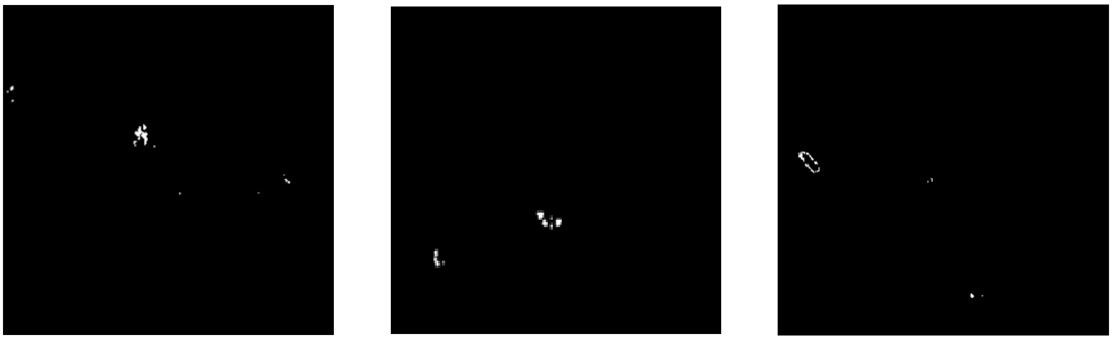

Figure 7. The luminance temperature binarization diagram of 21 channels divided by the OTSU algorithm. The white area of the image is the fire and the black area is the unburned background.

2.4.2. Select the Lowest Brightness Temperature

The pixel value of the fire point area in the segmentation result is inverted into the brightness temperature value, and all the pixels in the fire point area are traversed to find the fire point pixel at the lowest brightness temperature. The longitude and latitude coordinates of the pixel and the brightness temperature value are recorded. The brightness temperature value is the brightness temperature threshold of the fire map.

2.5. Build the Dataset of Wildfire Brightness Temperature Threshold

The nearest meteorological information, altitude, and NDVI were matched according to the latitude, longitude, and date of the fire. NDVI is monthly. Considering the effectiveness of fire point detection, the generation of NDVI this month lags behind that of the fire point. Therefore, the NDVI of the last month is used as the feature of the dataset in the experiment. For fires that are missing weather on the day, the nearest weather information to the date is matched. After collection and sorting, a total of 184 samples were obtained, and each sample contained 10 characteristic values: longitude (Lon), dimension (Lat), daily mean temperature (DMT), dew point temperature (DPT), mean wind speed (MeanWS), maximum sustained wind speed (MaxWS), daily maximum temperature (MaxT), daily minimum temperature (MinT), last month's NDVI (LNDVI), and altitude (ELE). Detailed explanations of each feature are shown in Table 1. Figure 8 shows the global distribution of all sample points in the MODIS_DT_Fire dataset.

Closer distances better ensure the accuracy of meteorological information. Among the 184 bright temperature threshold samples in the dataset, the farthest distance between the fire location and the weather station is 334.36 km, the closest distance is 5.51 km, and the average distance is 56.52 km. Out of all 184 sample sites used in the experiments, only 5 sample points have a distance of more than 150 km. (Figure 9). Since some of the fire sites are located in sparsely populated mountains or dense forests, the lack of closer meteorological observation stations is unavoidable, and we have tried to avoid such problems as much as possible in the dataset processing. For the timeliness of meteorological information, excluding "003-MOD021KM.A2018220.0550.061.2018220134027-12" and "059-MOD021KM.A2021210.0225.A2021210.0225.061.2021210134827-3", all the samples are real-time meteorological data from the nearest meteorological station on the day of fire data collection.

Table 1. Detailed explanation of the characteristics of the MODIS_DT_Fire dataset.

	Name (Unit)	Explanation	Range
Location	Longitude (degree)	Longitude of the firepoint pixel	−142.22~39.62
	Latitude (degree)	Latitude of the firepoint pixel	−14.98~66.05
Weather	Daily Average Temperature (Fahrenheit)	Meteorological information for the nearest weather station to the fire point image element for the nearest date	31.9~97.8
	Dewpoint Temperature (Fahrenheit)		−12.4~75.1
	Average Wind Speed (knots)		0~17.7
	Maximum Sustain Wind Velocity (knots)		3.9~33
	Daily Maximum Temperature (Fahrenheit)		39.2~116.2
	Daily Minimum Temperature (Fahrenheit)		17.1~88.9
Index of vegetation	Normalized Differential Vegetation Index (no unit)	Index to assess vegetation densities	0.136~0.865
Elevation Information	Altitude (meter)		0~5356
Brightness temperature threshold	Minimum Brightness Temperature of Ignition Point (centigrade)	The minimum brightness temperature at the ignition point	7.01~182.48

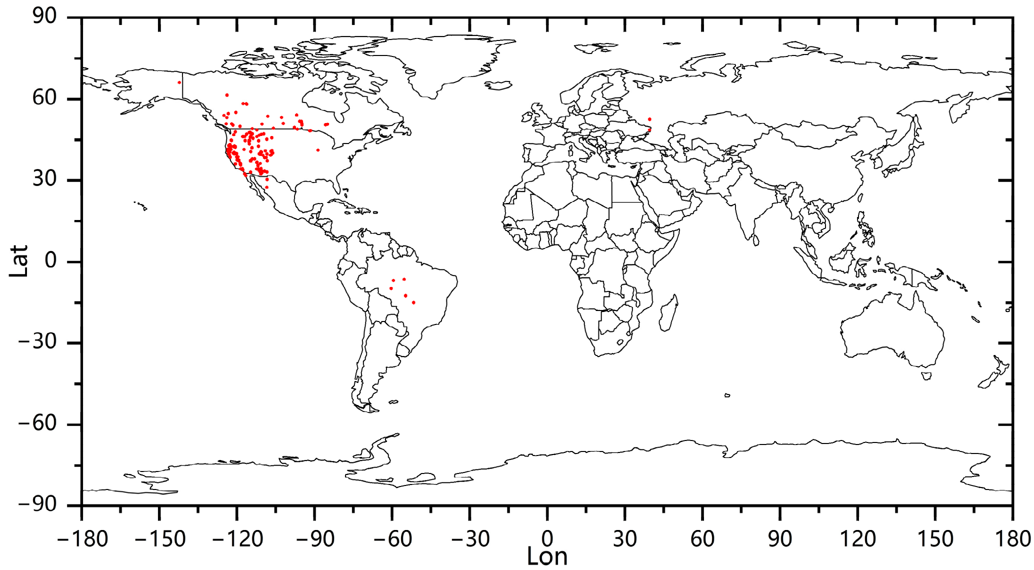

Figure 8. Global distribution map of all dynamic brightness temperature threshold sample points in the MODIS_DT_Fire dataset. The red dots are sample points.

2.6. Regression

For the newly established brightness temperature threshold regression dataset MODIS_DT_Fire, six classical regression models were used to verify the influence of each feature on the brightness temperature threshold.

2.6.1. Ridge Regression (RR)

Ridge regression (RR) is a partial estimation regression method dedicated to collinear data analysis, which is essentially an improved least squares estimation model [44]. RR gives up the unbiased nature of the least square method and obtains the regression coefficient at the cost of losing part of the information and reducing the accuracy. The fitting of ill-conditioned data is better than that of the least square method.

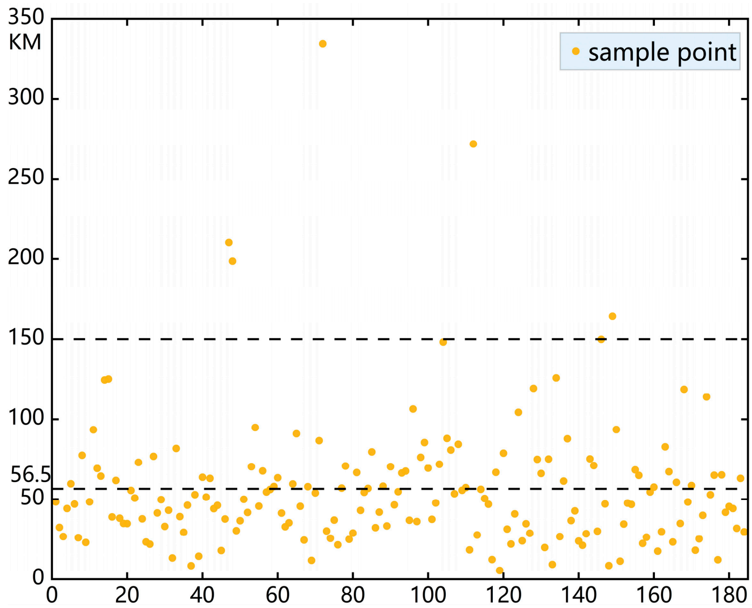

Figure 9. The distance between the location of wildfires in the MODIS_DT_Fire dataset and the meteorological station sites. The horizontal dashed line represented by 56.5 in the figure is the average of the distances.

2.6.2. Lasso Regression (LR)

Least absolute selection and shrinkage operator regression (lasso regression, LR) is a compression estimation method in the idea of reducing the variable set (order reduction) [45]. By constructing a penalty function, lasso regression can compress the coefficients of variables and make some regression coefficients become 0, thus achieving the purpose of variable selection.

2.6.3. Support Vector Machine Regression (SVR)

Support vector machine regression (SVR) works on the same principle as SVM [46]. The basic idea of SVR is to find the best-fitting line. In SVR, the best-fitting line is the hyperplane with the most points. The main advantage of SVR is that its computational complexity does not depend on the dimensions of the input space.

2.6.4. Random Forest Regression (RFR)

The random forest regression (RFR) model builds multiple unrelated decision trees by randomly extracting samples and features, and obtains prediction results through parallel methods [47]. Each decision tree can obtain a prediction result through the samples and features extracted, and the regression prediction result of the whole forest can be obtained by integrating the results of all trees and taking the average value.

2.6.5. eXtreme Gradient Boosting (XGR)

eXtreme gradient boosting (XGBoost) in the GBDT (gradient boosting decision tree, GBDT) model adds a rule item to the loss function to find the best solution and avoid overfitting [48,49]. The XGBoost model can be used for classification problems and regression problems according to the selection of different loss functions. In this study, XGBoost regression (XGR) was used to fit the minimum temperature dynamic distribution threshold curve of the fire points.

2.6.6. Categorical Boosting (CBR)

Categorical boosting (CatBoost) is a categorical framework GBDT based on symmetric decision trees, which has fewer parameters, supports categorical variables and high accuracy, and can handle categorical features efficiently and reasonably [50]. In addition, CatBoost regression (CBR) can also solve the problem of gradient bias and prediction shift, thus reducing the occurrence of overfitting and improving the accuracy and generalization ability of the model.

2.7. Experiment Details

2.7.1. Operating Environment

The running environment of the experiment was as follows: AMD Ryzen 7 4800H 8-core 16-thread 4.2 GHz; GPU: RTX 2060 6 G; operating system: Windows 11; memory: 16 G, hard disk: 512 G. Python version 3.8 was used for programming. The compiler used Visual Studio Code.

2.7.2. Model Parameter Settings

After parameter tuning, the optimal parameter configuration in each model is finally determined, as shown in Table 2.

Table 2. Parameter settings of each regression model.

Name	Parameters
Ridge Regression (RR)	Alpha = 0.01, tol = 10^{-5}, max_iter = 10,000
Lasso Regression (LR)	Alpha = 0.01, tol = 10^{-5}, max_iter = 10,000
Support Vector Machine Regression (SVR)	Kernel = "poly", degree = 2, epsilon = 0.01, gamma = "scale", max_iter = 10,000
Random Forest Regression (RFR)	n_estimators = 100, criterion = 'mse', min_samples_split = 2, min_samples_leaf = 100
XGBoost Regression (XGR)	learning_rate = 0.01, n_estimators = 1000, max_depth = 3, early_stopping_rounds = 100, eval_metric = "logloss", verbose = True
CatBoost regression (CBR)	learning_rate = 0.001, depth = 10, l2_leaf_reg = 0.01, grow_policy = 'Lossguide'

2.7.3. Dataset Partitioning

In order to obtain the experimental results closest to the real situation, the dataset is divided by the leave-one method. That is, only one sample is used as the test set each time, and all the remaining samples are used as the training set. All samples are iterated successively, and the score of the evaluation index is averaged as the final score of the model.

2.7.4. Evaluation Indicators

The root-mean-square error (RMSE), mean absolute error (MAE), and coefficient of determination are the most commonly used indicators to evaluate regression performance. Due to the data division method of the leave-one method, only one sample is used for the test set at a time, which makes it impossible to calculate. Therefore, this paper only adopts the indicators RMSE and MAE to conduct the evaluation.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{4}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{5}$$

3. Results

3.1. Evaluation Results of Six Regression Models

Based on the MODIS_DT_Fire dataset, six regression models (RR, LR, SVMR, RFR, XGR, CBR) were trained by the leave-one method in the experiment. In the leave-one method, the number of each test set is one, that is, RMSE = MAE. Figure 10 shows the scores of each model under the RMSE/MAE index. It can be seen that among the six regression models, the XGR model has the lowest RMSE/MAE score (0.0730) and the best-fitting effect on the wildfire threshold in the MODIS_DT_Fire dataset. There was not much difference between the scores of the other five models.

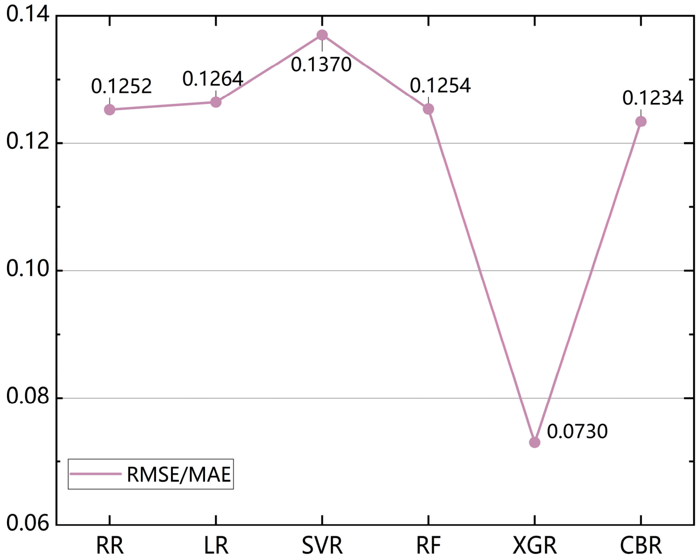

Figure 10. RMSE/MAE scores when six regression models were trained using the leave-one method.

In order to further verify the performance of the model, all samples were used to train the six regression models in the experiment. The RMSE and MAE scores of the models are shown in Figure 11. The results showed that under RMSE/MAE evaluation, XGR (0.1205/0.0915) and CBR (0.1163/0.0891) had the lowest score and the best effect, and the model performance was much higher than the other four types.

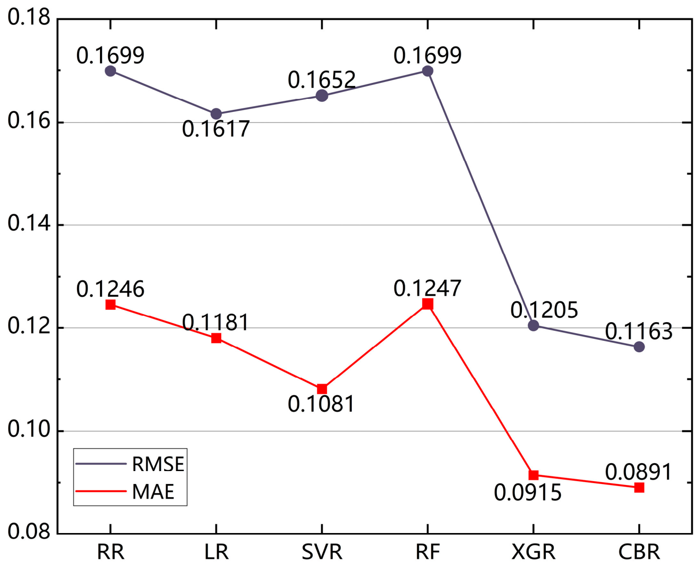

Figure 11. RMSE and MAE scores of the six regression models trained in all datasets.

3.2. Feature Sorting

Among the six regression models, the XGR and CBR models show better regression performance, so this paper uses these two models to sort the features. The weights of each feature are shown in Figure 12.

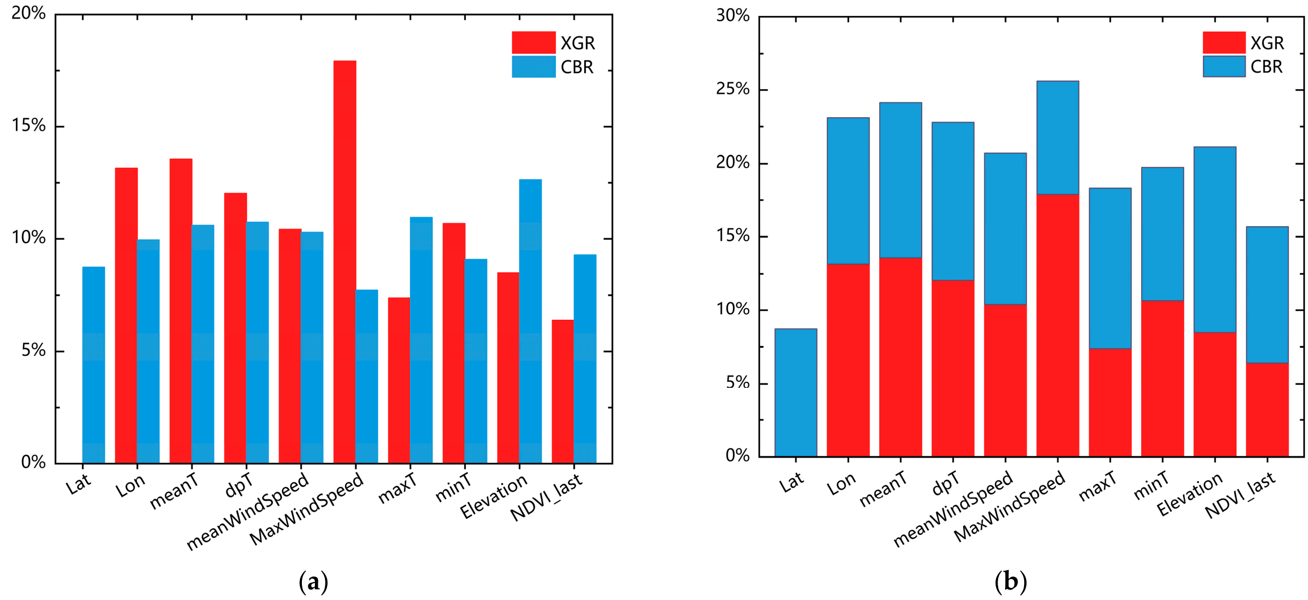

Figure 12. Generated feature weight diagram of XGR and CBR. (**a**) represents the weight values of each feature separately calculated by CBR model and XGR model. (**b**) represents the sum of weight values obtained by XGR and CBR models weighted 1:1; XGR, eXtreme gradient boosting regression, CBR: categorical gradient boosting.

Figure 12a represents the sum of the weight values obtained by the 1:1 weighting of XGR and CBR. In Figure 12b, the weight values in order from high to low are as follows: daily maximum wind speed, daily average temperature, longitude, dew point temperature, altitude, daily average wind speed, daily minimum temperature, daily maximum temperature, last month NDVI, and latitude. The blue section in Figure 12a shows the feature weight value separately calculated by the CBR model. The ranking from high to low is as follows: altitude (0.126), daily maximum temperature (0.110), dew point temperature (0.108), daily mean temperature (0.106), daily mean wind speed (0.103), longitude (0.100), last month NDVI (0.093), daily minimum temperature (0.091), latitude (0.087), and daily maximum wind speed (0.077). The red section in Figure 12a shows the feature weight value separately calculated by the XGR model. The ranking from high to low is as follows: daily maximum wind speed (0.179), daily mean temperature (0.136), longitude (0.131), dew point temperature (0.120), daily minimum temperature (0.107), daily mean wind speed (0.104), altitude (0.085), daily maximum temperature (0.0738), last month's NDVI (0.064), and latitude (0.00).

It can be seen that the weight of each feature calculated according to the CBR model is relatively average, and the weight of different features has little variation. The difference between the highest weight value (altitude, 0.126) and the lowest weight value (maximum daily wind speed, 0.077) is 0.049, and the standard deviation of the weight of 10 feature values is 0.013. For the XGR model, the calculated weights of each feature differ greatly. The difference between the highest weight value (maximum daily wind speed, 0.179) and the lowest weight value (latitude, 0.00) is 0.179, and the standard deviation of the weight of the 10 characteristic values is 0.046. In addition, the latitude index that plays a role in the CBR model does not make any contribution in the XGR model. This shows that the XGR model can achieve the same optimal regression effect as the MAE model with less feature information.

Therefore, considering that the XGR model had a good regression effect by using fewer features, this paper uses the XGR model to realize the regression prediction of the wildfire brightness temperature threshold in the DBTDW algorithm. According to the regression results of the XGR model, the maximum daily wind speed, daily average temperature, longitude, and dew point temperature are the four characteristics that have the greatest impact on the determination of the ignition temperature threshold.

4. Discussion

In order to realize the dynamic estimation of wildfire brightness temperature thresholds detected by remote sensing, we collected hundreds of MODIS remote sensing images of wildfires, and combined meteorological information, elevation data, and NDVI index to organize the MODIS_DT_Fire dataset, which to some extent makes up for the lack of the current research of wildfire dynamic brightness temperature threshold prediction datasets. At the same time, multi-band information is used in our dataset, making full use of the advantage of remote sensing images to provide multi-band data information. The collected dataset and XGR regression model provide a new method and idea for satellite remote sensing technology to detect wildfires.

In order to achieve accurate detection of wildfires and reduce missed and false detections, we used multiple bands to detect wildfires, but also combine other information affecting the occurrence of wildfires, such as the temperature, humidity, precipitation, altitude, etc. Previous studies have mentioned that climatic conditions [51] and altitude [52] are important conditions affecting the occurrence of wildfires, but the current remote sensing image data largely ignore this point and only use RGB images synthesized by several remote sensing bands for fire judgment. On the other hand, the dataset used to judge the occurrence of wildfires from meteorological conditions and other information ignores a large number of extensive multi-spectral images provided by remote sensing. Based on these two considerations, we formed the MODIS_DT_Fire dataset after a long period of collection and collation by combining their common advantages.

Regarding the issue that the brightness temperature threshold is difficult to determine in the application of large space–time and long timespan wildfire detection, we adopted six regression models to verify the correlation between each characteristic value in MODIS_DT_Fire and the brightness temperature threshold. The results show that the regression model can effectively predict the brightness temperature threshold in wildfire detection. The evaluation results of the XGR model show that the XGR can accurately predict the temperature threshold of wildfire brightness in a large range and over a large timespan. In addition, the XGR model can reach the processing speed of hundreds of items per second under the conditions of conventional performance computer computation, which has extremely high performance and price advantages compared with the deep learning model in fire point recognition [53,54].

In conclusion, the proposed XGR prediction model of the wildfire brightness temperature threshold based on the MODIS_DT_Fire dataset shows superior performance in wildfire detection. The new dataset collected in this study provides new ideas for scholars in the field of wildfire detection. In the future, we will further expand the breadth and quantity of the new dataset, and investigate other new features that can be further integrated to benefit fire detection. Meanwhile, other remote sensing researchers are welcome to mine and expand the MODIS_DT_Fire dataset.

5. Conclusions

In this paper, we propose a dynamic threshold of the brightness temperature wildfire detection model DBTDW based on MODIS images. Due to the complex geographical environment and long timespan, the traditional fire-point identification method based on a multi-channel fixed threshold easily causes missed and false detections, thus reducing the accuracy of remote sensing monitoring wildfire identification. In this paper, a new dataset (MODIS_DT_Fire) is constructed to judge the critical brightness temperature threshold of the fire point. Based on the MODIS band information, the dataset is further integrated with meteorological information from other sources, the NDVI index, and elevation information. Six regression models were used in the experiment to explore the correlation between the dimensionality of the dataset and the brightness temperature value. In order to verify the model performance, six models were trained on the dataset using the leave-one method, and the model performance was evaluated using RMSE (MAE). The results show that the regression model can effectively predict the brightness temperature by using the eigenvalues of each dimension in the dataset. The XGR model has the best fitting prediction effect in the leave-one-out method, and its RMSE/MAE score is 0.0730. In the regression training of all datasets, the CBR model showed the optimal brightness temperature threshold prediction effect (RMSE = 0.1163, MAE = 0.0891). Based on the weight values of each characteristic parameter of the XGR model, it is found that the maximum daily wind speed, daily average temperature, longitude, and dew point temperature are the four characteristics that have the greatest influence on the ignition temperature threshold. We have now made MODIS_DT_Fire public in the hope that more remote sensing practitioners will explore the potential of this dataset. In the future, we will further integrate other features that have an impact on wildfire detection, improve the regression model, and further improve the performance of wildfire detection models.

Author Contributions: Y.D.: conceptualization, data curation, formal analysis, investigation, methodology, validation, resources, writing—original draft. M.W.: software, supervision, visualization, writing—review and editing. Y.F.: writing—review and editing L.Z.: methodology, writing—review and editing. X.W.: writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China under Grant 71473034, and the Heilongjiang Provincial Natural Science Foundation of China under Grant LH2019G001.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset has been publicly released at https://pan.baidu.com/s/19zWAsDvs3q-lAKya5ilvdA?pwd=3cmu.

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

References

1. Dabanli, I. *The Relationship between Climate Change and Increasing Wildfires*; Forest Fires: Causes, Effects, Monitoring, Precautions and Rehabilitation Activities; Turkish Academy of Science: Ankara, Turkey, 2021; pp. 25–42.
2. Kountouris, Y. Human activity, daylight saving time and wildfire occurrence. *Sci. Total Environ.* **2020**, *727*, 138044. [CrossRef] [PubMed]
3. Hartter, J.; Hamilton, L.C.; Ducey, M.J.; Boag, A.E.; Salerno, J.D.; Christoffersen, N.D.; Oester, P.T.; Palace, M.W.; Stevens, F.R. Finding common ground: Agreement on increasing wildfire risk crosses political lines. *Environ. Res. Lett.* **2020**, *15*, 065002. [CrossRef]
4. Baker, S.J. Fossil evidence that increased wildfire activity occurs in tandem with periods of global warming in Earth's past. *Earth-Sci. Rev.* **2022**, *224*, 103871. [CrossRef]
5. Craig, C.A.; Allen, M.W.; Feng, S.; Spialek, M.L. Exploring the impact of resident proximity to wildfires in the northern Rocky Mountains: Perceptions of climate change risks, drought, and policy. *Int. J. Disaster Risk Reduct.* **2020**, *44*, 101420. [CrossRef]
6. Landry, J.-S.; Matthews, H.D.; Ramankutty, N. A global assessment of the carbon cycle and temperature responses to major changes in future fire regime. *Clim. Chang.* **2015**, *133*, 179–192. [CrossRef]
7. Tanentzap, A.J.; Burd, K.; Kuhn, M.; Estop-Aragones, C.; Tank, S.; Olefeldt, D. Aged soils contribute little to contemporary carbon cycling downstream of thawing permafrost peatlands. *Glob. Chang. Biol.* **2021**, *27*, 5368–5382. [CrossRef]
8. Mason, K.E.; Oakley, S.; Street, L.E.; Arróniz-Crespo, M.; Jones, D.L.; DeLuca, T.H.; Ostle, N.J. Boreal Forest Floor Greenhouse Gas Emissions Across a Pleurozium schreberi-Dominated, Wildfire-Disturbed Chronosequence. *Ecosystems* **2019**, *22*, 1381–1392. [CrossRef]
9. Kirdyanov, A.V.; Saurer, M.; Siegwolf, R.; Knorre, A.A.; Prokushkin, A.S.; Churakova, O.V.; Fonti, M.V.; Büntgen, U. Long-term ecological consequences of forest fires in the continuous permafrost zone of Siberia. *Environ. Res. Lett.* **2020**, *15*, 034061. [CrossRef]

10. Nedkov, R. Quantitative Assessment of Forest Degradation after Fire Using Ortogonalized Satellite Images from Sentinel-2. *Comptes Rendus l'Acad. Bulg. Sci.* **2018**, *71*, 83–86.
11. Sitnov, S.A.; Mokhov, I.I. A Comparative Analysis of the Characteristics of Active Fires in the Boreal Forests of Eurasia and North America Based on Satellite Data. *Izv. Atmos. Ocean. Phys.* **2018**, *54*, 966–978. [CrossRef]
12. Jang, E.; Kang, Y.; Im, J.; Lee, D.-W.; Yoon, J.; Kim, S.-K. Detection and Monitoring of Forest Fires Using Himawari-8 Geostationary Satellite Data in South Korea. *Remote Sens.* **2019**, *11*, 271. [CrossRef]
13. Wang, H.; Zhang, X.; Xue, W.; Qin, C.; Wu, Y.; Wang, S.; Qiu, P. Evaluation of forest fire damage based on Sentinel-2 images. In Proceedings of the International Conference on Environmental Remote Sensing and Big Data (ERSBD), Wuhan, China, 9 December 2021. [CrossRef]
14. Guzel, A.; Bicakli, K.; Bicakli, F.; Kaplan, G. Monitoring the Regeneration Process of Areas Destroyed by Forest Fires Aided by Google Earth Engine. *Kast. Univ. J. For. Fac.* **2021**, *21*, 122–130. [CrossRef]
15. Lizundia-Loiola, J.; Oton, G.; Ramo, R.; Chuvieco, E. A spatio-temporal active-fire clustering approach for global burned area mapping at 250 m from MODIS data. *Remote Sens. Environ.* **2020**, *236*, 111493. [CrossRef]
16. Ramo, R.; Roketa, E.; Bistinas, I.; van der Werf, G. African burned area and fire carbon emissions are strongly impacted by small fires undetected by coarse resolution satellite data. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2011160118. [CrossRef]
17. Seydi, S.T.; Akhoondzadeh, M.; Amani, M.; Mahdavi, S. Wildfire Damage Assessment over Australia Using Sentinel-2 Imagery and MODIS Land Cover Product within the Google Earth Engine Cloud Platform. *Remote Sens.* **2021**, *13*, 220. [CrossRef]
18. Talucci, A.C.; Loranty, M.M.; Alexander, H.D. Siberian taiga and tundra fire regimes from 2001–2020. *Environ. Res. Lett.* **2022**, *17*, 025001. [CrossRef]
19. Karagianni, A.C.; Lazaridou, M.A. Remote sensing techniques in monitoring areas affected by forest fire. In Proceedings of the 5th International Conference on Remote Sensing and Geoinformation of the Environment (RSCy), Paphos, Cyprus, 6 September 2017. [CrossRef]
20. Wu, R.; Zhao, J.; Zhang, H.; Guo, X.; Ying, H.; Deng, G.; Li, H. Wildfires on the Mongolian Plateau: Identifying Drivers and Spatial Distributions to Predict Wildfire Probability. *Remote Sens.* **2019**, *11*, 2361. [CrossRef]
21. Adekpedjou, A.; Niang, S.D. Semiparametric estimation with spatially correlated recurrent events. *Scand. J. Stat.* **2021**, *48*, 1097–1126. [CrossRef]
22. Bergonse, R.; Oliveira, S.; Goncalves, A.; Nunes, S.; da Camara, C.; Zezere, J.L. A combined structural and seasonal approach to assess wildfire susceptibility and hazard in summertime. *Nat. Hazards* **2021**, *106*, 2545–2573. [CrossRef]
23. Li, Z.Q.; Cihlar, J.; Moreau, L.; Huang, F.; Lee, B. Monitoring fire activities in the boreal ecosystem. *J. Geophys. Res.-Atmos.* **1997**, *102*, 29611–29624. [CrossRef]
24. Giglio, L.; Descloitres, J.; Justice, C.; Kaufman, Y.J. An enhanced contextual fire detection algorithm for MODIS. *Remote Sens. Environ.* **2003**, *87*, 273–282. [CrossRef]
25. Otsu, N. Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [CrossRef]
26. Jin, D.; Lee, K.-S.; Choi, S.; Seo, M.; Lee, D.; Kwon, C.; Kim, H.; Lee, E.; Han, K.-S. Determination of dynamic threshold for sea-ice detection through relationship between 11 μm brightness temperature and 11–12 μm brightness temperature difference. *Korean J. Remote Sens.* **2017**, *33*, 243–248.
27. Daxiang, X.; Debao, T.; Xiongfei, W.; Qiao, W. A Dynamic Threshold Cloud Detecting Approach Based On The Brightness Temperature From Fy-2 Vissr Data. In Proceedings of the 36th International Symposium on Remote Sensing of the Environment (ISRSE), Berlin, Germany, 11–15 May 2015.
28. Deng, Z.; Zhang, G. An Improved Forest Fire Monitoring Algorithm with Three-Dimensional Otsu. *IEEE Access* **2021**, *9*, 118367–118378. [CrossRef]
29. Giglio, L.; Schroeder, W.; Justice, C.O. The collection 6 MODIS active fire detection algorithm and fire products. *Remote Sens. Environ.* **2016**, *178*, 31–41. [CrossRef]
30. Gibson, R.; Danaher, T.; Hehir, W.; Collins, L. A remote sensing approach to mapping fire severity in south-eastern Australia using sentinel 2 and random forest. *Remote Sens. Environ.* **2020**, *240*, 111702. [CrossRef]
31. Jaafari, A.; Zenner, E.; Panah, M.; Shahabi, H. Hybrid artificial intelligence models based on a neuro-fuzzy system and meta-heuristic optimization algorithms for spatial prediction of wildfire probability. *Agric. For. Meteorol.* **2019**, *266*, 198–207. [CrossRef]
32. Jain, P.; Coogan, S.C.P.; Subramanian, S.G.; Crowley, M.; Taylor, S.; Flannigan, M.D. A review of machine learning applications in wildfire science and management. *Environ. Rev.* **2020**, *28*, 478–505. [CrossRef]
33. Milanovic, S.; Markovic, N.; Pamucar, D.; Gigovic, L.; Kostic, P.; Milanovic, S.D. Forest Fire Probability Mapping in Eastern Serbia: Logistic Regression versus Random Forest Method. *Forests* **2021**, *12*, 5. [CrossRef]
34. Huot, F.; Hu, R.L.; Goyal, N.; Sankar, T.; Ihme, M.; Chen, Y.-F. Next Day Wildfire Spread: A Machine Learning Dataset to Predict Wildfire Spreading From Remote-Sensing Data. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4412513. [CrossRef]
35. Sayad, Y.O.; Mousannif, H.; Moatassime, H.A. Predictive modeling of wildfires: A new dataset and machine learning approach. *Fire Saf. J.* **2019**, *104*, 130–146. [CrossRef]
36. Ba, R.; Chen, C.; Yuan, J.; Song, W.; Lo, S. SmokeNet: Satellite Smoke Scene Detection Using Convolutional Neural Network with Spatial and Channel-Wise Attention. *Remote Sens.* **2019**, *11*, 1702. [CrossRef]
37. Toulouse, T.; Rossi, L.; Celik, T.; Akhloufi, M. Automatic fire pixel detection using image processing: A comparative analysis of rule-based and machine learning-based methods. *Signal Image Video Process.* **2016**, *10*, 647–654. [CrossRef]

38. Toan, N.T.; Cong, P.T.; Huang, N.Q.V.; Jo, J. A deep learning approach for early wildfire detection from hyperspectral satellite images. In Proceedings of the IEEE 7th International Conference on Robot Intelligence Technology and Applications (RiTA), Daejeon, Republic of Korea, 1–3 November 2019.
39. Dun-zhu, Z.; Ba, L.; Wang, C. Study on the Application of EOS/MODIS Data under the Support of 3S Technology in the Monitoring of Forest Fire in Tibet. *Anhui Agric. Sci.* **2010**, *38*, 15714–15717. [CrossRef]
40. Jones, P.G.; Thornton, P.K. Representative soil profiles for the Harmonized World Soil Database at different spatial resolutions for agricultural modelling applications. *Agric. Syst.* **2015**, *139*, 93–99. [CrossRef]
41. Avellan, T.; Zabel, F.; Mauser, W. The influence of input data quality in determining areas suitable for crop growth at the global scale—A comparative analysis of two soil and climate datasets. *Soil Use Manag.* **2012**, *28*, 249–265. [CrossRef]
42. Smiatek, G.; Helmert, J.; Gerstner, E.-M. Impact of land use and soil data specifications on COSMO-CLM simulations in the CORDEX-MED area. *Meteorol. Z.* **2016**, *25*, 215–230. [CrossRef]
43. Todd-Brown, K.E.O.; Randerson, J.T.; Post, W.M.; Hoffman, F.M.; Tarnocai, C.; Schuur, E.A.G.; Allison, S.D. Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. *Biogeosciences* **2013**, *10*, 1717–1736. [CrossRef]
44. Hoerl, A.E.; Kennard, R.W. Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics* **1970**, *12*, 69–82. [CrossRef]
45. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B (Methodol.)* **1996**, *58*, 267–288. [CrossRef]
46. Vapnik, V.N. *The Nature of Statistical Learning Theory*, 2nd ed.; Springer: New York, NY, USA, 2000; p. 314.
47. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
48. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
49. Chen, T.; Guestrin, C.; Comp, M.A. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, 13 August 2016. [CrossRef]
50. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.; Gulin, A. CatBoost: Unbiased boosting with categorical features. In Proceedings of the 32nd Conference on Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 2–8 December 2018; Volume 31.
51. Ying, L.; Shen, Z.; Yang, M.; Piao, S. Wildfire Detection Probability of MODIS Fire Products under the Constraint of Environmental Factors: A Study Based on Confirmed Ground Wildfire Records. *Remote Sens.* **2019**, *11*, 3031. [CrossRef]
52. Alizadeh, M.R.; Abatzoglou, J.; Luce, C.; Adamowski, J.; Farid, A.; Sadegh, M. Warming enabled upslope advance in western US forest fires. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2009717118. [CrossRef]
53. Hally, B.; Wallace, L.; Reinke, K.; Jones, S.; Engel, C.; Skidmore, A. Estimating Fire Background Temperature at a Geostationary Scale-An Evaluation of Contextual Methods for AHI-8. *Remote Sens.* **2018**, *10*, 1368. [CrossRef]
54. Filizzola, C.; Corrado, R.; Marchese, F.; Mazzeo, G.; Paciello, R.; Pergola, N.; Tramutoli, V. RST-FIRES, an exportable algorithm for early-fire detection and monitoring: Description, implementation, and field validation in the case of the MSG-SEVIRI sensor. *Remote Sens. Environ.* **2016**, *186*, 196–216. [CrossRef]