

Research Article

Application of the Artificial Neural Network and Support Vector Machines in Forest Fire Prediction in the Guangxi Autonomous Region, China

Yudong Li,¹ Zhongke Feng ,¹ Shilin Chen,¹ Ziyu Zhao,¹ and Fengge Wang²

¹Precision Forestry Key Laboratory of Beijing, Beijing Forestry University, Beijing 100083, China

²Monitoring Center in Ministry of Emergency Management of China, Beijing 100054, China

Correspondence should be addressed to Zhongke Feng; zhongkefeng6205@126.com

Received 3 March 2020; Accepted 18 March 2020; Published 23 April 2020

Academic Editor: Maria Alessandra Ragusa

Copyright © 2020 Yudong Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The study of forest fire prediction is of great environmental and scientific significance. China's Guangxi Autonomous Region has a high incidence rate of forest fires. At present, there is little research on forest fires in this area. The application of the artificial neural network and support vector machines (SVM) in forest fire prediction in this area can provide data for forest fire prevention and control in Guangxi. In this paper, based on Guangxi's 2010–2018 satellite monitoring hotspot data, meteorology, terrain, vegetation, infrastructure, and socioeconomic data, the researchers determined the main forest fire driving factors in Guangxi. They used feature selection and backpropagation neural networks and radial basis SVM to build forest fire prediction models. Finally, the researchers use the accuracy, precision, and area under the characteristic curve (ROC-AUC) and other indicators to evaluate the predictive performance of the two models. The results showed that the prediction accuracy of the BP neural network and SVM is 92.16% and 89.89%, respectively. As both results are over 85%, the requirements of prediction accuracy is met. These results can be used for forest fire prediction in the Guangxi Autonomous Region. Specifically, the accuracy of the BP neural network was 0.93, which was higher than that of the SVM model (0.89); the recall of the SVM model was 0.84, which was lower than the BANN model (0.92), and the AUC value of the SVM model was 0.95, which was lower than the BP neural network model. The obtained results confirm that the BP neural network model can provide more prediction accuracy than support vector machines and is therefore more suitable for forest fire prediction in Guangxi, China. This research provides the necessary theoretical basis and data support for application in the field of forestry of the Guangxi Autonomous Region, China.

1. Introduction

Forest fires are one of the important disturbance factors in the global forest ecosystem [1], and it causes various degrees of negative impacts on the ecological environment, resources, human health, economy, and so on [2–8]. On March 30, 2019, there was a forest fire in Muli County, Liangshan Prefecture, Sichuan Province, China. This fire killed 31 people, and the total fire area was about 20 hectares [9]. Since July 2019, high temperatures and drought have caused forest fires in many places in Australia, killing 28 people, and over 10 million hectares of fire damage have been recorded [10]. The severe consequences of frequent forest fires mean that problems in forest fire management must be urgently resolved. Establishing

a forest fire prediction model is of great significance for reducing economic losses and casualties. Therefore, the determination of forest fire driving factors is of high significance, as well as the establishment of a high-precision forest fire prediction model.

A large amount of research has been done on forest fire prediction models. The model commonly used by most scholars is the logistic regression model [11–13]. Liu et al. used meteorological factors to predict the number of forest fires through exponential equations [14]. In the past, scholars have also used zero expansion models [15–17] to develop fire prediction models, for example, Guo FT (2010) used ordinary least square regression (OLS), Poisson, and zero-inflated Poisson (ZIP) models to predict the number of forest fires in

the Greater Xing'an Mountains, Heilongjiang Province, China. At the same time, a negative binomial (NB) regression model and a zero-inflated negative binomial (ZINB) model were used to simulate the relationship between forest fires and meteorological factors. Argañaraz et al. (2015) used weighted regression tree models (also known as boosted regression trees) to determine the occurrence of fires.

With the continuous development of artificial intelligence technology, the use of machine learning algorithms to build forest fire prediction models has drawn increasing attention from the scientific community [18–26]. Artificial neural networks are a highly nonlinear dynamic system, which can approximate and simulate any nonlinear function of nonlinear dynamic phenomena such as forest fires with strong fault tolerances [27]. Compared to the traditional multiple linear regression models or parametric regression models, artificial neural networks have stronger nonlinear mapping capabilities and can also benefit from self-learning and adaptive capabilities. Many scholars have done relevant research on specific areas, for example, Alonso et al. [28] introduced an intelligent system for forest fire risk prediction based on a multilayer perceptron network. Sakr et al. [29] used two meteorological parameters, relative humidity and accumulated precipitation, to predict whether forest fires in developing countries would occur using both SVM and ANN. Hu [30] took the forest fire situation in Guilin and Guangzhou as examples and established a BP neural network prediction model with high accuracy. Wang et al. [31] used particle swarm optimization to optimize the BP neural network to fit and predict the forest farms of Nanjing Forestry University. Support vector machines are widely used in businesses and academia as a method of data mining and knowledge exploration [29, 32–35]. In comparison with existing statistical methods, SVM can quickly and efficiently implement “transduction inference” from both training and prediction samples using a simple and robust algorithm. The support vector machine method is rarely used in forest fire prediction. Scholars engaged in related research mainly include Cortez and Morais [36] and Al-Janabi et al. [20]. Paulo Cortez used five different DM algorithms to predict the burning areas in the northeast of Portugal and showed the significance of SVM prediction in this field of research. Xu [19] used SVM and semipositive definite programming modeling to select the optimal kernel function of the support vector machine to establish an SVM model for forest fire prediction. Al-Janabi [20] used five different SC technologies such as the SVM algorithm to predict areas of fire interactions and finally determined that the SVM algorithms can provide accurate predictions.

China's Guangxi Autonomous Region is an area known for a high incidence rate of forest fires. In the past, He and Lu [37, 38] used empirical methods or established linear regression models [39] to study the relationship between forest fire occurrence rules and driving factors in Guangxi. However, due to the complex nonlinear relationship between the occurrence of forest fires and the influencing factors, no one has yet established a suitable and accurate prediction model for the area. Therefore, based on factors such as weather and topography, this study uses Matlab and

other software to establish both a neural network and a support vector machine model for forest fires in the Guangxi Zhuang Autonomous Region. This study will also calculate the prediction accuracy of both these respective models. Through the comparative analysis of model fitting results, this study will judge the adaptability of the two algorithms in forest fire prediction in the Guangxi forest area.

2. Materials and Methods

2.1. Study Area. Guangxi Zhuang Autonomous Region is located in South China, with a geographic location between $21^{\circ} 54' - 26^{\circ} 24' N$ and $104^{\circ} 28' - 112^{\circ} 04' E$. The land area is 237,600 square kilometers, and the sea area is about 40,000 square kilometers. Guangxi belongs to the subtropical monsoon climate zone, with a warm climate, abundant rain, and sufficient light. Summer sunshine time is long, the temperature is high, and precipitation is high; winter sunshine time is short, and the weather is dry and warm. This place has an annual average temperature of $17.5 - 23.5^{\circ} C$, average annual rainfall 841.2–3387.5 mm, and annual sunshine 1213.0–2135.2 hours. Guangxi is generally a mountainous and hilly basin landform, with a basin-like shape as a whole, and its terrain rises from southeast to northwest [23]. Guangxi is rich in forest resources. It is one of the important forest areas in the south of China. Its forest land area is 15.2717 million hectares. Forest volume reaches 602 million cubic meters in Guangxi, ranking seventh in the country. The dominant tree species in Guangxi are *Taxodium*, eucalyptus, and pine. Guangxi is a high-incidence area of forest fires. According to incomplete statistics, there were 19379 forest fires in Guangxi from 2001 to 2018, with an average of 1077 forest fires per year. It is an area with severe forest fire hazards in China. Figure 1 shows survey of the research area.

2.2. Data Resources. The data in this article include six parts, which are fire data, meteorological data, terrain data, vegetation data, infrastructure data, and socioeconomic data. The fire point data come from the Fire Prevention and Management Department of the Ministry of Emergency Management of China, which includes hotspot data (longitude, latitude, and image time) of satellite monitoring in the Guangxi Autonomous Region from 2010 to 2018. The meteorological data are derived from the China Meteorological Data Network for a total of eight years of daily value data sets, which include eight factors such as the pressure, temperature, relative humidity, and precipitation of this station. DEM data are obtained through the geospatial data cloud website; the vegetation factor is represented by NDVI (Vegetation Normalization Index), and the monthly vegetation index comes from the resource environment data cloud platform. The basic geographic data come from the National Basic Geographic Database of 1:250,000 on the National Geographic Information Resource Catalog System website. This data include points, lines, and areas of railways, highways, water systems, and residential areas. Socioeconomic data include population density and GDP per capita. It uses population and environmental data cloud platforms

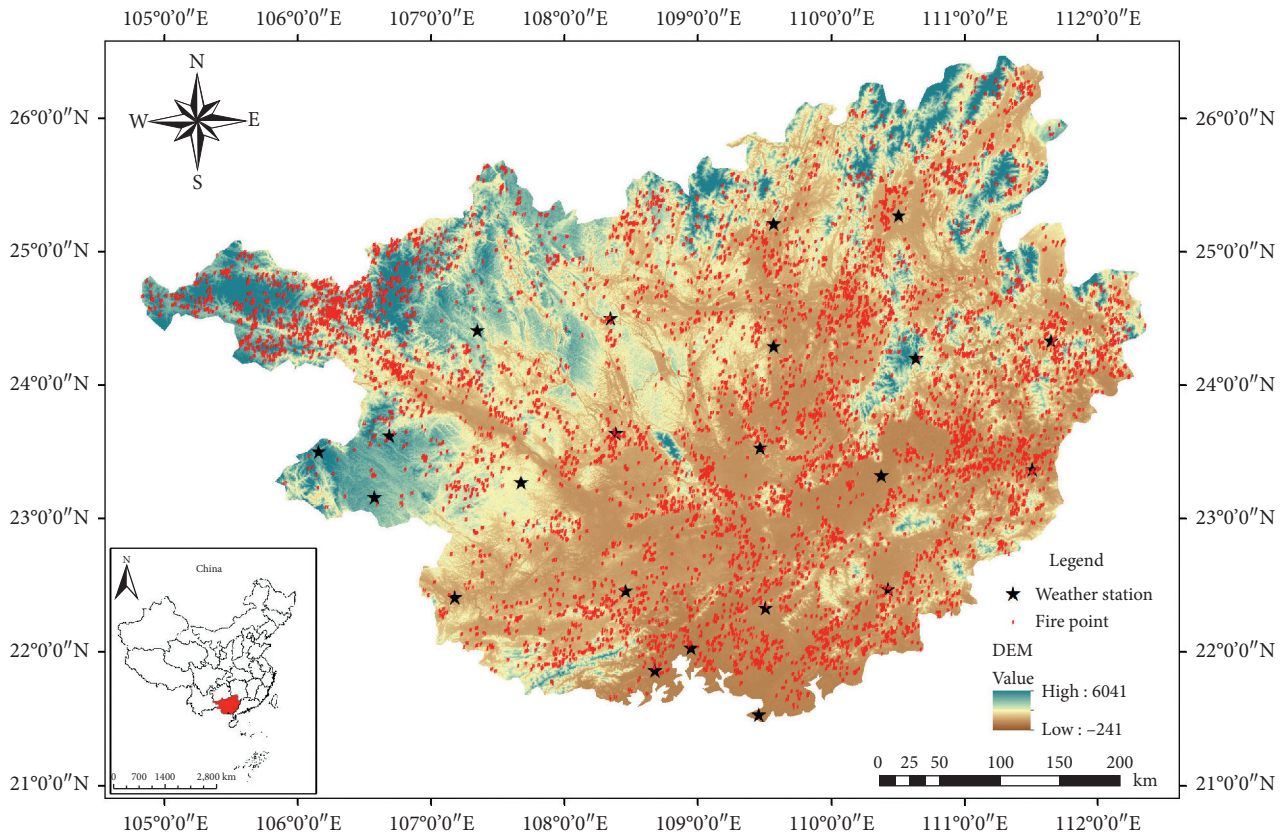


FIGURE 1: Survey of the study area.

to obtain population and GDP spatial distribution kilometer grid data.

2.3. Data Processing

2.3.1. Variable Processing. The dependent variable in the study is a binary variable, so we need to use ArcGIS to create a certain percentage of random points (i.e., nonfire points) and assign the fire point to 1 and the nonfire point to 0 [40]. In order to ensure that the data are not excessively scattered, random points are selected at a 1 : 1 ratio according to experience [41], and in principle, double randomness in space and time should be followed [42]. In this study, slope, aspect, and special holidays were set as categorical variables, and the rest were continuous variables. The classification of slope and aspect is shown in Table 1 and 2, respectively. Since Chinese traditional festivals burn paper to commemorate lost loved ones, Chinese New Year’s Eve, the first day of the first lunar month to the fourth day of the lunar month, the fifteenth day of the first lunar month, the Qingming Festival, and the Chinese Yuan Festival (July 15th) are designated as special holidays expressed as 1; nonspecial holidays are represented as 0. The random points in this study were created by ArcGIS software and based on the 2015 national land use data. After removing random points that fell on waters, urban land, and the ocean, the number of fire points and random points was 13,178. After processing the data, a total of 26 independent variables and their states are finally obtained, as shown in Table 3.

TABLE 1: Slope classification.

Slope range (degrees)	Classification description
0~5	0
5~15	1
15~25	2
25~35	3

TABLE 2: Aspect classification.

Aspect	Aspect range (degrees)	Classification description
Gentle slope	-1	1
North	0~22.5/337.5~360	2
Northeast	22.5~67.5	3
East	67.5~112.5	4
Southeast	112.5~157.5	5
South	157.5~202.5	6
Southwest	202.5~247.5	7
West	247.5~292.5	8
Northwest	292.5~337.5	9

The data in this article contain a total of 26,355 sample points from 2010 to 2018. First of all, we need to clean these data, that is, to remove abnormal samples in the original data set (including some samples with missing data and samples with observations that obviously exceeded the normal range) and add target category labels. The 26 independent variables were then preprocessed.

TABLE 3: Overview of independent variables.

Influencing factors	Independent variable	Symbol	Variable type
Location	Longitude (°)	<i>Lon</i>	Continuous variable
	Latitude (°)	<i>Lat</i>	Continuous variable
	Altitude (m)	<i>Alt</i>	Continuous variable
	Slope (°)	<i>Slo</i>	Categorical variable
	Aspect	<i>Asp</i>	Categorical variable
Meteorology	Average surface temperature (°C)	<i>Avst</i>	Continuous variable
	Daily maximum surface temperature (°C)	<i>Mast</i>	Continuous variable
	Daily minimum surface temperature (°C)	<i>Mist</i>	Continuous variable
	Cumulative precipitation at 20-20 (mm)	<i>Pre</i>	Continuous variable
	Average station pressure (hPa)	<i>Spr</i>	Continuous variable
	Average relative humidity (%)	<i>Arh</i>	Continuous variable
	Minimum relative humidity (%)	<i>Mrh</i>	Continuous variable
	Hours of sunshine (h)	<i>Suh</i>	Continuous variable
	Average temperature (°C)	<i>Ate</i>	Continuous variable
	Daily maximum temperature (°C)	<i>Mate</i>	Continuous variable
	Daily minimum temperature (°C)	<i>Mite</i>	Continuous variable
	Average wind speed (m/s)	<i>Aws</i>	Continuous variable
	Maximum wind speed (m/s)	<i>Mws</i>	Continuous variable
Infrastructure	Distance between fire point and railway (m)	<i>Ral</i>	Continuous variable
	The shortest distance from the fire point to the water system (m)	<i>Wat</i>	Continuous variable
	Distance from fire point to highway (m)	<i>Hig</i>	Continuous variable
	Closest distance from fire point to residential area (m)	<i>Set</i>	Continuous variable
Social humanity	Population	<i>Pop</i>	Continuous variable
	GDP	<i>GDP</i>	Continuous variable
Vegetation	Special festival	<i>Sfe</i>	Categorical variable
	NDVI	<i>NDVI</i>	Continuous variable

2.3.2. *Normalization.* The dimensions of each forest fire driving factor are different, and the data levels are different. In order to eliminate the dimension, large differences in the magnitude of the input and output data are avoided and the contributions of various factors are balanced; we normalize the data and convert all the data into numbers between [0,1]. The function form is as follows:

$$x_i^* = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \quad (1)$$

where x_i and x_i^* are the values before and after data normalization and x_{\max} and x_{\min} are the maximum and minimum values of the full sample data, respectively [43].

2.4. Research Method

2.4.1. *Relief Algorithm.* The Relief algorithm is an efficient feature weighting algorithm that simplifies feature sets and is

often applied to binary classification problems. The main idea of the algorithm is to weight the features according to the correlation of each feature to the category, and the weight is calculated from the ability of the feature to distinguish the two types of samples in the neighborhood. Then, according to the weight threshold that has been set, the features with weights less than the threshold will be removed, and we will finally get the optimal feature subset [44]. The content of the Relief algorithm is as follows: we randomly select a sample R from the training set, taking k nearest neighbor samples from the same sample set of R and then finding k nearest neighbor samples from other different types of sample sets, and finally the weight of the feature is updated according to equation (2) which is shown as follows [45]:

$$W(A) = W(A) - \sum_{j=1}^k \frac{\text{diff}(A, R, H_j)}{mk} + \sum_{C \notin \text{class}(R)} \frac{p(C)/1 - p(\text{class}(R)) \sum_{j=1}^k \text{diff}(A, R, M_j(C))}{mk}, \quad (2)$$

where $p(C)$ is the class ratio, $p(\text{class}(R))$ is the class ratio of a sample selected randomly, $\text{diff}(A, R_1, R_2)$ represents the distance between samples R_1 and R_2 on feature A , and

$$\text{diff}(A, R_1, R_2) \begin{cases} \frac{|R_1[A] - R_2[A]|}{\max(A) - \min(A)} & \text{if } A \text{ is continuous,} \\ 0, & \text{if } A \text{ discrete and } R_1[A] = R_2[A], \\ 1, & \text{if } A \text{ discrete and } R_1[A] \neq R_2[A], \end{cases} \quad (3)$$

2.4.2. Backpropagation Neural Network (BPNN). The BP neural network is a kind of multilayer feedforward neural network. Its main characteristics are forward signal transmission and error backpropagation. In the forward transfer, the input signal is processed layer by layer from the input layer through the hidden layer to the output layer. The neuron status of each layer only affects the neuron status of the next layer. If the output layer does not get the desired output, it switches to backpropagation and adjusts the network weights and thresholds according to the prediction error, so that the predicted output of the BP neural network is constantly approaching the expected output. The topology of the BP neural network is shown in Figure 2

The iterative formula of the BP algorithm is expressed as

$$x_{k+1} = x_k - a_k g_k, \quad (4)$$

where x_k is a matrix of weights and thresholds, g_k represents the gradient of the current performance function, and a_k represents the learning rate. Taking the calculation process of a three-layer BP network algorithm as an example, the input node x_m , the hidden node y_n , and the output node z_i are obtained. The network connection weight between the input layer node and the hidden layer node is w_{nm} , and the other network connection weight between the hidden layer node and the output layer node is v_{in} . If the expected value of the output node is t_i , the calculation formula of the model is as follows:

The output of the hidden node is

$$y_n = f\left(\sum_m w_{nm} x_m - \theta_n\right) = f(\text{net}_n). \quad (5)$$

The output of the output node is

$$z_i = f\left(\sum_n v_{in} y_n - \theta_i\right) = f(\text{net}_i). \quad (6)$$

The error of the output node is

$$\frac{1}{2} \sum_i \left(t_i - f\left(\sum_n v_{in} f\left(\sum_m w_{nm} x_m - \theta_n\right) - \theta_i\right) \right)^2. \quad (7)$$

Then, the following steps are performed:

- (1) The error function differentiates the output nodes;

$M_j(C)$ is the j -th nearest neighbor sample in $C \notin \text{class}(R)$, as shown in the following equation:

- (2) Network error function is used to differentiate nodes in hidden layers;
- (3) The threshold is corrected;
- (4) The derivative of the transfer function $f(x)$ is obtained [46].

2.4.3. Support Vector Machine (SVM). Support vector machine (SVM) is mainly used for pattern classification and nonlinear regression. It is a general learning algorithm based on the principle of structural risk minimization. Its basic idea is to establish a classification hyperplane as a decision surface, so that the isolated edge between positive and negative examples is maximized. This provides good generalization performance. The main reason for using support vector machines is to deal with various nonlinear problems and to flexibly use kernel functions to improve the ability to transform data from high-dimensional space. Taking a binary classification support vector machine as an example, a training set $T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (X \times Y)^l$ is given, where $x_i \in X = R^n$, $y_i \in \{1, -1\}$ ($i = 1, 2, \dots, l$) and x_i is a feature vector. The penalty parameter C and the kernel function $K(x, x')$ were selected to construct and solve the optimization problem [47]:

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j a_i a_j K(x, x') - \sum_{j=1}^l \alpha_j, \\ \text{s.t.} \quad & \sum_{i=1}^l y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, \dots, l. \end{aligned} \quad (8)$$

And, the optimal solution: $\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)^T$ is obtained. Then, a positive component of α^* , $0 \leq \alpha_j^* \leq C$ is selected and the threshold value was calculated based on this:

$$b^* = y_j - \sum_{i=1}^l y_i \alpha_i^* K(x_i - x_j). \quad (9)$$

Finally, the decision function is constructed:

$$f(x) = \text{sgn}\left(\sum_{i=1}^l \alpha_i^* y_i K(x, x_i)\right) + b^*. \quad (10)$$

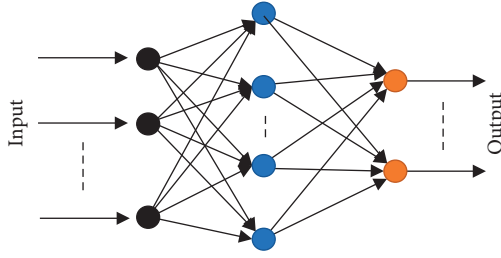


FIGURE 2: BP neural network structure.

2.4.4. Multiple Collinearity Test. Multicollinearity means that there is a high correlation between predictors in a linear regression model [48]. Before the model prediction, since the problem of collinearity among multiple variables is very common, if there is strong collinearity among the variables, the accuracy of the prediction of the prediction model may be reduced. Variance inflation factor (VIF) was used to test the predictive variables of the model, and collinearly significant variables were excluded. The remaining variables entered the next variable screening process. The expression of the variance expansion coefficient is

$$VIF = \frac{1}{1 - R_j^2}, \quad (i = 1, 2, \dots, n). \quad (11)$$

Generally, $VIF = 10$ is used as the standard. When $VIF > 10$, the collinearity between independent variables is significant, and the corresponding independent variables can be gradually eliminated to improve the accuracy of the model. In the experiment, normalized 26 independent variables were selected for linear modeling. Based on the results of the multicollinearity test, the variables with a $VIF > 10$ value were eliminated: Avst, Mast, Mist, Mrh, Ate, Mate, and Mite. The VIF values of the remaining 19 variables are shown in Table 4.

2.4.5. Quality Measures

Confusion Matrix. Confusion matrix, also called error matrix, is a standard format for expressing accuracy evaluation. It is represented by a matrix of n rows and n columns. Compared with the calculated [49], its form is shown in Table 5.

Furthermore, classification performance indicators such as accuracy, precision, recall, and f1 value are used to evaluate the model effect.

- (1) Accuracy: the ratio of the number of correct samples (TP and TN) to the total number of samples. Generally speaking, the higher the accuracy, the better the classifier. The formula is as follows:

$$P = \frac{TP + TN}{TP + FP + TN + FN}. \quad (12)$$

- (2) Precision: characterizes the classification effect of the classifier (precision effect), which predicts the correct frequency value in instances where the prediction is a positive sample:

$$P = \frac{TP}{TP + FP}. \quad (13)$$

- (3) Recall (recall rate): the recall (recall) effect of a certain class is to predict the correct frequency in the examples labeled positive samples:

$$R = \frac{TP}{TP + FN}. \quad (14)$$

- (4) f1 value: the value used to measure precision and recall, and it is the harmonic mean of these two values:

$$f1 = \frac{(1 + a^2)PR}{a^2(P + R)}. \quad (15)$$

ROC Curve. The receiver operating characteristic (ROC) curve is a method for judging the prediction effect of a model, and the prediction accuracy of the model is determined by the value of the area under the curve (AUC). The area under the curve (AUC) ranges from 0.5 to 1; the larger the value, the higher the degree of fit of the model.

3. Results

This research used Matlab 2019, a programming language for algorithm implementation. Matlab is a commercial mathematical software produced by MathWorks in the United States. It is an advanced technical computing language and an interactive environment for algorithm development, data visualization, data analysis, and numerical calculation. It mainly includes Matlab and Simulink [50]. In order to evaluate the characteristic factors and their performance based on quality metrics, this study divided the data set into two parts. The first dataset is a random extract comprising 70% of the preprocessed sample data acting as the training set. The remaining 30% of data will act as the test set.

3.1. Feature Selection. Since the multicollinearity test can only remove strong collinearity feature factors, it cannot effectively reduce the model calculation rate and reduce errors. Therefore, this paper uses the Relief algorithm to filter the remaining 19-dimensional data and remove irrelevant features, as well as reduce the dimensions to improve the model operation rate and the classification performance of the classifier. The researchers first used the Relief algorithm to calculate the weight coefficients of each feature and then arranged the feature variables in a descending order of the weight coefficients to obtain the correlation weight coefficient W of each feature and the target category. Subsequently, the weight threshold W was set to 0.005 to eliminate features smaller than the threshold to achieve an optimal feature subset. Figure 3 shows the 19 feature subsets in a descending order generated by the Relief algorithm. From Figure 3, the sunshine time weight coefficient is the largest and has the greatest correlation with the target category. The sunshine time has the greatest impact on the occurrence of

TABLE 4: Variable VIF values.

Independent variable	VIF
<i>Lon</i>	2.22
<i>Lat</i>	1.79
<i>Pre</i>	1.19
<i>Spr</i>	2.42
<i>Arh</i>	8.20
<i>Suh</i>	3.89
<i>Aws</i>	2.64
<i>Mws</i>	2.41
<i>Alt</i>	2.70
<i>Slo</i>	1.43
<i>Asp</i>	1.00
<i>Pop</i>	2.99
<i>GDP</i>	2.75
<i>NDVI</i>	1.23
<i>Ral</i>	1.34
<i>Wat</i>	1.11
<i>Hig</i>	1.10
<i>Set</i>	1.10
<i>Sfe</i>	1.04

TABLE 5: Confusion matrix form.

Prediction (column)/label (row)	Positive sample	Negative sample
Positive sample	TP	FN
Negative sample	FP	TN

forest fires, so it is taken as the first feature of the final feature subset. In contrast, the effect of the aspect weight coefficient is minimal as it has the least correlation with the target category; therefore, this variable is eliminated. Finally, the eight feature subsets of sunshine hours, average relative humidity, cumulative precipitation at 20-20 hours, average wind speed, maximum wind speed, closest distance from the fire point to the residential area, latitude, and special festivals were entered into the neural network model and support vector machine model establishment.

3.2. Applying Predictors

3.2.1. BPNN. In this study, the input layer was set to eight indicators (*Suh*, *Arh*, *Pre*, *Aws*, *Mws*, *Set*, *Lat*, and *Sfe*) after feature selection, and the output layer was 1 or 0 (to show whether a forest fire occurs or not). The output layer and categorical variable *Sfe* were processed using one-hot coding. The number of input neurons was 8, and the number of output neurons was 2. For the number of nodes in the hidden layer, a trial and error method was implemented to select the number of neurons corresponding to the least number of trainings and the smallest error. Subsequently, the optimal number of hidden layer nodes was finally determined to be 10. Therefore, the topology of the BP neural network in this study was finally determined to be 8:10:2. Researchers used the adaptive learning rate momentum factor gradient training method as the training function.

The BP neural network model used 70% of the training set to predict the remaining test set. The comparison between the predicted value and the actual value is shown in Figure 4.

Figure 4 shows that the number of errors between the predicted value and the measured value are small, and the prediction result is good. The results obtained after training are inversely normalized, and the relative errors between the results and the standard results are calculated. The calculated relative error histogram is shown in Figure 5. This figure shows that the relative errors of the training samples and the test samples are mainly scattered between -0.97 and 0.97 , and the errors are relatively small and concentrated. After model training, the accuracy rate of the model test set reached 92.16%, and the mean square error (MSE) is 0.25. Therefore, the model effect is ideal.

3.2.2. SVM. In this study, the SVM was built using the LIBSVM package in Matlab, and the model was constructed using a radial basis function (RBF) kernel function for processing nonlinear data. The researcher obtained a model with a radial basis function and parameters by training the SVM of the training group data. Following this, the test set can be inputted for a fit test. Finally, the researcher used the model performance evaluation index to judge the validity of the model.

The researchers used the grid search method and cross-validation to select parameters and determined the penalty parameter C and the kernel parameter (g), where the number of cross-validation (k) was set to 10. Figure 6 shows an SVC parameter selection result map (contour map) and an SVC parameter selection result map (3D view). After calculation, the accuracy of the grid search method reached 93.42% and the accuracy of cross-validation reached 82.60%.

It can be seen from Figure 6 that the optimal values of C and g are 0.5 and 9, respectively. After the parameters are set to the optimal values, SVM modeling is performed and the predicted values are obtained. Figure 7 shows a comparison chart between actual and predicted values.

After optimization, the number of support vectors is 5166, the number of support vectors at the boundary is 3292, and the total number of support vectors is 5166. The number of support vectors is inferred to be a response to whether the model is overlearning. Of the 18,449 support vectors in the training data, 5166 support vectors have not experienced learning.

After model training, the accuracy of the training set is 94.92%, the accuracy of the test set is 89.89%, and the mean square error (MSE) is 0.35. Therefore, it can be concluded that the model is performing at near optimal levels.

3.3. Performance Analysis. Researchers built a confusion matrix on the prediction results of the BP neural network and the support vector machine model. Figure 8 shows the confusion matrix visualization. Results of the confusion matrix and Figure 8 display the final accuracy rate of the BPNN model as 92.16%, the total number of correctly predicted samples as 7,286, and the number of incorrectly

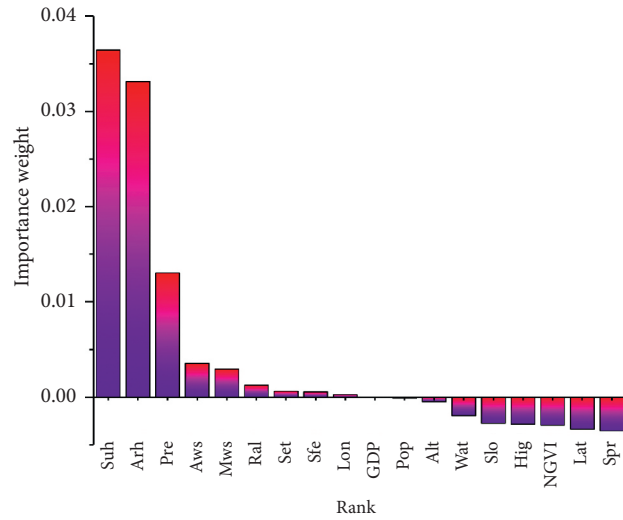


FIGURE 3: Important ranking of the 19 feature subsets.

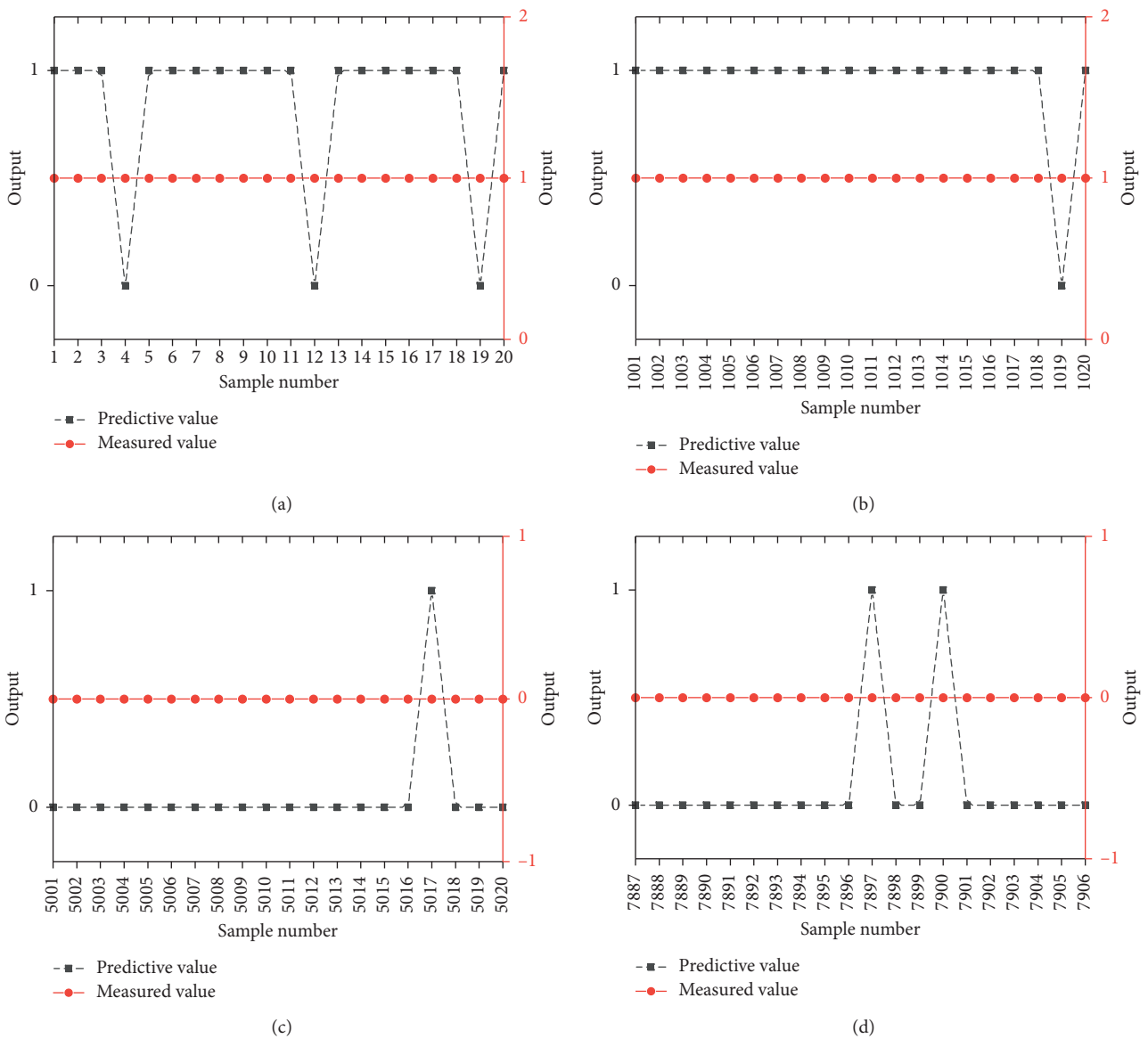


FIGURE 4: Comparison chart of the predicted and actual values of the BP neural network (a part of the sample): (a) number: 1–20, (b) number: 1001–1020, (c) number: 5001–5020, and (d) number: 7887–7906. Due to the large sample size, only a part of the sample comparison chart is displayed.

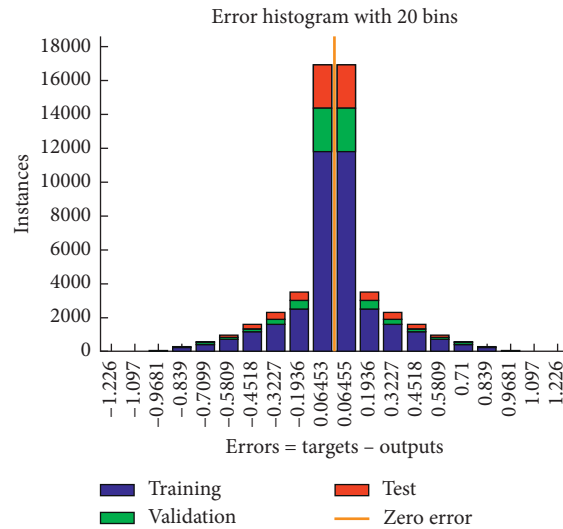


FIGURE 5: Error histogram.

predicted samples as 620. The accuracy of SVM is 89.89%, and the total number of correctly predicted samples is 7,107. The total number of prediction errors is 799, meaning the accuracy of the BP neural network model prediction is higher. This study used accuracy, precision, recall, f1, and MSE (mean squared error) indicators to analyse and compare the model’s generalization ability in more detail, while using the ROC curve and the AUC value to evaluate the prediction accuracy of the model.

Table 6 shows the specific values of the five indicators of the model test sets, and Figure 9 shows the comparison results of the evaluation indicators of the BPANN and SVM models. Table 6 shows that the mean square error of the BP neural network is less than 0.1 of the SVM model. In addition, the accuracy of the BP neural network is 0.93, which is 0.03 higher than that of the SVM model. At the same time, for the prediction results, the overall accuracy rate of BPANN is higher than SVM 0.02. Figure 9 shows that the BPANN indicators are slightly higher than the SVM. The AUC values of both models in Figure 10 are higher than 85%, indicating that both models have high accuracy in predicting forest fire occurrence in the Guangxi Autonomous Region. In addition, it can be seen from the ROC curve that the prediction effect of the BP neural network model is higher than that of the SVM model.

4. Discussion

This paper analyzed the factors that affect the occurrence of fires through feature selection. For the 26 selected forest fire driving factors, the first five factors that have the greatest impact are all meteorological factors. Therefore, it can be concluded that meteorological factors are the most important cause of fires in the Guangxi Autonomous Region, China. Meteorological factors have a direct relationship with natural or man-made forest fires. Among meteorological factors, the number of hours of sunshine is the most important factor. In the daytime, sunshine is long, the

temperature is high, and the relative humidity in the air is low, which means forest fires are more likely. During morning, evening, or night, the temperature is low and the relative humidity in the air is high, meaning forest fires are more unlikely. At the same time, precipitation is also related to humidity and directly affects the occurrence of fires [21, 51]. Human factors also play an important factor contributing to the occurrence of forest fires. These factors include the distance from fire points to residential areas and special festival factors. Frequent human activities, such as burning paper, reclamation, wasteland, or arson will cause the frequency of forest fires to gradually increase [52]. In the selection of forest fire driving factors, this study first used multicollinearity tests and then selected features. In the test of collinearity, a linear regression model is used to eliminate variables with high correlation between explanatory variables. However, due to the limited sample size and the representativeness of forest fire drivers, the results may miss some important factors related to forest fires [53]. Therefore, it is necessary to further analyse the explanatory variables before elimination in the future. More important factors of the dependent variable should not be excluded and entered into the next modeling classification. In addition, for the method of feature selection, this study uses the Relief algorithm to assign weights to features [54]; however, this algorithm cannot remove feature redundancy, hence it is necessary to optimize the algorithm or adopt other algorithms for this research. The setting of the weight threshold is the experience value chosen by the researcher, and a more scientific selection method needs to be discussed and implemented in the future.

For the application of these two models in forest fire prediction, Hong et al. [55] used support vector machine algorithms to predict forest fires in Dayu County, southwest of Jiangxi Province, China. The results show that the AUC value of the support vector machine is 0.75, while the AUC value of the support vector machine model in this experiment is 0.95. This result may be due to Haoyuan Hong’s

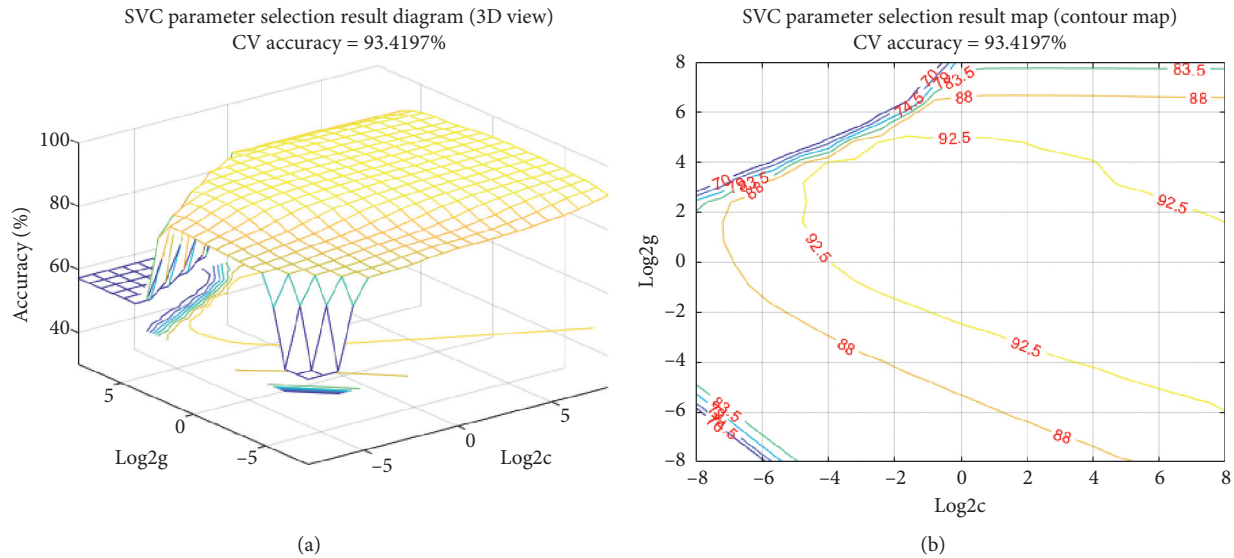


FIGURE 6: SVC parameter selection result graph. (a) 3D view. (b) Contour map.

prediction of 184 (fire and nonfire) data for Jiangxi Province, whereas this study selected 26355 samples. A larger training sample size provides more accurate and reliable results. Bisquert et al. [56] used artificial neural network models to predict forest fire risks in Galicia (northwestern Spain) and obtained an accuracy rate of 76% for the artificial neural network models. In comparison, the accuracy of the artificial neural network in this study was 92.16%. The forest fire driving factors used by Mar Bisquert included fire history, the enhanced vegetation index (EVI), and ground temperatures for each unit, whereas over twenty factors were used in this study. Multidimensional variables make model training more accurate. From a comprehensive point of view, the results of this study has reached a higher accuracy rate. The possible reason is that the researcher has comprehensively considered many forest fire driving factors, including 26 hours of sunshine, rainfall, and the distance from fire to residential sites. Secondly, for these factors, this study used a feature selection algorithm suitable for binary classification to determine the weight of each factor, which can more scientifically screen the driving factors that have less correlation with forest fire occurrence.

For the linear inseparability of the data, researchers used BP neural networks and support and vector machine models for classification and prediction. Support vector machines can map input vectors to high-dimensional spaces and better construct optimal classification surfaces through kernel functions. In this study, a radial basis function was used, and the grid search method was implemented to traverse all parameter combinations in the search range to optimize the kernel parameters and therefore improve the prediction accuracy. The analysis results showed that the prediction accuracy of RBF SVM exceeds 85%, which is significantly high. The model has a strong learning ability and good generalization ability, and this makes it suitable for predicting the occurrence of forest fires. However, the model has some shortcomings. For example, when the model

complexity is increased, the model calculation speed is significantly reduced. This is also a reason for the limitation of support vector machines, that is, solving the convex quadratic programming problem takes up a large amount of time due to the complexity, making this method impractical for large-scale sample data sets [19, 57]. It is recommended that future researchers further study and optimize the algorithm for this purpose. In addition, this study used the radial basis kernel function SVM for modeling based on previous experience, and the results are promising. However, in subsequent studies, multiple methods such as polynomial kernel functions should be used for comparison, and the support vector machine model with the best effect should be selected for prediction.

BP neural networks are a type of one-way propagation multilayer feedforward network [58], which has strong nonlinear mapping abilities, including self-learning and adaptive technology. In the course of the experiment, since there is no clear selection method for the number of nodes in the hidden layer, the previous experience generally sets the number of layers to the number of test classifications. In this experiment, the final hidden layer node number is determined by the empirical value and the trial and error method. The optimal number of model iterations is 54 times, and the data sample calculation time is 7 seconds. The model calculation speed is very fast, and the prediction accuracy is high. This combines to provide excellent predictions of forest fires. However, in this study, for the selection of the initial weight and threshold of the BP neural network, this paper used a random setting of the model. In order to better improve the prediction accuracy, genetic algorithms or particle swarm optimizations can be used in future research to optimize the initial weight and threshold of the network. From the overall analysis, the prediction accuracy of both models is greater than 85%; therefore, it has a good applicability for forest fire prediction. However, for the prediction of forest fire occurrence in the Guangxi Autonomous

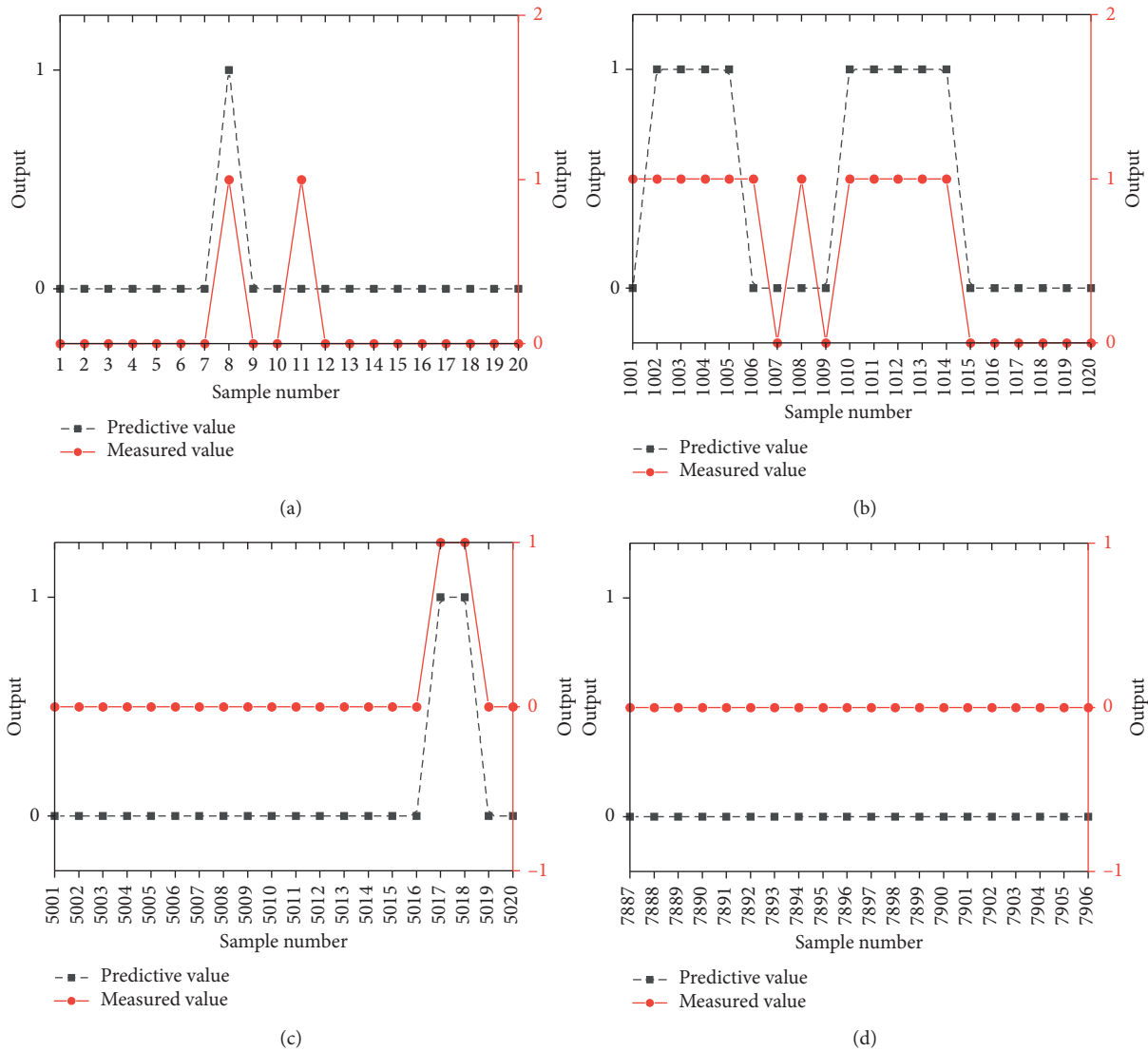


FIGURE 7: Comparison chart of SVM predicted and actual values (a part of the sample): (a) number: 1–20, (b) number: 1001–1020, (c) number: 5001–5020, and (d) number: 7887–7906.

Region, the generalization ability of the BP neural network is slightly higher than that of support vector machine; hence, the BP neural network model is more suitable for predicting the occurrence of forest fire in the Guangxi Autonomous Region, China.

5. Conclusions

This study initially conducted feature selections of forest fire driving factors. These were eight variables of sunshine, average relative humidity, cumulative precipitation at 20–20, average wind speed, maximum wind speed, closest distance of fire points to residential areas, latitude, and special festivals. These variables acted as the model inputs, and whether forest fires occurred was the output variable. Researchers established multilayer feedforward artificial neural networks and radial basis SVM models to predict forest fires in the

Guangxi Autonomous Region, China. The results showed that the accuracy of the backpropagation neural network model test set is 92.16%, and the mean square error is 0.25; the accuracy of the SVM model test set is 89.89%, and the mean square error is 0.35. The overall model training result is that the accuracy of the backpropagation neural network is 92.16%, and the AUC value is 0.99; the support vector machine is 89.89%, and the AUC value is 0.95. It can be seen from the results that both models have high prediction accuracy for Guangxi and both are applicable to the Guangxi Autonomous Region. However, it is concluded that the BP neural network model is better than the support vector machine model. BP neural networks have strong self-learning, self-adaptive abilities in addition to fast calculation speeds for large samples, which allows the best prediction models in this field of research. These results can provide a reference for future modeling of Guangxi forest fires.

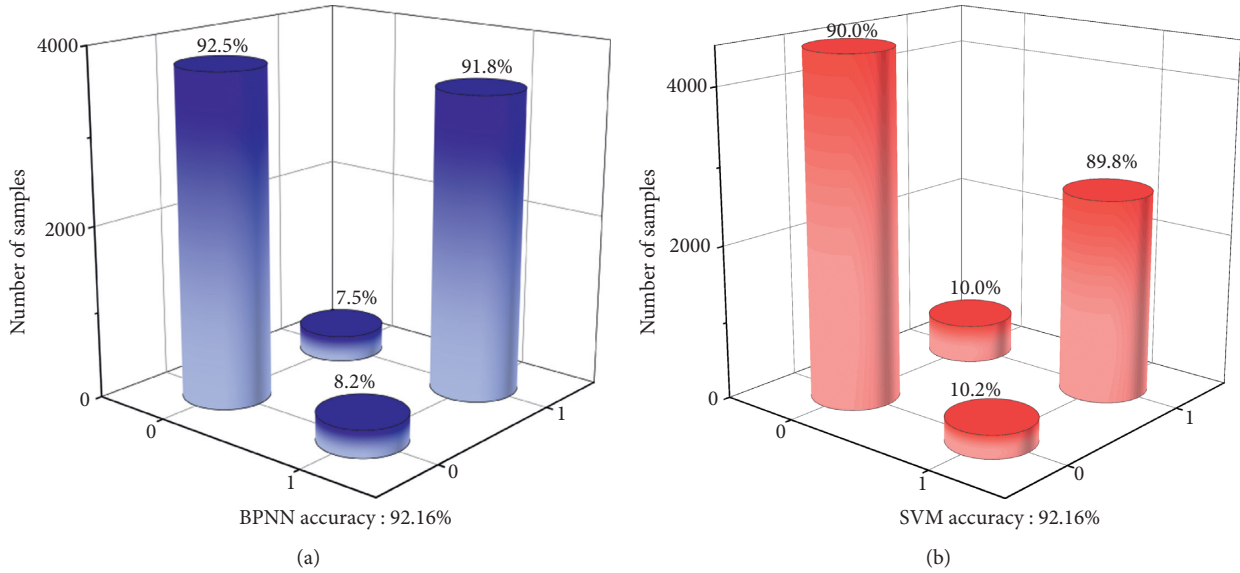


FIGURE 8: Confusion matrix visualization result. (a) BPNN. (b) SVM.

TABLE 6: Comparison of BPNN and SVM model evaluation index results.

Index	BPNN	SVM
MSE	0.25	0.35
Accuracy	0.92	0.89
Precision	0.93	0.89
Recall	0.92	0.84
f1	0.92	0.87

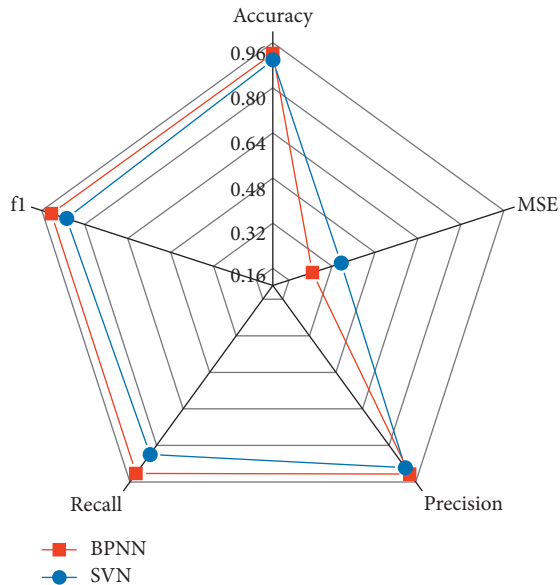


FIGURE 9: Comparison of evaluation results of BPANN and SVM models.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

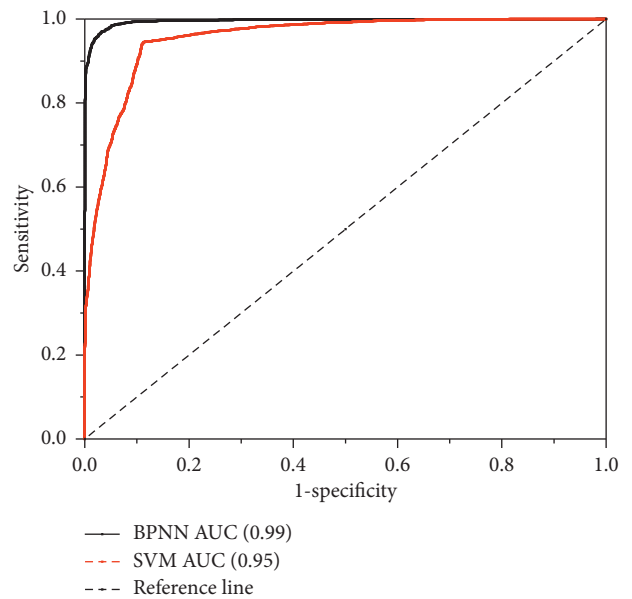


FIGURE 10: ROC curve.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to acknowledge support from the Beijing Key Laboratory for Precision Forestry, Beijing Forestry University, as well as all the people who have contributed to this paper. This research was jointly supported by the medium long-term project of “Precision Forestry Key Technology and Equipment Research” (no. 2015ZCQ-LX-01) and the National Natural Science Foundation of China (no. U1710123).

References

- [1] F. T. Guo, H. Q. Hu, Z. Hma, and Y. Zhang, "Applicability of different models in simulating the relationships between forest fire occurrence and weather factors in Daxing'an Mountains," *Chinese Journal of Applied Ecology*, vol. 21, pp. 159–164, 2010.
- [2] G. Ferrauto, R. M. S. Costa, P. Pavone, and G. L. Cantarella, "Human impact assessment on the Sicilian agroecosystems through the evaluation of melliferous areas," *Annals of Botany*, vol. 3, pp. 237–244, 2013.
- [3] R. M. S. Costa, T. van Andel, P. Pavone, and S. Pulvirenti, "The pre-Linnaean herbarium of Paolo Boccone (1633–1704) kept in Leiden (The Netherlands) and its connections with the imprinted one in Paris," *Plant Biosystems—An International Journal Dealing with All Aspects of Plant Biology*, vol. 152, no. 3, pp. 489–500, 2018.
- [4] A. Cuspilici, P. Monforte, and M. A. Ragusa, "Study of Saharan dust influence on PM 10 measures in Sicily from 2013 to 2015," *Ecological Indicators*, vol. 76, pp. 297–303, 2017.
- [5] A. Duro, V. Piccione, M. A. Ragusa, and V. Veneziano, "New environmentally sensitive patch index-ESPI-for MEDALUS protocol," *AIP Conference Proceedings*, vol. 1637, pp. 305–312, 2014.
- [6] B. S. Halpern, S. Walbridge, K. A. Selkoe et al., "A global map of human impact on marine ecosystems," *Science*, vol. 319, no. 5865, pp. 948–952, 2008.
- [7] F. Borgwardt, L. Robinson, D. Trauner et al., "Exploring variability in environmental impact risk from human activities across aquatic ecosystems," *Science of The Total Environment*, vol. 652, pp. 1396–1408, 2019.
- [8] M. H. Mona, H. A. El-Naggar, E. E. El-Gayar et al., "Effect of human activities on biodiversity in nabq protected area, south sinai, Egypt," *The Egyptian Journal of Aquatic Research*, vol. 45, no. 1, pp. 33–43, 2019.
- [9] Z. Zhuo, *Experimental and Modelling Study on Downslope Forest Fire Spread*, University of Science and Technology of China, Beijing, China, 2019.
- [10] K. Koksall, J. McLennan, D. Every, and C. Bearman, "Australian wildland-urban interface householders' wildfire safety preparations: "Everyday life" project priorities and perceptions of wildfire risk," *International Journal of Disaster Risk Reduction*, vol. 33, pp. 142–154, 2019.
- [11] Z. W. Su, A. Q. Liu, F. Guo et al., "Driving factors and spatial distribution pattern of forest fire in Fujian Province," *Journal of Natural Disasters*, vol. 25, pp. 110–119, 2016.
- [12] M. Yu, *The Establishment and Study of Regional Forest Fire Prediction and Forecasting Model*, Beijing Forestry University, Beijing, China, 2016.
- [13] Y. Song, *Research on Forest Fire Drivers and Models in Heilongjiang Province*, Northeast Forestry University, Heilongjiang, China, 2018.
- [14] K. Z. Liu, L. F. Shu, F. J. Zhao et al., "Research on spatial distribution of forest fire based on satellite hotspots data and forecasting model," *Journal of Forestry Engineering*, vol. 2, pp. 128–133, 2017.
- [15] F. T. Guo, H. Q. Hu, S. Jin et al., "Relationship between forest lighting fire occurrence and weather factors in Daxing'an Mountains based on negative binomial model and zero-inflated negative binomial models," *Chinese Journal of Plant Ecology*, vol. 21, pp. 159–164, 2010.
- [16] J. Zhang, H. Y. Zhao, M. X. Zhang et al., "Research on generalized linear models applied to forest fire forecast," *Journal of Forestry Engineering*, vol. 2, pp. 135–142, 2017.
- [17] B. Q. Liao and J. Wei, "Logistic and ZIP regression model for forest fire data," *Fire Safety Science*, vol. 17, pp. 143–149, 2008.
- [18] H. L. Liang, Y. Y. Lin, G. Yang et al., "Application of random forest algorithm on the forest fire prediction in tahe area based on meteorological factors," *Scientia Silvae Sinicae*, vol. 52, pp. 89–98, 2016.
- [19] Z. Q. Xu, *Forest Fire Area Prediction Based on Support Vector Machine*, Beijing Forestry University, Beijing, China, 2012.
- [20] S. Al-Janabi, I. Al-Shourbaji, and M. A. Salman, "Assessing the suitability of soft computing approaches for forest fires prediction," *Applied Computing and Informatics*, vol. 14, no. 2, pp. 214–224, 2018.
- [21] V. Sevinca, O. Kucukb, and M. Goltasc, "A Bayesian network model for prediction and analysis of possible forest fire Causes," *Forest Ecology and Management*, vol. 457, pp. 1–11, 2020.
- [22] T. Artés, A. Cencerrado, A. Cortés, and T. Margalef, "Time aware genetic algorithm for forest fire propagation prediction: exploiting multi-core platforms," *Concurrency and Computation: Practice and Experience*, vol. 29, 2016.
- [23] D. T. Bui, Q. T. Bui, Q. P. Nguyen, B. Pradhan, H. Nampak, and P. T. Trinh, "A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area," *Agricultural and Forest Meteorology*, vol. 233, pp. 32–44, 2017.
- [24] M. Denham, A. Cortés, T. Margalef, and E. Luque, "Applying a dynamic data driven genetic algorithm to improve forest fire spread prediction," in *International Conference on Computational Science*, vol. 5103, pp. 36–45, Springer Berlin Heidelberg, Heidelberg, Germany, 2008.
- [25] H. Hong, S. A. Naghibi, M. M. Dashtpaderdi, H. R. Pourghasemi, and W. Chen, "A comparative assessment between linear and quadratic discriminant analyses (LDA-QDA) with frequency ratio and weights-of-evidence models for forest fire susceptibility mapping in China," *Arabian Journal of Geosciences*, vol. 10, p. 167, 2017.
- [26] A. M. Özbayog'lu and R. Bozer, "Estimation of the burned area in forest fires using computational intelligence techniques," *Procedia Computer Scienc*, vol. 12, pp. 282–287, 2012, <https://doi.org/10.1016/j.procs.2012.09.070>.
- [27] J. B. Yang and X. Q. Ma, "On the basis of artificial neural network to forecast the forest fire in guangdong province," *Scientia Silvae Sinicae*, vol. 41, pp. 127–132, 2015.
- [28] B. A. Alonso, O. Fontenla-Romero, B. Guijarro-Berdiñas et al., "A neural network approach for forestal fire risk estimation," in *Proceedings of the 15th European Conference on Artificial Intelligence*, IOS, Lyon, France, pp. 643–647, July 2002.
- [29] G. E. Sakr, I. H. Elhajj, and G. Mitri, "Efficient forest fire occurrence prediction for developing countries using two weather parameters," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 5, pp. 888–894, 2011.
- [30] C. Hu, *The Regional Forest Fire Research Based on BP Neural Network*, Zhejiang Ocean University, Zhejiang, China, 2015.
- [31] L. Wang, Y. Hao, and M. Wen, "A multi-factor forest fire risk rating prediction model based on particle swarm optimization algorithm and back-propagation neural network," *Journal of Forestry Engineering*, vol. 4, pp. 137–144, 2019.
- [32] Z. Q. Xu, X. Y. Su, and Y. Zhang, "Forest fire prediction based on support vector machine," *Chinese Agricultural Science Bulletin*, vol. 28, pp. 126–131, 2012.
- [33] B. C. Ko, K.-H. Cheong, and J.-Y. Nam, "Fire detection based on vision sensor and support vector machines," *Fire Safety Journal*, vol. 44, no. 3, pp. 322–329, 2009.

- [34] D. W. Xie and S. L. Shi, "Prediction for burned area of forest fires based on SVM model," *Applied Mechanics and Materials*, vol. 513–517, pp. 4084–4089, 2014.
- [35] J. Zhao, Z. Zhang, S. Han, C. Qu, Z. Yuan, and D. Zhang, "SVM based forest fire detection using static and dynamic features," *Computer Science and Information Systems*, vol. 8, no. 3, pp. 821–841, 2011.
- [36] P. Cortez and A. Morais, "A data mining approach to predict forest fires using meteorological data," *New Trends in Artificial Intelligence*, pp. 512–523, 2007.
- [37] Y. He, *Occurrence Regularity of Forest Fire in Guangxi and its Correlation with Meteorological Factors*, Central South University of Forestry & Technology, Changsha, China, 2018.
- [38] F. Lu, "Study on the occurrence of forest fires in Guangxi," *Guangxi Forestry Science*, vol. 40, pp. 189–191, 2011.
- [39] K. Z. Liu, *Fire Environment and Forecasting Models of Summer Forest Fire in Daxing'an Mountain District*, Chinese Academy of Forestry, Beijing, China, 2018.
- [40] F. Guo, Z. Su, G. Wang et al., "Understanding fire drivers and relative impacts in different Chinese forest ecosystems," *Science of the Total Environment*, vol. 605–606, pp. 411–425, 2017.
- [41] F. Guo, G. Wang, Z. Su et al., "What drives forest fire in Fujian, China? Evidence from logistic regression and Random Forests," *International Journal of Wildland Fire*, vol. 25, no. 5, pp. 505–519, 2016.
- [42] Y. Chang, Z. Zhu, R. Bu et al., "Predicting fire occurrence patterns with logistic regression in Heilongjiang Province, China," *Landscape Ecology*, vol. 28, no. 10, pp. 1989–2004, 2013.
- [43] W. Y. Ma, Z. K. Feng, Z. X. Cheng et al., "Study on driving factors and distribution pattern of forest fires in Shanxi province," *Journal of Central South University of Forestry & Technology*, vol. 40, pp. 1–13, 2020.
- [44] M. Y. He and H. Zhou, "Relieff-MFO multi-label feature selection algorithm," *Computer Engineering and Design*, vol. 40, pp. 3469–3473, 2019.
- [45] W. Jing and W. P. Ding, "SVM algorithm based on feature selection of differential space fusion," *Computer Engineering and Design*, vol. 40, pp. 2235–2241, 2019.
- [46] L. Q. Feng, Q. F. Liu, and Z. H. Lu, "Research on the prediction model of potato cultivation environment regulation and control based on neural network," *Computer Era*, vol. 12, pp. 19–27, 2019.
- [47] X. C. Wang, *43 Cases of MATLAB Neural Network Analysis*, Bei hang University Press, Beijing, China, 2013.
- [48] D. Chen, "Prediction of forest fire occurrence in Daxing'an mountains based on logistic regression model," *Forest Resources Management*, vol. 1, pp. 116–122, 2019.
- [49] Y. Guo, *Research on Automatic Classification Algorithm of High - Resolution Image Based on SVM and its System Realization*, Hebei University of Science and Technology, Shijiazhuang, China, 2017.
- [50] M. Zhang, *Research on 3D Reconstruction from Contour Map and Program Based on MATLAB and VRML*, Shan Dong University, Jinan, China, 2007.
- [51] P. H. S. Mota, S. J. S. S. d. Rocha, N. L. M. d. Castro et al., "Forest fire hazard zoning in Mato Grosso state, Brazil," *Land Use Policy*, vol. 88, 2019.
- [52] A. C. Zeng, Q. J. Cai, Z. W. Su et al., "Seasonal variation and driving factors of forest fire in Zhejiang Province, China, based on MODIS satellite hot spots," *Chinese Journal of Applied Ecology*, vol. 31, pp. 399–406, 2020.
- [53] A. Senawi, H.-L. Wei, and S. A. Billings, "A new maximum relevance-minimum multicollinearity (MRmMC) method for feature selection and ranking," *Pattern Recognition*, vol. 67, pp. 47–61, 2017.
- [54] O. Reyes, C. Morell, and S. Ventura, "Scalable extensions of the ReliefF algorithm for weighting and selecting features on the multi-label learning context," *Neurocomputing*, vol. 161, pp. 168–182, 2015.
- [55] H. Hong, P. Tsangaratos, I. Ilia, J. Liu, A.-X. Zhu, and C. Xu, "Applying genetic algorithms to set the optimal combination of forest fire related variables and model forest fire susceptibility based on data mining models. The case of Dayu County, China," *Science of the Total Environment*, vol. 630, pp. 1044–1056, 2018.
- [56] M. Bisquert, E. Caselles, J. M. Sánchez, and V. Caselles, "Application of artificial neural networks and logistic regression to the prediction of forest fire danger in Galicia using MODIS data," *International Journal of Wildland Fire*, vol. 21, no. 8, p. 1025, 2012.
- [57] S. Y. Feng and X. W. Gong, "Application of support vector machines and artificial neural networks in the risk prediction of advanced saphenous vein graft disease after coronary artery bypass grafting," *Chinese Journal of Health Statistics*, vol. 36, pp. 493–496, 2019.
- [58] H. Y. Shen, Z. X. Wang, C. Y. Gao et al., "Determining the number of BP neural network hidden layer units," *Journal Of Tianjin University of Technology*, vol. 24, pp. 13–15, 2008.