Research Article
Fire Prediction Based on CatBoost Algorithm

Fangrong Zhou,[1] **Hao Pan,**[1] **Zhenyu Gao,**[1] **Xuyong Huang,**[1] **Guochao Qian,**[1] **Yu Zhu,**[2] **and Feng Xiao** [ID][3]

[1]*Joint Laboratory of Power Remote Sensing Technology, Electric Power Research Institute, Yunnan Power Grid Company Ltd., Kunming 650217, China*
[2]*China Academy of Space Technology Institute of Spacecraft System Engineering, Beijing 100094, China*
[3]*School of Computer Science, Guangdong University of Technology, Guangzhou, China*

Correspondence should be addressed to Feng Xiao; 2112005051@mail2.gdut.edu.cn

In recent years, increasingly severe wildfires have posed a significant threat to the safe and stable operation of transmission lines. Wildfire risk assessment and early warning have become an important research topic in power grid risk assessment. This study proposes a fire prediction model on the basis of the CatBoost algorithm to effectively predict the fire point. Five wildfire risk factors, including vegetation factors, meteorological factors, human factors, terrain factors, and land surface temperature, were combined using the feature selection method on the basis of the gradient boosting decision tree model and principal component analysis to achieve dimensionality reduction of redundant data and create a fire prediction model. The MODIS fire point product is used as the model evaluation data. The verification result uses the AUC value as the evaluation factor. The accuracy of the model is 0.82, and the AUC value is 0.83. The obtained fire point evaluation results are in good agreement with the actual fire points. Results show that this model can effectively predict the risk of wildfires.

1. Introduction

Mountain fire disaster is an essential factor that destroys the forest ecosystem and affects the safe and stable operation of the power grid [1, 2]. Mountain fires accounted for 60% of all the emergencies that have changed the stable operation of the power grid in recent years [2]. According to statistical analysis over the years, most reclosing of transmission line trips caused by wildfire disasters will fail, which seriously affects the quality of life in the area and causes substantial economic losses to relevant departments.

Most regions in southern China are located in forest areas, with dense forests, complex terrain, and dry climate, which provide a good material basis for the occurrence of mountain fire disasters, leading to frequent mountain fire disasters and posing a considerable threat to the safe and stable operation of the power grid [3, 4]. Mountain fire disasters have become an important factor that affects the safe and stable operation of the power supply system.

Therefore, effectively predicting the fire risk of woodland, grassland, and cultivated land that may occur in the future and making corresponding warnings are considerably significant to maintain the stable operation of the power grid [5].

Scholars at home and abroad are mainly divided into two directions in the research of wildfire risk: purely using meteorological data for wildfire risk assessment and combining the tripping mechanism of transmission lines, vegetation factors, and human factors to classify wildfire risk levels. At present, meteorological departments and forestry departments mainly assess the risk of wildfires from the perspective of meteorology [6]. In 1995, Wang et al. [7] and others proposed a new technology for forest fire risk assessment based on meteorological elements such as temperature, humidity, precipitation, and wind speed, but it is only suitable for large-area forest fire risk forecasting. Literature [8] built a graph model-based overhead transmission line wildfire risk prediction model based on the

meteorological factors, combined with surface combustion factors and historical fire factors. This method has been effectively applied to a certain southern power grid. Literature [9] uses forest fire danger meteorological grades to assess the probability of wildfires and establishes a risk assessment model for transmission lines with temporal and spatial distribution characteristics. Literature [10] established a risk assessment model from two aspects: the risk of wildfire disasters and the vulnerability of transmission lines. Literature [11] combined the relationship between normalized differential vegetation index (NDVI), satellite remote sensing fire point, rainfall, and other factors with the occurrence of wildfires on transmission lines and proposed a wildfire risk assessment model for transmission lines, but only monthly risk assessment. In fact, fires are very closely related to human activities. Literature [11] proposed a fire prediction model that combines meteorological data and human activities. The model is applied in areas with severe fire disasters, and it has good prediction accuracy. Literature [12], based on historical meteorological data, vegetation, data and terrain data, used partial least squares method PLS to select the main wildfire forecasting factors and established an optimized power grid wildfire risk early warning model. Literature [13] designed a forest fire early warning model based on mobile edge computing (MEC) by acquiring ground surface parameters, which can be used to effectively predict wildfires.

In order to more fully combine meteorological data and human factors, this study is based on the MODIS fire point data of a southern province from 2015 to 2019 combined with meteorological data, terrain data, land surface temperature (LST), human factors, and vegetation factors to analyze the influencing factors of mountain fire disasters and establish a CatBoost model to predict fire points. Effective prediction and early warning of fire points are significantly important to reduce the loss of wildfire disasters.

2. Analysis and Data Acquisition of Influencing Factors of Mountain Fire Disasters

The occurrence of mountain fire disasters is comprehensively affected by a variety of factors. According to the analysis of relevant literature and the research on the principles of mountain fires [14], the occurrence of mountain fire disasters is not random, and specific laws have been passed in relation to this situation. This article divides the factors that affect the appearance of wildfires into five aspects: vegetation factors, human factors, surface temperature, terrain factors, and meteorological factors. This research aims to realize large-scale wildfire assessment through multisource remote sensing data and combined meteorological data. The specifically related factors among the five factors that affect the occurrence of wildfire disasters are as follows.

2.1. Remote Sensing Data

2.1.1. Vegetation Factors. Vegetation is the material basis for the occurrence of wildfire disasters. In this study, the influencing factors of plant on wildfire disasters are refined into normalized difference infrared index 7 (NDII7) and normalized differential vegetation index (NDVI). The NDII7 is a critical wildfire risk assessment factor. Qin [15] proved that the NDII7 can characterize the vegetation fuel moisture content and then evaluate the mountain fire risk. The NDVI is used as a criterion for judging surface vegetation and estimate the growth status and density of plant. The occurrence of mountain fire disasters is closely related to the growth status and density of vegetation. Wang et al. [14] judged the event of wildfire disasters and estimated the area of the fired area according to the change of plant NDVI at adjacent time points.

The acquisition of NDVI comes from the MOD13A1 vegetation information product of MODIS provided by the NASA website (https://ladsweb.modaps.eosdis.nasa.gov/), with a spatial resolution of 1000 m. The global NDVI information is updated every 16 days. NDII7 is derived from the MOD09A1 product provided on the website as previously mentioned. The temporal resolution of this product is 8 days. After the product is obtained, NDII7 is calculated according to the calculation formula obtained by Qin [15] and others:

$$\text{NDII7} = \frac{\rho_2 - \rho_7}{\rho_2 + \rho_7}. \tag{1}$$

2.1.2. Human Factors. The occurrence of wildfire disasters is highly correlated with the time of people's frequent activities. Statistics show that the occurrence of wildfire disasters shows a significant upward trend every Friday and every day from 13:00 to 16:00 from January to April [2]. The uncertainty of human factors is relatively considerable. This study extracts the influencing factors of wildfire disasters as land type, distance from roads, and distance from cultivated land. These data directly indicate the inevitability of human activities and can be used as the influencing factors of wildfire disasters. This notion indirectly suggests the impact of humans on fire. Land types are classified into cultivated land, forest land, grassland, water area, residential land, and unused land according to the 30 m classification data of the global surface.

2.1.3. LST. Surface temperature affects the occurrence of forest fires because it will indirectly affect the moisture content of the combustibles of vegetation. In areas with a relatively dense vegetation, the evaporation of the surface is relatively small because the surface temperature is low, thereby leading to the high moisture content of the combustibles. Mountain fire disasters are less likely to occur [16].

By contrast, if the surface temperature is high, then it is easy to cause mountain fire disasters.

The LST data come from the MOD11A1 product, with a spatial resolution of 1000 m and a temporal resolution of 1 day.

2.1.4. Terrain Factors.
Elevation, slope, and aspect are fixed static variables, and many researchers classify them as the fundamental factors leading to wildfire disasters. The ups and downs of terrain will cause different vegetation coverage and meteorological conditions, including rainfall, water content, dense vegetation, vegetation types, and growth conditions; thus, the probability of wildfire disasters will naturally vary. The spatial resolution of terrain data is 30 m. Currently, NASA website (https://ladsweb.modaps.eosdis. nasa.gov/) provides downloading of SRTM 30 m resolution digital elevation data.

2.2. Meteorological Data.
The probability of mountain fire disaster is highly correlated with meteorological factors. Meteorological factors, such as rainfall, average relative humidity, maximum temperature, average temperature, minimum temperature, maximum wind speed, and maximum wind direction [15], have a significant influence on the occurrence of wildfire disasters. The meteorological data come from the China Meteorological Data Network (http://data.cma.cn/), which is the cumulative annual value data set (2015–2019) of China.

3. Information Extraction

3.1. Fire Point Information Extraction.
The fire point data come from the fire point product of MODIS C6 (2015–2019) provided by https://firms.modaps.eosdis.nasa.gov/, and its spatial resolution is 500 m. This study extracts the fire point data according to the fire point collection time and confidence level provided by the product. Detailed MODIS C6 product information is shown in Table 1. This study will extract high-confidence fire data with a confidence of more than 90% as the input data of the fire information to improve the quality of the extracted fire information.

3.2. Nonfire Point Information Extraction.
This study first determines the distance of 35 pixels (17,500 m) from the buffer radius of the fire point through the semivariogram function [17] on the basis of the fire point data to eliminate the influence of time and then extracts it from the ring buffer (17,500–18,000 m). Thereafter, all the nonfire point data in a month are obtained. Finally, the daily fire point data corresponding to the fire point data are extracted from the corresponding monthly nonfire point data according to the daily fire point data.

4. Input Data Preprocessing

4.1. Spatial Interpolation of Meteorological Data.
The meteorological data downloaded from the China Meteorological Data Network are monitored by various meteorological stations and are spatially discrete. The meteorological data need to be spatially interpolated to achieve the continuity of the meteorological data in the study area. This study uses Anusplin software to interpolate meteorological data, which has a good effect. Qian et al. [18] compared the interpolation accuracy of Anusplin software with that of Ordinary Kriging and reverse distance weights and found that the interpolation error of the former is the smallest. The interpolation principle is mainly to use ordinary and local thin disk spline functions. The advantage of this method is primarily that it allows the introduction of multiple influence factors as covariates. This study introduces elevation data to significantly reduce the influence of elevation on temperature data changes.

4.2. Data Undersampling.
This study will use the ensemble resampling [15] algorithm for undersampling the data to ensure the consistency of the model training samples, that is, the proportion of fire-spot samples and nonfire-spot samples is the same. This algorithm can correctly solve the problem of data loss in the undersampling. Such an algorithm uses ensemble to sample with various models. Each model is undersampling. The undersampling results of multiple models are integrated, and the data distribution will not be changed. The sampling effect is better than the current numerous oversampling and undersampling techniques.

4.3. Normalization of Real Factor Data.
Among the influencing factors of mountain fire disasters, some variables are of real number type. Before the CatBoost model is trained, such input data must be normalized to ensure the dimensionlessness of the data, such as the following: distance from the road (x_1, m), distance from cultivated land (x_2, m), land surface temperature (x_3), NDVI (x_4), NDII7 (x_5), DEM (x_6, m), precipitation (x_7, mm/day), maximum temperature (x_8, °C), average relative humidity (x_9, %), average temperature (x_{10}, °C), lowest temperature (x_{11}, °C), and maximum wind speed (x_{12}, m/s). These input variables will be normalized to zero mean. The advantage of this method is that if abnormal points occur, then a small number of strange points will not have a significant effect on the average value; thus, the variance of the variance is little. Z-score normalization is also called standardization. This method maps data to a distribution with a mean of zero and a standard deviation of one. With regard to the above x_i, formula (2) is used to standardize the data, and the obtained new variable data x_i' is used as the input data of the model:

$$x_i' = \frac{x_i - \text{mean}}{\sigma}, \tag{2}$$

where $x_i \, (i = 1, 2, 3, \ldots, 12)$ is the original wildfire disaster impact factor, mean; σ is the average value and standard deviation corresponding to each element; and x_i' is the standardized wildfire disaster impact factor.

4.4. Discrete Factor One-Hot Encoding.
The discontinuous values, such as land type, slope, and aspect, have no significance. This study will perform one-hot encoding to

TABLE 1: MODIS C6 product details.

Brightness	Scan	Track	ACQ_Date	Satellite	Confidence	Bright_T31	FRP
312.4	1.1	1.1	2015/1/1	Terra	72	290.3	11.6
305.3	1.3	1.1	2015/1/1	Terra	61	289.2	9.1
304.3	1.3	1.1	2015/1/1	Terra	59	292.3	6.9
326.2	1.3	1.1	2015/1/1	Terra	84	289.8	34.1
307.8	1.3	1.1	2015/1/1	Terra	66	288.3	11.2
312	3.3	1.7	2015/1/1	Aqua	72	285.9	67.5
305	4.3	1.9	2015/1/1	Aqua	57	290.3	42.7
304.5	1	1	2015/1/2	Aqua	48	292.2	4.9
324.5	1	1	2015/1/2	Aqua	83	291.6	22.5
309.7	1	1	2015/1/2	Aqua	35	293.6	5.8
346.2	1	1	2015/1/2	Aqua	94	294	55.8
305.6	1	1	2015/1/2	Aqua	62	287.8	6.7
315.4	1	1	2015/1/2	Aqua	73	293.2	12.9
301.1	1	1	2015/1/2	Aqua	44	289.6	3.9
...	...	...	...	...	...	...	...
300.3	1	1	2019/12/30	Terra	30	286.9	4.3
304.7	1.1	1	2019/12/30	Terra	60	292.1	5.6
308.1	1	1	2019/12/31	Aqua	61	295.5	5
302.9	1	1	2019/12/31	Aqua	24	290.4	3
300.1	1	1	2019/12/31	Aqua	13	285.9	3.8
302.4	1.2	1.1	2019/12/31	Aqua	49	290	5

eliminate the influence between the numerical values. The significant advantage of this method is that it is easy to deal with noncontinuous values, and the model input data are also expanded to a certain extent.

4.5. Feature Selection Method Based on the Gradient Boosting Decision Tree (GBDT) Model.

Features must be selected because of the large number of variables in this study, and some variables have little effect on the occurrence of wildfires. Feature selection is the process of choosing factors that are highly correlated with the appearance of fires. The feature selection method based on the GBDT model is a commonly used feature selection method based on the tree model. The principle is to use the node magazines in each decision tree to calculate the importance of features. The final feature importance is the average of the feature importance of all decision trees. In this study, the cross-validation method is used to select the factors whose feature importance is more significant than 0.3. Then, the dimensionality reduction is performed according to the principal component analysis (PCA). The ranking of the importance index of wildfire impact factors is shown in Table 2.

4.6. PCA: Principal Component Analysis.

Among the influencing factors of mountain fire disaster, a specific correlation exists between elevation, slope, aspect, maximum temperature, average temperature, minimum temperature, and surface temperature. This study uses the currently widely used linear dimensionality reduction algorithm (PCA) to reduce the dimensionality of all influencing factors of wildfire disasters and eliminate redundant data. The advantage of this algorithm is its ability to retain the original data quality of the sample. In this mechanism, the model training data are compressed as much as possible, and the

factors with high principal components for model training are determined.

The mathematical model of the PCA algorithm in this study is as follows.

$X = x_1, x_2, \ldots, x_m$ is the impact factor of wildfire disaster, where the dimension of X is m, which is the number of impact factors. The projection of x_i on the hyperplane in the new hyperdimensional space is $W^T x_i$. The principle is to increase the variance between all sample points to ensure that the projections between all sample data are separated as much as possible. XX^T can be obtained according to the following formula:

$$XX^TW = \lambda W. \tag{3}$$

After the sample feature matrix XX^T is decomposed, the eigenvalues of each factor $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_m$ are obtained, and the corresponding eigenvectors of the first I samples $W = (W_1, W_2, \ldots W_i)$ are the required mountain fires of the principal components of the disaster impact factor. This paper retains 99% of the main information of the original feature. The latitude of the principal component m is 18. Compared with the feature selection based on the GBDT model, the feature dimension is reduced by 13.

5. Fire Point Prediction Model Based on CatBoost Algorithm

5.1. CatBoost Model.

CatBoost is an algorithm that combines GBDT and categorical features. This approach is an improved implementation under the framework of the GBDT algorithm. CatBoost is based on oblivious trees with few parameters and supports categorical variables and high accuracy sexual GBDT framework. The main pain point is to efficiently and rationally deal with categorical features. CatBoost is composed of categorical variables and boost.

TABLE 2: Importance index of wildfire impact factors.

Land surface temperature	37.61
Distance from the road	7.31
Maximum wind speed	6.51
Slope	5.15
Aspect (NE)	0.63
Aspect (SW)	0.58
Aspect (N)	0.54
Precipitation	0.37
Wind direction (SW)	0.29
Aspect (E)	0.16
Wind direction (SE)	0.07
Wind direction (NNE)	0.00
Wind direction (ENE)	0.00
Normalized infrared index 7	9.83
Maximum temperature	6.76
Normalized vegetation index	6.24
Wind direction (WSW)	0.67
Aspect (N)	0.62
Aspect (NW)	0.55
Grass	0.44
Aspect (W)	0.35
Wind direction (SSE)	0.28
Wind direction (W)	0.16
Wind direction (E)	0.06
Wind direction (NNW)	0.00
Wind direction (N)	0.00
Elevation	9.69
Distance from cultivated land	6.63
Average relative humidity	5.84
Wind direction (S)	0.65
Aspect (SE)	0.62
Woodland	0.55
Cultivable land	0.42
Wind direction (WNW)	0.32
Wind direction (SSW)	0.27
Wind direction (ESE)	0.12
Wind direction (NW)	0.04
Wind direction (NE)	0.00
Aspect (flag)	0.00

This mechanism also deals with gradient bias and prediction shift problems, thereby improving the generalization ability and robustness of the algorithm [19, 20]. This study considers many categorical features, such as rainfall, wind direction, slope direction, and land type. CatBoost can be used to quickly process nonnumerical features. When the CatBoost algorithm processes categorical features, it puts all sample data sets into the algorithm for learning. Then, CatBoost randomly arranges all these sample data sets and filters out samples with the same category from all features. When numerically transforming the characteristics of each sample, the target value of the sample is first calculated before the sample, and the corresponding weight and priority are added [21, 22]. The specific formula is shown in the following:

$$x_k^i = \frac{\sum_{j=1}^n \{x_j^i = x_k^i\} \cdot y_i + ap}{\sum_{j=1}^n \{x_j^k = x_k^i\} + a}, \qquad (4)$$

where p represents the added prior value and the weight coefficient greater than zero. An a priori value is added to significantly reduce the noise points caused by low-frequency features to effectively minimize the overfitting of the model and improve the generalization ability.

5.2. Fire Point Model Training and Optimization

5.2.1. Model Training.
The five-year MODIS monitoring fire point data of a southern province from 2015 to 2019 and the nonfire point data extracted by the method described in this study are selected as the sample set. The fire-spot data with a confidence level of less than 90% is eliminated to improve the quality of the fire-spot samples. The sample data after data oversampling, normalization, one-hot encoding, feature selection, and PCA dimensionality reduction are substituted into the CatBoost model for training. Approximately 70% of the data are randomly selected for model training and 30% for model testing. The temporal resolution of NDII7, NDVI, and land surface temperature in the input feature variables of the model are 8 days, 16 days, and 1 day, respectively. The input data of NDII7, NDVI, and land surface temperature select the data of the previous time phase before the fire to prevent the input vegetation data and land surface temperature from being affected by the fire and failing to achieve the effect of fire prediction. The data of human factors and terrain factors are unchanged, while the input time phase of weather data is consistent with the fire data. Figure 1 is the time phase relationship diagram of the input feature variables of the CatBoost model, and Figure 2 is the basic flow chart of model training.

5.2.2. Model Optimization.
This study uses grid search combined with tenfold cross-validation to optimize the primary hyperparameters of the CatBoost model, including iterations, learning_rate, max_depth, criterion, and feature importance, to improve the accuracy of model fire prediction. Tenfold cross-validation divides the sample data into ten mutually exclusive training subsets. Each time nine subsets are selected as training data, and the remaining subset is used as test data. The multiple rounds of training are repeated to ensure that each subset is as the test set, the ten test results are obtained, and the average of the ten test results is the accuracy of the model. The hyperparameters obtained through a grid search can effectively improve the prediction effect of the model [23].

After model optimization, the best hyperparameters of the fire point prediction model are shown in Table 3.

5.2.3. Model Evaluation.
This study uses accuracy, precision, recall, F1-score, and AUC value to make a comprehensive evaluation of the model prediction accuracy and address the classification problem of unbalanced data of fire point prediction. The confusion matrix of the fire point and nonfire point data sets in this article is shown in Table 4.

The evaluation index of the fire point prediction model can be obtained according to the confusion matrix.
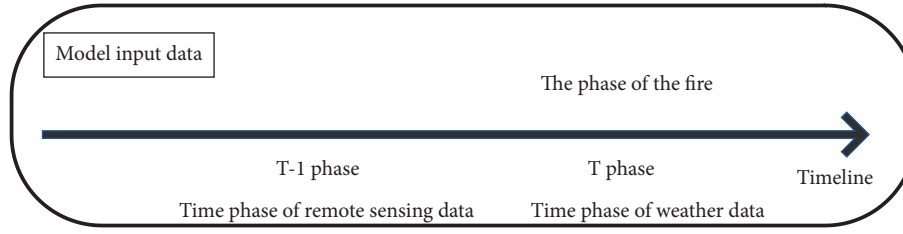
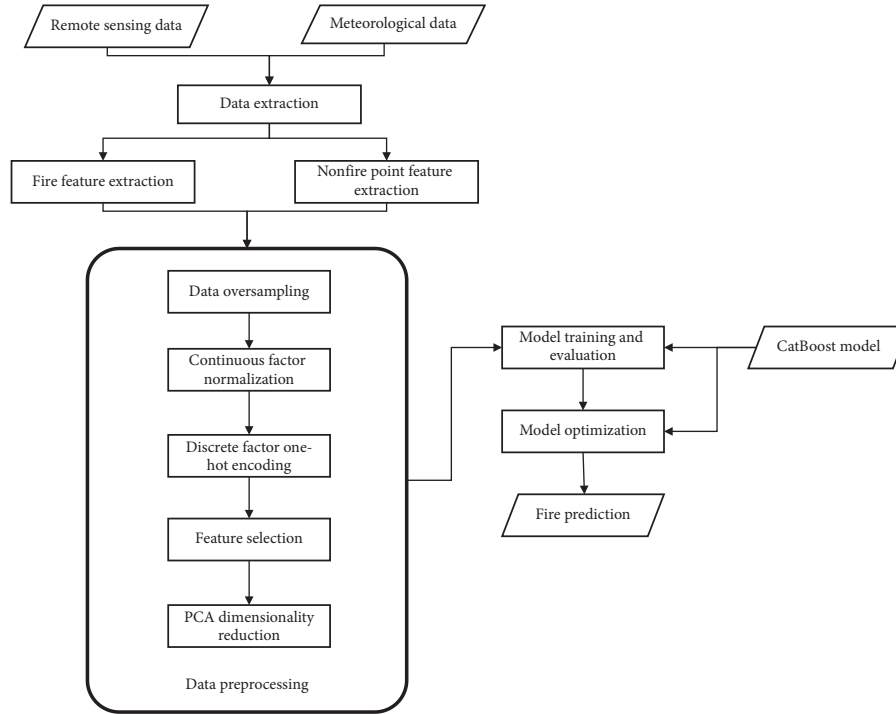Figure 1: Time-phase relationship diagram of the model input characteristic variables.

Figure 2: Flow chart of the fire prediction model.

Table 3: Best hyperparameters of the fire prediction model.

Iterations	Learning_rate	Max_depth	Criterion	Feature importance
1000	0.05	9	Gini	0.3

Table 4: Confusion matrix of fire and nonfire points.

	Predicted fire point	Predicted nonfire point
Real fire point	TN (true negative)	FP (false positive)
Real nonfire point	FN (false negative)	TP (true positive)

TN: the actual value is the fire point, and it is also predicted as the fire point. FN: the actual value is a nonfire point, but it is predicted to be a fire point. FP: the actual value is a fire point, but it is predicted to be a nonfire point. TP: the actual value is a nonfire point, and it is also predicted as a nonfire point.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FN + FP + TN)}. \tag{5}$$

$$\text{Recall} = \frac{TP}{(TP + FN)}. \tag{7}$$

$$\text{Precision} = \frac{TP}{(TP + FP)}. \tag{6}$$

$$\text{F1-score} = 2\frac{(\text{precision} * \text{recall})}{(\text{precision} + \text{recall})}. \tag{8}$$

TABLE 5: CatBoost fire point prediction model results.

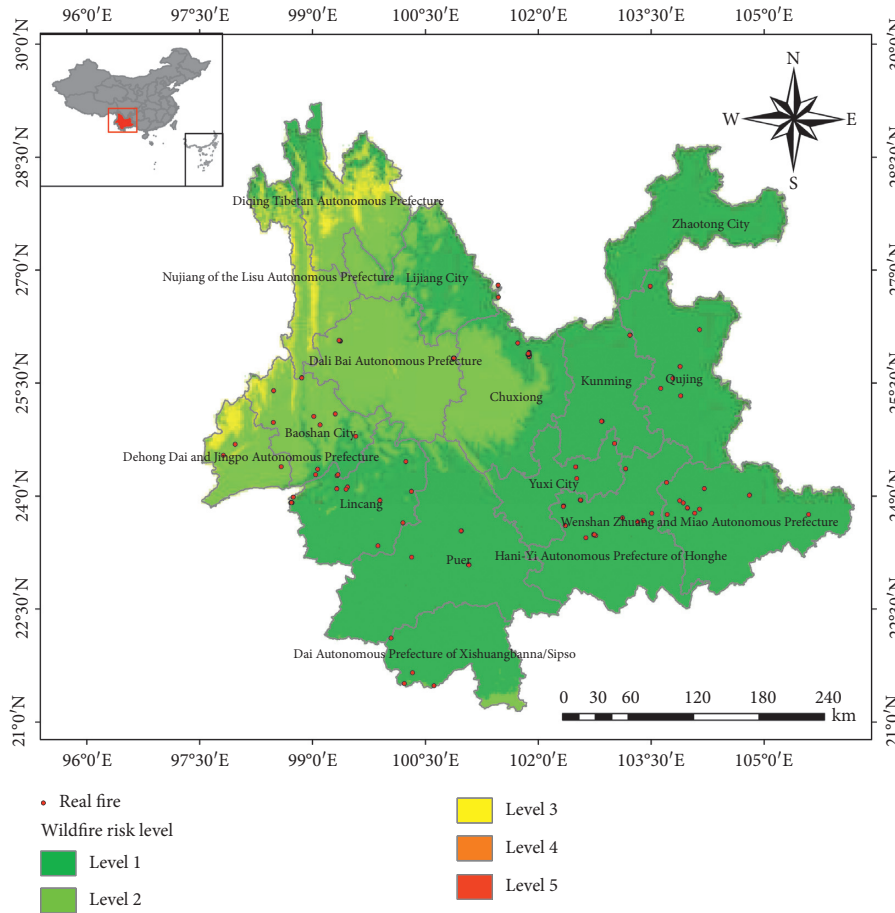|	Precision	Recall	F1-score	Support	AUC
Nonfire point	0.83	0.87	0.78	889	
Fire point	0.81	0.82	0.83	273	
Accuracy			0.79	1162	
Overall	0.82	0.84	0.80	1162	0.83

FIGURE 3: The risk forecast of wildfires in Yunnan Province on March 15, 2020.

AUC value: the AUC value is the area value under the ROC curve, which can quantitatively reflect the model performance measured on the basis of the ROC curve. The abscissa of the ROC curve is the false positive rate, $FPR = FP/(FP + TN)$, and the ordinate is the true positive rate, $TPR = TP/(TP + FN)$.

This study uses the best hyperparameters obtained from the model optimization in Section 4.2 to predict the fire point of the sample data set. The final five model evaluation indicators are shown in Table 5.

The results shown in Table 5 demonstrate that the CatBoost fire point prediction model after model optimization has a nonfire point precision of 0.83, recall of 0.87, and F1-score of 0.78 and a fire point precision of 0.81, recall of 0.82, and F1-

score of 0.83. The final accuracy is 0.79, the overall precision is 0.82, recall is 0.84, the F1-score is 0.80, and the AUC value is 0.83. The fire prediction results indicated that the model's prediction of the fire starts with a good predictive effect, and the risk of wildfires can be effectively predicted.

In order to more intuitively reflect the effect of the model in predicting the risk of wildfires, this article draws the comparison between the wildfire risk prediction maps and real fire spots in Yunnan Province on March 15, 2020, April 15, 2020, and May 15, 2020. The resolution of the wildfire risk prediction map is 500 meters, as shown in Figures 3–5. It can be seen that more than 80% of the real fire points fall in the high-risk area of the prediction map, which further verifies the model's effectiveness.
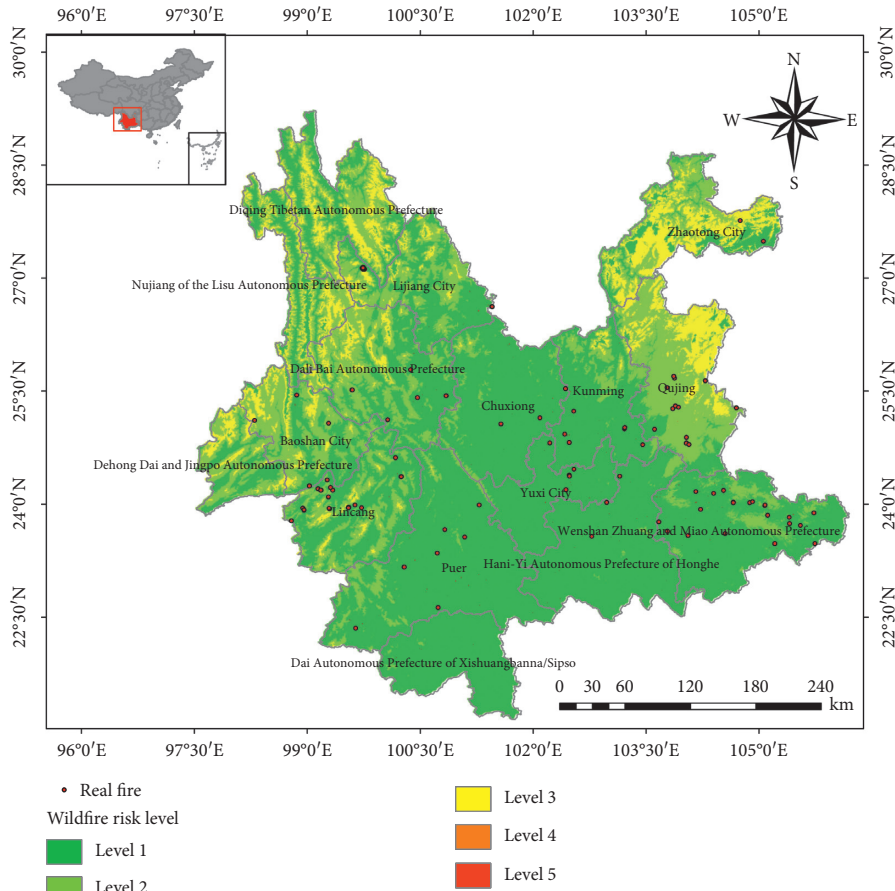
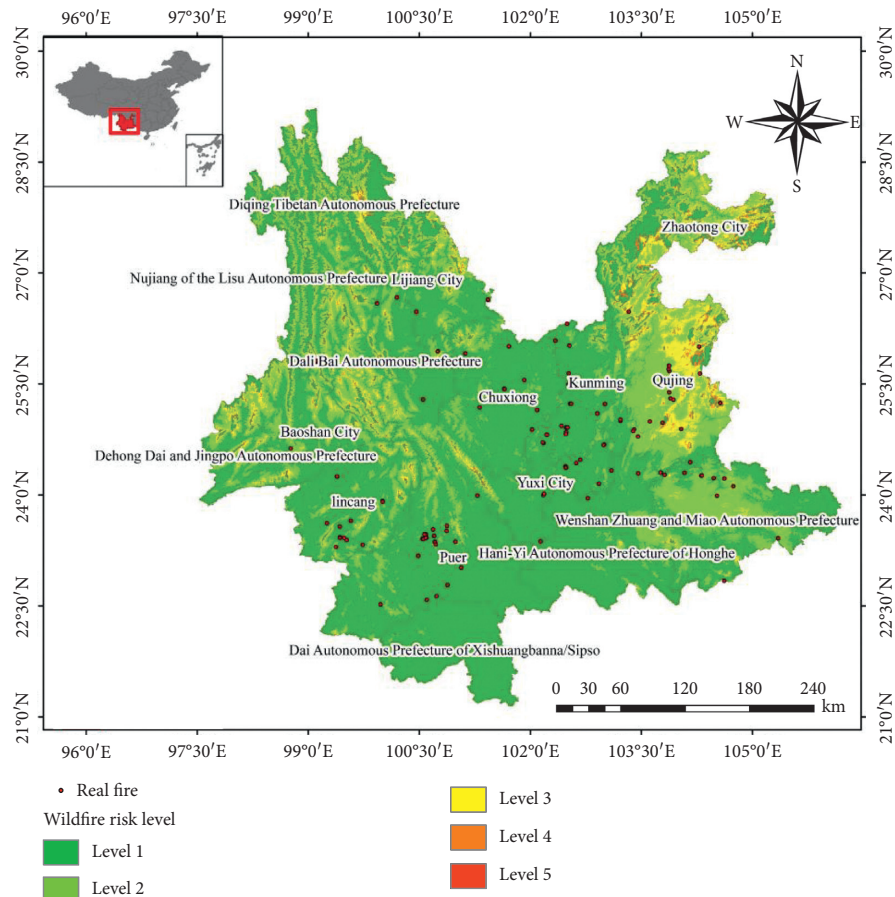Figure 4: The risk forecast of wildfires in Yunnan Province on April 15, 2020.

Figure 5: The risk forecast of wildfires in Yunnan Province on May 15, 2020.

6. Conclusions and Prospects

This study uses MODIS fire data, combined with vegetation factors, human factors, meteorological factors, surface temperature, and terrain factors, based on feature selection and PCA dimension reduction to find out the influencing factors that are highly correlated with the occurrence of wildfires. The research proposes a method based on Cat-Boost algorithmic fire prediction model. This model can effectively predict fire points, is helpful in preventing wildfire risks, and has a specific guiding role for the electric power department to avoid risks of fire and make appropriate early warning arrangements in advance.

Although this article has achieved positive research results, it still has some deficiencies and areas worthy of in-depth study. The research conducted in this study is only based on the first-level classification of land types to make fire forecasts and does not make precise fire forecasts under a single ground type. Under the secondary classification of land types, the establishment of different fire prediction models is based on each specific feature to achieve more precise and accurate fire prediction in the direction of further in-depth research.

Data Availability

This article contains data to support the results of this research. Some data cannot be provided because it involves the coordinate data of power grid poles.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] L. Shun-gen, "Forest-fire rule, cause and preventive measure," *China New Technologies and Products*, vol. 3, no. 252, pp. 354-355, 2011.

[2] X. Chen, R. N. Ling, J. Chen et al., "A temporal and spatial clustering analysis and its application with wildfire disasters in hubei power grid based on satellite hot spot data," *Bulletin of Surveying and Mapping*, vol. 268, no. 9, pp. 14–18, 2015.

[3] G. Lei, W. He, and J. Lin, "Realization and application of integrated monitoring system for anti-landscaping in overhead transmission line," *Electrical Engineering*, vol. 12, no. 12, pp. 112–115, 2013.

[4] H. Xu, W. Liao, K. Li, and T. Guo, "Measures for preventing mountain fire in transmission line," *Electric Safety Technology*, vol. 5, no. 5, pp. 51–54, 2013.

[5] M. Liu, Z. Shao, S. Guan et al., "Study on models and methods of risk assessment of transmission line fault caused by fire," *Power System Protection and Control*, vol. 43, no. 15, pp. 28–35, 2015.

[6] X. Qin, Z. Zhang, and Z. Li, "Forecasting methodology of national-level forest fire risk rating," *Remote Sensing Technology and Application*, vol. 486, no. 5, pp. 500–504, 2008.

[7] Y. Wang, H. Cheng, and W. Duan, "A new forecast method of forest fire risk rating based on remote sensing data of land surface temperature and humidity," *Journal of West China Forestry Science*, vol. 43, no. 6, pp. 97–103, 2014.

[8] E. Zhou, Y. Huang, J. Chen, X. Tian, R. Wei, and Y. Zhou, "Graph theory based fire risk prediction level model of overhead transmission lines," *Southern Power System Technology*, vol. 4, no. 4, 2020.

[9] X. Xiong, Y. Zeng, J. Wang, and H. Li, "Risk assessment of power transmission channels in forest regions based on spatial-temporal features of forest fire," *Power System Protection and Control*, vol. 46, no. 4, pp. 1–9, 2008.

[10] Z. Feng, Y. Lu, Y. Sun, Q. Xu, and P. Jiang, "Assessment of power grid risk caused by wildfire disaster in Yunnan Province," *Journal of Natural Disasters*, vol. 23, no. 5, 2014.

[11] J. Lu, J. Guo, Li Yang, F. Tao, and J. Zhang, "Research and application of fire forecasting model for electric transmission lines incorporating meteorological data and human activities," *Mathematical Problems in Engineering*, vol. 2016, Article ID 9828676, 10 pages, 2016.

[12] E. Zhou, S. Hu, L. Zhang, R. Wei, H. Wang, and Y. Fan, "Characteristics and risk warning technology of wildfire disaster in power grid," *Electric Power Engineering Technology*, vol. 39, no. 3, pp. 2096–3203, 2020.

[13] C. Cheng, H. Zhou, X. Chai et al., "Adoption of image surface parameters under moving edge computing in the construction of mountain fire warning method," *PLoS One*, vol. 15, no. 5, p. e0232433, 2020.

[14] H. Wang, J. Luo, T. Xu et al., "Investigation and analysis of the status quo of natural disaster prevention technology in China power grid," *Automation of Electric Power Systems*, vol. 34, no. 23, pp. 5–10, 2010.

[15] X. Qin, "Study on forest fire early warning and monitoring methodology using remote sensing and geography information system techniques," PhD thesis, Resource Information Institute of Chinese Academy of Forestry, Chinese Academy of Forestry, Beijing China, 2005.

[16] H. K. Preisler, R. E. Burgan, J. C. Eidenshink, J. M. Klaver, and R. W. Klaver, "Forecasting distributions of large federal-lands fires utilizing satellite and gridded weather information," *International Journal of Wildland Fire*, vol. 18, no. 5, pp. 508–516, 2009.

[17] R. Fraser, Z. Li, and J. Cihlar, "Hotspot and NDVI differencing synergy (HANDS) A new technique for burned area mapping over boreal forest," *Remote Sensing of Environment*, vol. 74, no. 3, pp. 362–376, 2000.

[18] Y. Qian, H. Lv, Y. Zhang et al., "Application and assessment of spatial interpolation method on daily meteorological elements based on ANUSPLIN software," *Journal of Meteorology and Environment*, vol. 26, no. 2, 2010.

[19] B.-h. Huang, H. Zhang, Z.-j. Sun et al., "Forest fire danger factors and their division in Shandong based on GIS and RS," *Chinese Journal of Ecology*, vol. 34, no. 5, pp. 1464–1472, 2015.

[20] A. L. F. Chan, "An application of classification analysis for skewed class distribution in therapeutic drug monitoring-he case of vancomycin," in Proceedings of the 2004 IDEAS Workshop on Medical Information Systems: The Digital Hospital (IDEAS-H'04), Beijing, China, September 2004.

[21] Dorogush A. V., Ershov V., Gulin A.. CatBoost: Gradient Boosting with Categorical Features Support [EB/OL]. (2018-11-24). [2019-03-19], https://arxiv.org/abs/1810.11363.

[22] Prokhorenkova L., Gusev G., Vorobev A., et al.. CatBoost: Unbiased Boosting with Categorical Features [EB/OL]. (2019-01-02). [2019-03-22], https://arxiv.org/abs/1706.09516.

[23] J. Wang, L. Zhang, G. Chen et al., "A parameter optimization method for an SVM based on improved grid search algorithm," *Applied Science and Technology*, vol. 391, no. 3, pp. 28–31, 2012.