



Forest Fire Susceptibility Modeling Using a Convolutional Neural Network for Yunnan Province of China

Guoli Zhang¹ · Ming Wang^{1,2} · Kai Liu^{1,2}

© The Author(s) 2019

Abstract Forest fires have caused considerable losses to ecologies, societies, and economies worldwide. To minimize these losses and reduce forest fires, modeling and predicting the occurrence of forest fires are meaningful because they can support forest fire prevention and management. In recent years, the convolutional neural network (CNN) has become an important state-of-the-art deep learning algorithm, and its implementation has enriched many fields. Therefore, we proposed a spatial prediction model for forest fire susceptibility using a CNN. Past forest fire locations in Yunnan Province, China, from 2002 to 2010, and a set of 14 forest fire influencing factors were mapped using a geographic information system. Over-sampling was applied to eliminate the class imbalance, and proportional stratified sampling was used to construct the training/validation sample libraries. A CNN architecture that is suitable for the prediction of forest fire susceptibility was designed and hyperparameters were optimized to improve the prediction accuracy. Then, the test dataset was fed into the trained model to construct the spatial prediction map of forest fire susceptibility in Yunnan Province. Finally, the prediction performance of the proposed model was assessed using several statistical measures—Wilcoxon signed-rank test, receiver operating characteristic curve, and area under the curve (AUC). The results confirmed the

higher accuracy of the proposed CNN model (AUC 0.86) than those of the random forests, support vector machine, multilayer perceptron neural network, and kernel logistic regression benchmark classifiers. The CNN has stronger fitting and classification abilities and can make full use of neighborhood information, which is a promising alternative for the spatial prediction of forest fire susceptibility. This research extends the application of CNN to the prediction of forest fire susceptibility.

Keywords China · Convolutional neural network · Forest fire susceptibility · Geographic information system · Machine learning

1 Introduction

Forest fires have caused considerable losses in global forest resources and people's lives and property, seriously impacting the global ecological balance, and have received considerable attention from countries worldwide. In recent years, with global warming, industrialization, and human interventions, the frequency and severity of forest fires have been increasing significantly in many parts of the world (Crimmins 2006; Running 2006; Hantson et al. 2015). Regional forest fire susceptibility is often affected by many factors and has typical nonlinear and complex characteristics; therefore, it is still a difficult task to develop forest fire prediction models with satisfactory accuracy (Ngoc Thach et al. 2018). Various approaches have been developed for modeling forest fire susceptibility, ranging from physics-based methods to statistical and machine learning (ML) methods (Dimuccio et al. 2011; Tien Bui et al. 2017; Leuenberger et al. 2018; Hong et al. 2019; Jaafari et al. 2019). Compared to traditional

✉ Ming Wang
wangming@bnu.edu.cn

¹ State Key Laboratory of Earth Surface Processes and Resource Ecology / Academy of Disaster Reduction and Emergency Management, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

² Key Laboratory of Environmental Change and Natural Disasters, Ministry of Education, Beijing Normal University, Beijing 100875, China

qualitative and statistical analysis methods, ML approaches have shown the ability to provide better results for the spatial prediction of wildfires (Bar Massada et al. 2013). In the last decade, various ML algorithms—such as artificial neural networks (Dimuccio et al. 2011; Bisquert et al. 2012; Satir et al. 2016), random forests (RF) (Oliveira et al. 2012; Arpaci et al. 2014; Pourtaghi et al. 2016), support vector machine (SVM) (Hong et al. 2018), multilayer perceptron neural network (MLP) (Vasconcelos et al. 2001; Satir et al. 2016), kernel logistic regression (KLR) (Hong et al. 2015; Bui, Le, et al. 2016), naive Bayes (Elmas and Sönmez 2011; Jaafari et al. 2018), gradient boosted decision trees (Sachdeva et al. 2018), and particle swarm optimized neural fuzzy (Tien Bui et al. 2017)—have been successfully developed and widely applied for producing wildfire susceptibility maps. Comparative studies of multiple ML algorithms have also been employed (Pourtaghi et al. 2016; Cao et al. 2017; Ngoc Thach et al. 2018; Kim et al. 2019). Therefore, advanced ML approaches are very promising for forest fire spatial prediction. However, the ML approaches mentioned above are pixel-based classifiers with shallow architectures, which do not make use of the spatial patterns that are implicit in images (Zhang et al. 2018). In addition, these classifiers directly classify the input data without feature extraction, and representative features cannot be mined from the input data to improve classification accuracy.

Deep learning (DL) methods (Hinton and Salakhutdinov 2006; Lecun et al. 2015) have recently received more attention and achieved remarkable success. Deep learning algorithms attempt to discover multiple representation levels (Schmidhuber 2015) and have been broadly applied in areas such as object recognition and detection, speech recognition, and natural language processing (Lecun et al. 2015). The convolutional neural network (CNN) (LeCun et al. 1998), which has been recognized as one of the most successful and widely used DL algorithms, has produced significant improvements in the latest studies in areas such as disaster damage detection (Muhammad et al. 2018; Vetrivel et al. 2018), remotely sensed image classification (Liu and Abd-Elrahman 2018; Zhang et al. 2018), and landslide susceptibility mapping (Wang et al. 2019). However, none of these studies evaluated the effectiveness of CNN in the prediction of forest fire susceptibility. The first law of geography (Tobler 1970) emphasizes that near things are more closely related than distant things. Whether a pixel is an ignition point should not only consider the situation of the pixel itself, but also consider other pixels within a certain range around the pixel. While pixel-based classifiers may overlook certain information in spatial patterns, the contextual-based CNN explores the complex spatial patterns that are implicit in images (Zhang et al. 2018). The CNN can make full use of contextual

information (that is, neighborhood information) and can discover multiple levels of representations from input data, which is more suitable for the evolution of fire event spatial characteristics. The DL process reveals the deep features and can distinguish the differences between different geographical units. Therefore, there is a certain practical significance in studying the application of the CNN algorithm in forest fire susceptibility analysis.

Forest fire susceptibility, in this article, is defined as the probability estimation of fire occurrence in a region. The main objective of this study is to utilize contextual-based CNN with deep architectures for the spatial prediction of regional forest fire susceptibility in Yunnan Province, China. The forest fire susceptibility model was established based on a CNN and the hyperparameters of the model were optimized to improve the prediction accuracy. The performance of the proposed model was compared with benchmark methods using several statistical measures—Wilcoxon signed-rank test (WSRT), receiver operating characteristic (ROC) curve, and area under the curve (AUC).

2 Study Area and Data Collection

Yunnan Province is located in southwestern China (Fig. 1). It is a mountainous region with over 90% mountain and plateau landscape interspersed with less than 10% small, scattered valley basins. The terrain slopes downward from northwest to southeast, and elevation ranges from 0 to 6135 m. The region has a plateau-type tropical monsoon climate, and average temperatures in the summer and winter are 19–22 °C and 6–8 °C, respectively. The distribution of precipitation throughout the seasons and regions is extremely uneven. The winter and spring seasons from November to April of the following year are dry seasons with precipitation accounting for only 20% or less of the 1100 mm annual precipitation. Yunnan has abundant and diverse forest resources, including tropical rainforests, seasonal rainforests, subtropical evergreen broad-leaved forests, and temperate coniferous forests. Compared with other regions in China, the frequency of forest fires in Yunnan is relatively high (Ying et al. 2018). The occurrence of forest fires in Yunnan has a strong seasonal pattern, mainly concentrated in the spring from mid-February to mid-May.

Compiling a forest fire inventory is a mandatory task for forest fire susceptibility modeling (Tien Bui et al. 2017). In this study, a forest fire event map was prepared using multiple resources including historical fire reports and the interpretation of satellite images (see Cao et al. (2017) for details). From 2002 to 2010, a total of 7675 fires occurred, and the number of fires in the spring was 4428, accounting

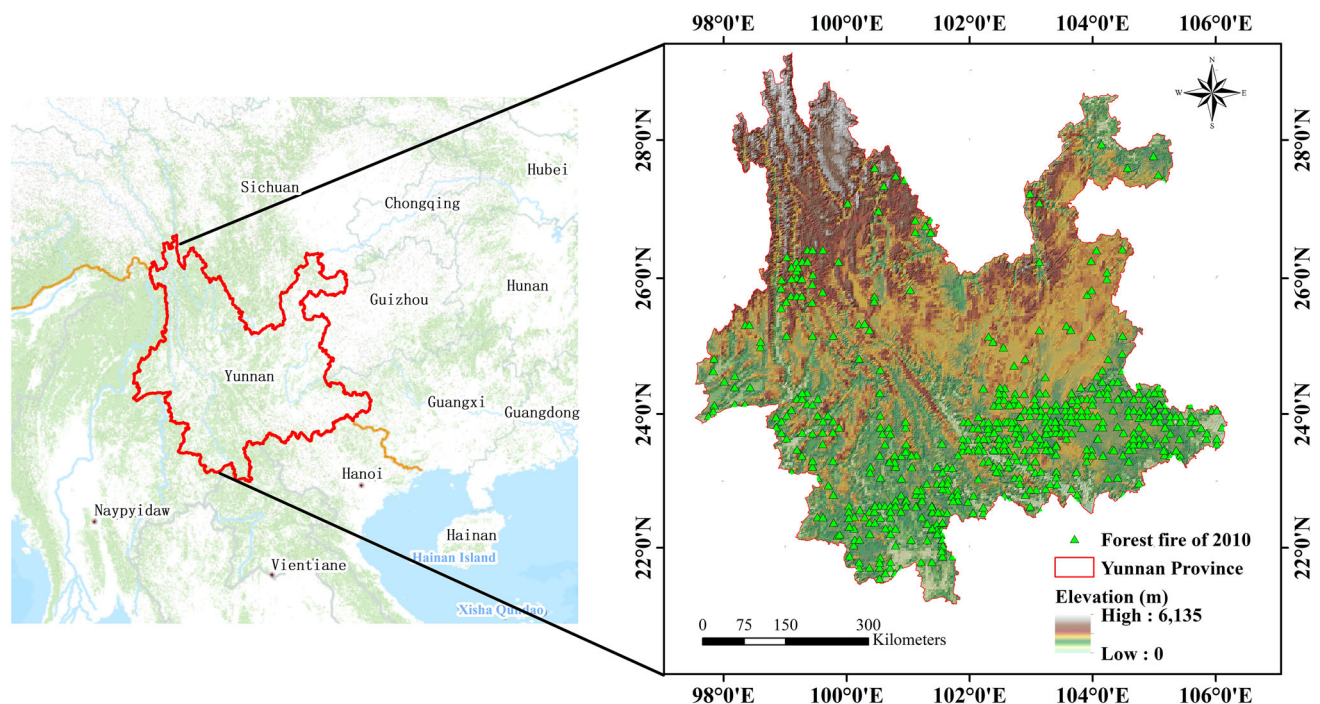


Fig. 1 Location of Yunnan Province, the study area, in China

for 58% of the total number of fires. The occurrence of forest fires is affected by many factors, and selecting the appropriate influencing factors is important (Pew and Larsen 2001; Guo et al. 2016). Fourteen forest fire influencing factors were selected, and all datasets were prepared in the form of raster/vector data, as described and listed in Table 1.

3 Methodology

This section first provides a preview of the CNN algorithm, and then elaborates the procedure of the susceptibility model development through a series of processes including data preprocessing and sample library generation, model architectures design and parameter adjustment, and performance evaluation. Finally, the detailed methodological flowchart is described.

3.1 Preview of Convolutional Neural Network

Convolutional neural network (CNN) is one of the most notable DL approaches and has exhibited robust performance in feature learning for image classification and recognition. It is a feed-forward neural network whose parameters are trained by using the classic stochastic gradient descent based on the backpropagation algorithm (Hu et al. 2015).

Generally, the CNN consists of several building blocks—convolutional, pooling, and fully connected layers (Yamashita et al. 2018). The different types of processing layers play different roles. The convolutional layers, which perform linear convolution operations between the input tensor and a set of filters, output the feature maps. Typically, each feature map is then followed by a nonlinear activation function. The rectified linear unit (ReLU), which performs the nonlinear transformation of the feature map generated by the convolution layer and introduces nonlinearity into the system, is the most commonly used activation function.

The purpose of the convolution operation is to extract different input layer features and achieve weight sharing. The input and output of each stage are sets of arrays called feature maps. For example, if the input is a 2-dimensional image x , the input is first decomposed into a sequential array $x = \{x_1, x_2, \dots, x_N\}$. The convolutional layer is defined as:

$$y_j = f\left(b_j + \sum_i k_{ij} * x_i\right)$$

where y_j denotes the j th output for the convolutional layer and x_i denotes each input feature map. k_{ij} denotes the convolutional kernel with the i th input map x_i . $*$ Denotes the discrete convolution operator, b_j denotes a trainable bias, and f is the nonlinear activation.

The pooling layers perform a subsampling operation to reduce the dimensions of the feature maps. According to

Table 1 Data description of forest fire influencing factors

No.	Data	Scale/resolution original	Unit	Original data format	Source
1	Elevation	30 m	m	Raster	ASTER GDEM ^a
2	Slope	30 m	degree	Raster	
3	Aspect	30 m	degree	Raster	
4	Average temperature	0.312° (lat/long)	°C	NetCDF	CFSR ^b
5	Average precipitation	0.312° (lat/long)	kg/m ²	NetCDF	
6	Surface roughness	0.312° (lat/long)	Ratio	NetCDF	
7	Average wind speed	0.312° (lat/long)	m/s	NetCDF	
8	Maximum temperature	0.312° (lat/long)	°C	NetCDF	
9	Specific humidity	0.312° (lat/long)	Ratio	NetCDF	
10	Precipitation rate	0.312° (lat/long)	mm/h	NetCDF	Vegetation map (Zhang 2007)
11	Forest coverage ratio	1:1,000,000	Ratio	Vector	
12	NDVI	500 m	Ratio	Raster	
13	Distance to roads	1:40,000,000	km	Vector	
14	Distance to rivers	1:40,000,000	km	Vector	Basic geographic data ^d
					Basic geographic data ^e

ASTER GDEM advanced spaceborne thermal emission and reflection radiometer global digital elevation map; CFSR climate forecast system reanalysis; NetCDF network common data form; NDVI normalized difference vegetation index

^a<https://search.earthdata.nasa.gov>

^b<https://rda.ucar.edu>

^c<http://www.gscloud.cn>

^d<http://www.ngcc.cn>

^e<http://www.ngcc.cn>

the maximum and average functions, the pooling layer can be divided into the max-pooling and average-pooling layers. The fully connected layers, which are the flat feed-forward neural network layers, provide high-level abstraction features. They are often used at the end of the network architecture and create the final nonlinear combinations of features for making the predictions by the network. The activation function for the last fully connected layer needs to be selected reasonably based on given tasks. The softmax or sigmoid function can be used to compute the posterior probability for each grid cell.

3.2 Data Preprocessing and Sample Library Generation

First, appropriate forest fire influencing factors were selected, and variables raster datasets (VRDs) and ignition raster datasets (IRDs) were constructed through preprocessing. Then, the sample libraries were collected from the established IRDs and VRDs using an appropriate sampling method.

3.2.1 Forest Fire Influencing Factors

Four categories of forest fire influencing factors were considered, including topography-related, climate-related, vegetation-related, and human activities-related variables.

ArcGIS 10.5 was employed for handling geographic data. The effect of the topography has been considered as a significant feature in forest fire assessment (Renard et al. 2012; Adab et al. 2013). Three topography-related influencing factors—elevation, slope, and aspect—were retrieved from the digital elevation model (DEM). The DEM was produced by the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) GDEM version 2 (with a 30 m pixel size). Surface roughness was obtained from the Climate Forecast System Reanalysis (CFSR). The climatic characteristics of an area affect the occurrence and intensity of forest fires (Moritz et al. 2012). Six meteorologically related influencing factors—average temperature, average precipitation, average wind speed, maximum temperature, specific humidity, and precipitation rate—were obtained from the CFSR. The spatial resolution is approximately 0.3 degree ($0.312^\circ \times 0.312^\circ$) with a 6 h temporal resolution. These influencing factors were calculated as the spring means from March through May. Six meteorologically related factors were finally mapped with the inverse distance weighted (IDW) interpolation method. The Moderate Resolution Imaging Spectroradiometer (MODIS) normalized difference vegetation index (NDVI) has also been identified as an important variable in forest fire modeling (Bajocco et al. 2015). The NDVI values reflect the vegetation's health and essentially the fuel load distribution (Yi et al. 2013). The MODIS NDVI monthly

synthetic products with a 500 m resolution were obtained from the Geospatial Data Cloud website. The clip and raster calculator tools in ArcGIS10.5 were applied to calculate the annual spring NDVI of Yunnan for the period 2002–2010. The NDVI map for each year was calculated by taking the average NDVI of the spring months (March, April, and May). The forest coverage data for Yunnan that were used in this research were derived from the vegetation map of the People's Republic of China (1:1,000,000). The forest coverage ratio map was calculated by the ratio of the forest area of each pixel to the total area of the pixel. The distance from the river network and the distance from the road network (highways, main roads, and local roads) were obtained by applying the buffer tool in the proximity toolbox, which calculates the Euclidean distance from the road and river networks. Then, maps of the proximity to roads and rivers were produced. Figure 2 shows 11 maps of the 14 influencing factors in 2010. Not shown in Fig. 2 are maximum temperature, specific humidity, and precipitation rate, which are discussed in more detail in Sect. 4.1.

3.2.2 Influencing Factor Evaluators

Variable selection is particularly important in the prediction of forest fire susceptibility. The high dimensionality of the training dataset may complicate the prediction process and decrease the prediction accuracy. In this study, multicollinearity analysis and an information gain ratio (IGR) were selected to evaluate the forest fire influencing factors. Multicollinearity analysis (O'Brien 2007) was applied to estimate the correlation between the forest fire influencing factors. Two measures of variance—inflation factors (VIF) and tolerances (TOL)—were used to identify the degree of multicollinearities among the forest fire influencing factors. A factor is considered to be multicollinear if its tolerance is less than 0.1 or the VIF value is greater than 10 (Colkesen et al. 2016). The IGR is an effective approach for selecting an optimal subset of variables that can represent the whole dataset to improve the prediction performance in forest fire susceptibility mapping (Dash and Liu 1997; Jaafari et al. 2018). The average merit (AM) reveals the importance of forest fire influencing factors in predicting forest fire occurrence (Jaafari et al. 2018).

3.2.3 Building Variables Raster Datasets and Ignition Raster Datasets

The min–max normalization process was conducted for the influencing factor maps to avoid the potential bias caused by the unbalanced magnitudes of factors. Rasterization was performed by the ArcGIS model builder tool to convert all maps to a raster format with the same pixel size (5000 m × 5000 m), the same data type (8 bit-unsigned integer), and

the same coordinate system (WGS 1984 Web Mercator), resulting in 18,718 cells in total. Finally, the composite bands tool was used to combine all the factor maps in a year into one raster, which was named the VRD. All the data of influencing factors from 2002 to 2010 were processed in the same way and a total of nine VRDs were finally established.

Class imbalance refers to the number of events in the nonfire class being much greater than that in the fire class (López et al. 2013). Class imbalance has a detrimental impact on the classification performance of the CNN and oversampling has been proved to be a robust solution for solving this problem in DL (Buda et al. 2018). In addition, according to the spatial characteristics of forest fire events, a certain area near the pixel of a fire vector point may be a fire-prone area. Therefore, to solve the class imbalance, buffer analysis was used with the existing fire vector points (Cao et al. 2017) to increase the number of events in the fire class. The pixels in the 5 km buffer zone were resampled to 1 (fire), and the pixels outside the buffer zone were resampled to 0 (nonfire). Then, the generated raster data—the IRD—were obtained. All fire datasets from 2002 to 2010 were processed, and nine IRDs for the corresponding years were obtained. All IRDs use the same pixel size, data type, and coordinate system as the VRDs.

3.2.4 Sample Library Generation

Since the CNN conducts its effective training in a fully supervised manner, the training and validation sample libraries must be constructed first. A binary classification method was adopted for the susceptibility analysis of forest fires, and samples were classified to either the forest fire class or the nonfire class, with 1 representing the fire class and 0 representing the nonfire class.

The CNN has many learnable parameters to estimate; thus, this predictor requires more data to achieve sufficient training. Table 2 shows that there is a large annual variation in the actual number of fire points in 2002–2009 (data from 2010 were used as the test dataset and therefore not included in the training and validation process). A simple random sampling method may lead to sample quantity imbalances, while small samples would lead to overfitting the model. Therefore, to generate more samples and keep the sample quantity balanced, proportional stratified sampling was adopted. The total number of forest fires was multiplied by the sampling rate (0.8) to determine the number of fire samples in every year. The same number of nonfire samples was randomly selected, which constitutes a total of 49,706 training samples.

After the number of samples needed for each year is determined (Table 2), a list of XY coordinates for fire or nonfire class was randomly generated and the locations of

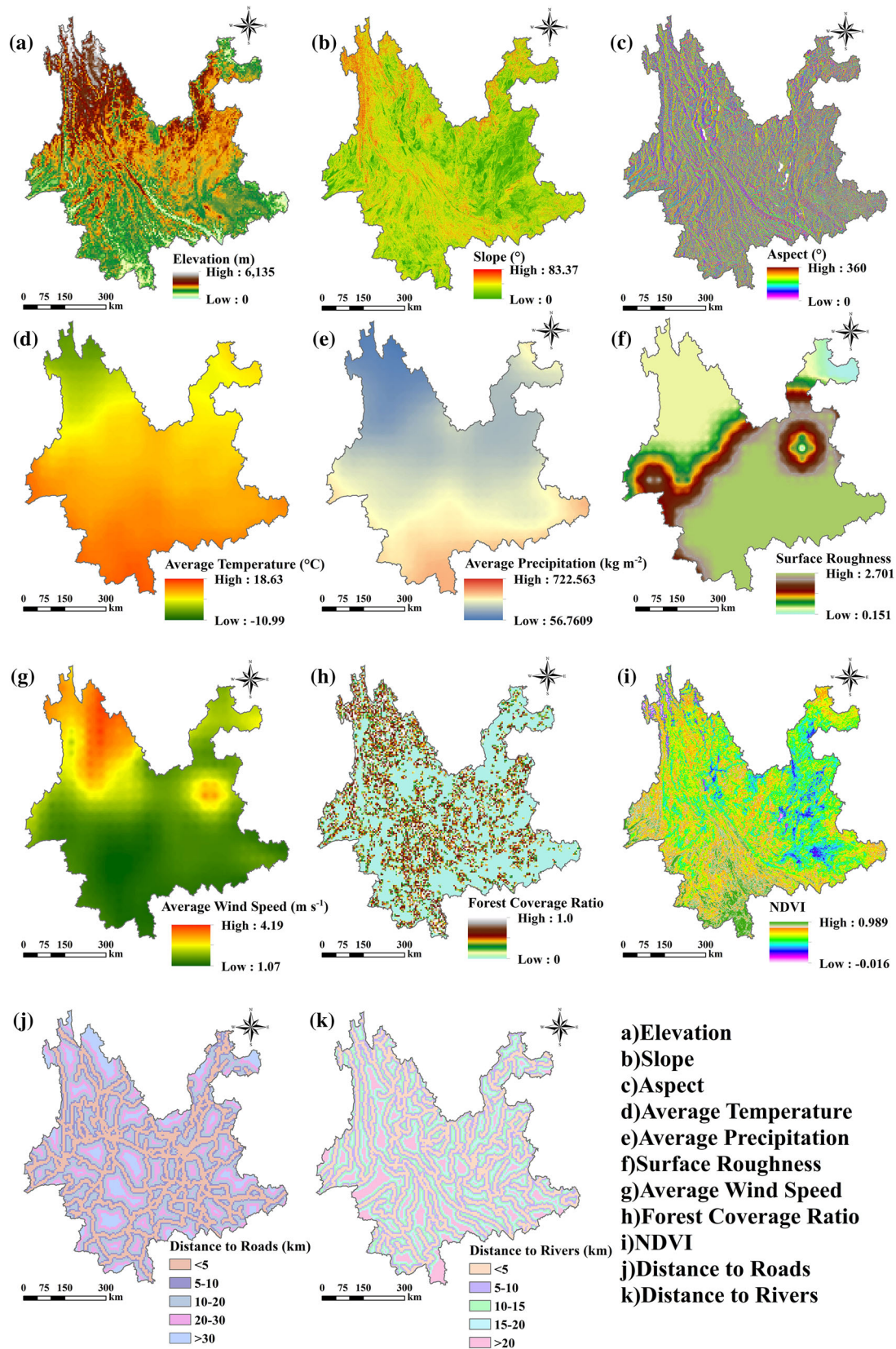


Fig. 2 Forest fire influencing factor maps in 2010 for Yunnan Province, China. *NDVI* normalized difference vegetation index

the sample points were recorded—the list contains the same number of coordinates as the sample quantity of each year (Table 2) (XY was defined using the row and column number of the IRD). Then the same number of windows were defined, each with the size of 25×25 pixels and the center of the window was the XY coordinates in the XY coordinate list. Then, the VRD in the window with 25×25 pixels was extracted as a 3-dimensional array, $n \times n \times c$, where n denotes the row and column of each input patch and c represents the number of forest fire influencing factors. Each sample consisted of two parts: a 3-dimensional array containing influencing factors; and a corresponding ground truth label from the IRD in the central XY coordinates. The dataset was divided randomly into two parts: (1) 80% as the training samples (49,706 pixels); and (2) the remaining 20% as the validation samples (12,426 pixels).

3.3 The Proposed Convolutional Neural Network Architectures and Related Parameters

The architecture of the proposed convolutional neural network (CNN) model was completed by referring to the AlexNet model (Krizhevsky et al. 2013) and the architecture and hyperparameters were tuned based on our datasets. As mentioned above, each input patch was a 3-dimensional data representation of size $n \times n \times c$, taking $25 \times 25 \times 11$ as an example. Figure 3 shows the main architecture of the prediction model of the CNN for forest fire susceptibility. There were a total of three convolution layers, two pooling layers, and three fully connected layers. The first three consecutive convolution layers had 64, 128, and 256 kernels, with uniform kernel sizes of 3×3 . Each convolution layer was followed by an activation function (ReLU) and a pooling layer. Zero padding was employed to retain in-plane dimensions. All of the pooling layers perform max-pooling and summarize a 2×2 neighborhood with a stride

Table 2 Training sample statistics

Year	Actual fire points	Samples in the fire buffer	Sampling rate	Number of fire samples	Number of nonfire samples
2002	158	1536	0.8	1229	1229
2003	267	2687	0.8	2150	2150
2004	522	4234	0.8	3387	3387
2005	620	4708	0.8	3766	3766
2006	625	4801	0.8	3841	3841
2007	835	6092	0.8	4874	4874
2008	339	3014	0.8	2411	2411
2009	491	3994	0.8	3195	3195
total	3857	31,066		24,853	24,853

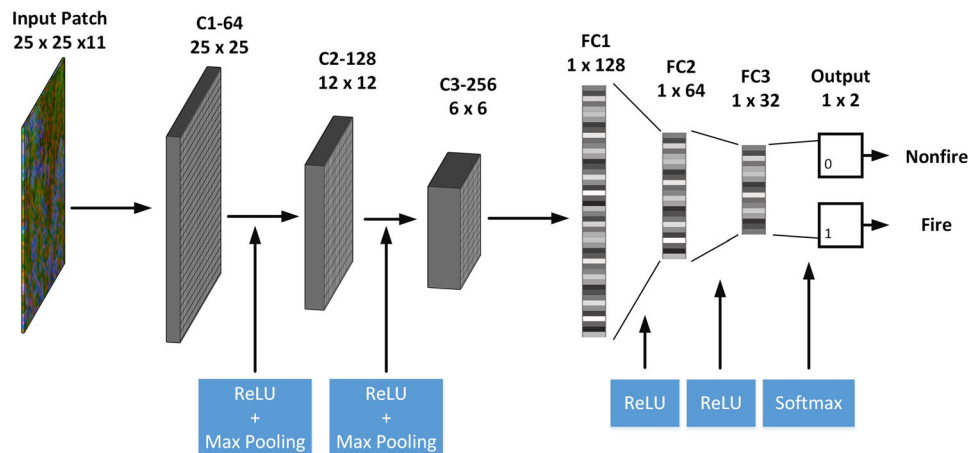


Fig. 3 The architectural design of the proposed convolutional neural network (CNN) model (C1–C3 are convolutional layers and FC1–FC3 are fully connected layers. The values on the right and below C1–C3 indicate the number of filters and their sizes. The values below FC1–FC3 indicate their dimensions, that is, the number of neurons in the fully connected layer.). *ReLU* rectified linear unit

Table 3 A list of parameters and hyperparameters utilized in the convolutional neural network (CNN) model

	Parameters	Hyperparameters
Convolution layer	Kernels	Kernel size: 3×3 ; number of kernels: 64, 128, 256; stride=1; padding; activation function=ReLU
Pooling layer	None	Pooling methods: max-pooling; filter size: 2×2 ; stride=2
Fully connected layer	Weights	Number of weights; activation function: Softmax
Others		Model architecture; initialize weights; optimizer; loss function; window size; epochs; learning rate; dropout

ReLU rectified linear unit

of 2 pixels. At the end, the next three weight layers were fully connected layers with 128, 64, and 32 neurons each. Finally, the output of the last fully connected layer was fed into a 2-way classifier with an activation function named softmax, which computes the probabilities for the two classes' labels.

The parameters in CNN, which are automatically learned during the training process, refer to kernels in the convolution layers and weights in the fully connected layers. Training the CNN network is a process to find appropriate parameters to minimize the error between the predicted results and the ground truth labels on a training dataset. The CNN converted each input patch from the original pixel values to the final probability classification results, and the parameters were calculated by a loss function through feed-forward propagation. The learnable parameters were updated according to the loss value by using the stochastic gradient descent based on the back-propagation algorithm.

The hyperparameters (Table 3) are the variables that need to be set before the training process begins. Dropout is a recently introduced regularization technique and the fully connected layers are followed by dropout rates of 0.5 to mitigate overfitting. Bergstra and Bengio (2012) argued that random searches are more efficient for hyperparameter optimization than grid searches and manual searches. A random search was used to optimize the hyperparameters and improve the accuracy and speed of the model. Adam, an efficient stochastic optimization algorithm based on the gradient (Kingma and Ba 2014), was selected as the optimizer.

3.4 Performance Evaluation

The evaluation criteria are a key factor in assessing the classification performance and guiding the classifier modeling (Sokolova and Lapalme 2009). In this article, a two-class classification method is modeled to predict forest fire susceptibility. Thus, five statistical measures including overall accuracy, specificity, sensitivity, positive predictive

value (PPV), and negative predictive value (NPV) are employed to appraise the classification capability (Tien Bui et al. 2017). The five statistical measures are computed in the following manner:

$$\text{Overall accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}};$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}; \quad \text{PPV} = \frac{\text{TP}}{\text{FP} + \text{TP}};$$

$$\text{NPV} = \frac{\text{TN}}{\text{FN} + \text{TN}}$$

where TP (true positive) and TN (true negative) are the number of samples that are correctly classified as positive (fire class) and negative (nonfire class) observations, respectively. FP (false positive) and FN (false negative) are the number of samples that are misclassified. Sensitivity is the percentage of positive (fire class) observations that are correctly classified whereas specificity is the percentage of negative (nonfire class) observations that are correctly identified.

The ROC curve has been increasingly utilized to evaluate and validate the global performance assessment of the prediction models in ML and data mining research (Pourtaghi et al. 2016; Satir et al. 2016). It depicts the trade-offs between the TPs and FPs rather than arbitrarily selecting a particular threshold (Freeman and Moisen 2008). A ROC plot is constructed by plotting the TP rate (TPrate, sensitivity) on the Y-axis against the FP rate (FPrate, 1.0—specificity) for all possible thresholds from 0 to 1 on the X-axis. The ROC plot for a good classifier tends to rise sharply at the origin and then level off near the maximum value of 1. A trivial classifier will result in a ROC graph near the diagonal where the TP rate is equal to the FP rate for all thresholds. The AUC is generally considered to be an important index to quantitatively assess the overall accuracy of the classifier's performance. An AUC value near 0.5 means that the predictive ability of the model is completely random and a value of 1.0 represents a perfect

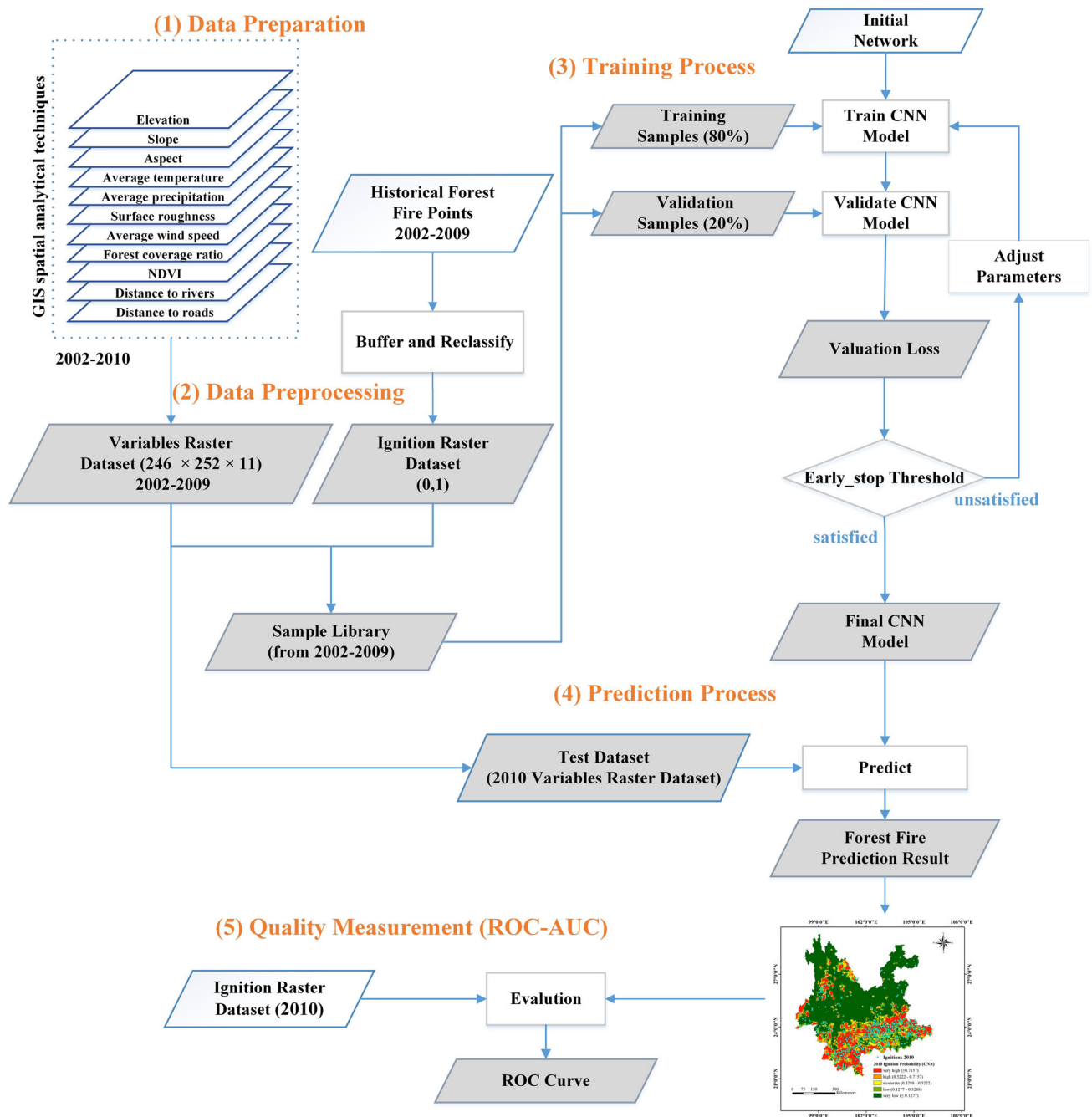


Fig. 4 Methodological flowchart employed in this study. *CNN* convolutional neural network; *ROC* receiver operation characteristic

prediction without misclassification. The closer the AUC value is to 1, the better the performance of the forest fire prediction model. The AUC measure is computed by obtaining only the area of the graphic:

$$AUC = \frac{1 + TP_{rate} - FP_{rate}}{2}$$

3.5 Technical Process for Predicting Forest Fire Susceptibility

The technical process can be summarized into the following five steps. The detailed procedure of this study is depicted in Fig. 4.

Step 1 Construct the fire and nonfire inventory maps and create maps of the influencing factors that can potentially influence the ignition susceptibility.

Table 4 Multicollinearity analysis of forest fire influencing factors

No.	Forest fire influencing factor	TOL	VIF
1	Elevation	0.288	3.47
2	Slope	0.819	1.221
3	Aspect	0.999	1.001
4	Average temperature	0.157	6.372
5	Average precipitation	0.4	2.5
6	Surface roughness	0.199	5.016
7	Average wind speed	0.106	9.443
8	Forest coverage ratio	0.881	1.134
9	NDVI	0.608	1.645
10	Distance to roads	0.948	1.055
11	Distance to rivers	0.932	1.073
12	Maximum temperature	0.014	69.653
13	Specific humidity	0.017	58.696
14	Precipitation rate	0.062	16.078

NDVI normalized difference vegetation index; *TOL* tolerances; *VIF* variance inflation factor

Step 2 Preprocess all datasets and generate the training/validation samples.

Step 3. Design the architecture of the CNN model, optimize the CNN hyperparameters, and train the CNN classifier.

Step 4 Predict forest fire susceptibility using the VRD of 2010.

Step 5 Evaluate the performance of the proposed model.

The CNN predication model was constructed under a graphics processing unit (GPU) acceleration environment using the Keras DL framework that uses TensorFlow as a backend, which is a Python-based DL library. The system configuration used in the lab environment is as follows:

Intel Core i7 CPU, 16 GB RAM, Windows10 OS, and an NVIDIA GeForce GTX 1070 with 12 GB of onboard memory.

4 Results

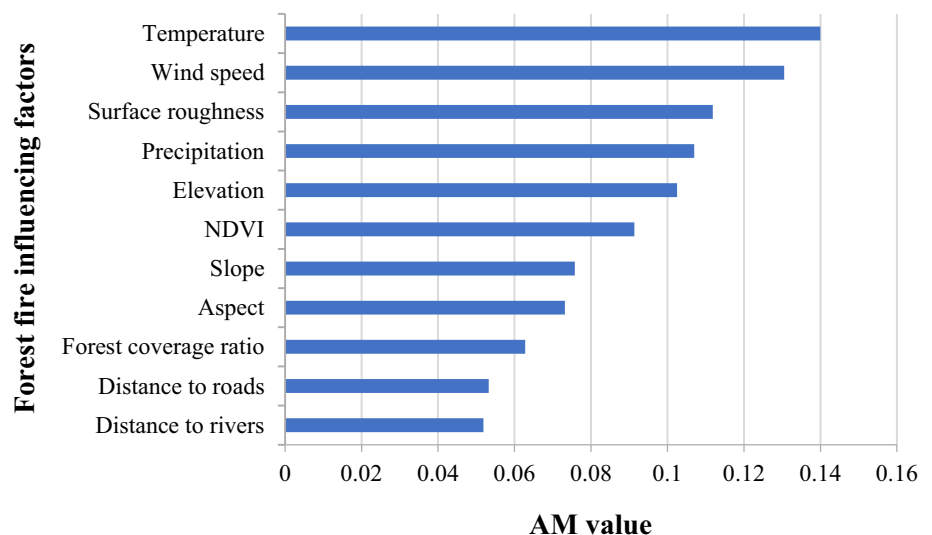
This section first shows the results of multicollinearity analysis and an information gain ratio (IGR) for the selection of forest fire influencing factors. The loss and the accuracy in the training/validation phases were tracked. Then, the test dataset was fed into the trained model and the prediction map of ignition probabilities was constructed by the CNN model. Finally, the performance of the proposed model was compared with benchmark methods.

4.1 Relative Importance Analysis of Influencing Factors

According to the results of a multicollinearity analysis of the 14 forest fire influencing factors in Table 4, three factors—precipitation rate, specific humidity, and maximum temperature—did not satisfy the critical values, suggesting the existence of multicollinearity and should be excluded from further analyses.

For the IGR method, the factors with a higher value of average merit (AM) indicate a stronger prediction ability of the model. However, factors with AM values equal to or less than 0 indicate a “null” contribution to the forest fire susceptibility model and should be excluded from further analysis (Bui, Tuan, et al. 2016). The results listed in Fig. 5 show that the AM values of all remaining 11 influencing factors are greater than 0, indicating that all these influencing factors contribute to the model and should be retained in the following prediction process. Temperature

Fig. 5 Average merit (AM) of each forest fire influencing factor using the information gain ratio (IGR) method



is the most important factor for forest fire susceptibility, with the highest AM value of (0.139). It is followed by wind speed (0.131), surface roughness (0.112), precipitation (0.107), and elevation (0.102).

4.2 Model Accuracy

The training process was divided into training and validation phases. The validation dataset was used to monitor the model classification performance after the end of each epoch, and the validation results were used as the basis for whether the training process should be terminated earlier or the hyperparameters should be fine-tuned. The loss and accuracy were two important indicators for evaluating the effect of model training. Callback functions were applied to adjust the training state and statistics during model training, including “EarlyStopping,” “ReduceLROnPlateau,” and “ModelCheckpoint.”

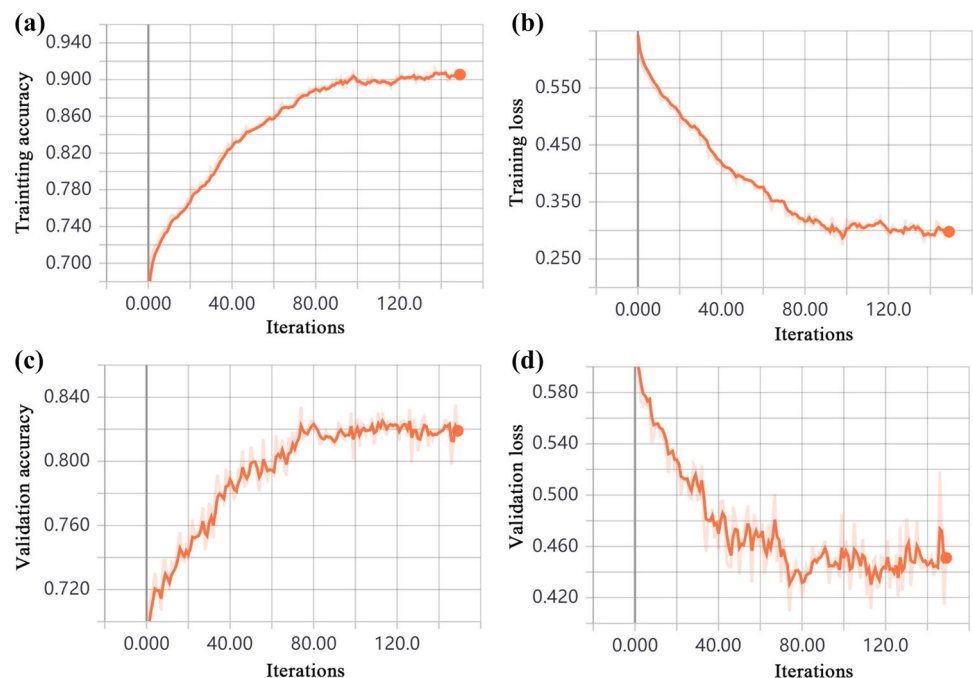
Figure 6 shows the loss and the training/validation accuracy using TensorBoard for visualization. After the training phase, the training accuracy was close to 91%, the training loss no longer decreased after the 100th epoch and tended to fit, and the minimum loss value was 0.3. After the 80th epoch, the validation loss no longer decreased and tended to converge. However, the validation loss rose slightly after the 120th epoch, suggesting overfitting may have occurred. Immediately, the EarlyStopping function ended the training process to decrease the phenomenon of overfitting. The validation accuracy of the model reached 82%, and the minimum validation loss was 0.45.

The slow decline curve of validation loss indicates that the model was well fitted. Overfitting did not occur because the training process was effectively terminated by the “EarlyStopping” function. After the training process, the model corresponding to the epoch with the lowest loss of the validating dataset in Check_pointer was selected as the final classification model.

4.3 Susceptibility Map

After the training process, a classification model was achieved. A test dataset—the VRD of 2010—was then used at the end of the project to evaluate the performance of the final model. The VRD of 2010 was fed into the classification model to predict forest fire susceptibility. The model resulted in probabilities for both the fire and nonfire classes on each pixel in the generated prediction map. Notably, although the CNN was designed to predict a single label from a small input patch of size $25 \times 25 \times 11$, the CNN was trained to predict all pixels in the new VRD since the sliding window was densely overlapped and covered the entire VRD during the reasoning phase. The sum of the probabilities of the fire and nonfire class was 1 on each pixel. The probability of the fire class was chosen as the final predicted probability value, and then we produced a map of the probability values for all pixels by converting the array data into an image using the libtiff package in Python. The forest fire susceptibility map was finally produced by dividing the image into five levels using the natural breaks method in ArcGIS10.5. Five

Fig. 6 a, b Represent the training accuracy and the convergence graph of the training loss, respectively; c, d represent the validation accuracy and the convergence graph of the validation loss, respectively



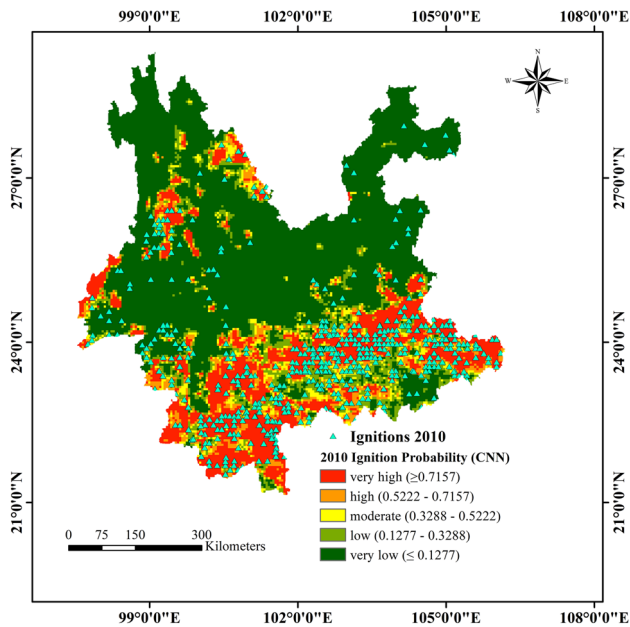


Fig. 7 Forest fire susceptibility map derived from the convolutional neural network (CNN) model for Yunnan Province, China, in 2010

susceptible classes were identified—very low, low, moderate, high, and very high—for constructing the forest fire susceptibility map (Fig. 7).

From the predicted susceptibility results, we can see that the highest forest fire ignition susceptibilities are mainly distributed in the south and northwest of Yunnan Province. The areas with the lowest forest fire ignition susceptibilities are in the central and northeastern parts of Yunnan.

4.4 Model Comparison

The usability of the proposed model was compared with benchmark methods random forests (RF), support vector machine (SVM), multilayer perceptron neural network (MLP), and kernel logistic regression (KLR). The four benchmark methods were built and implemented using the scikit-learn package, and the grid search method was used to optimize the hyperparameters. The main hyperparameters utilized in the benchmark methods are listed in Table 5. Recursive feature elimination with cross-validation was employed to perform automatic tuning of the number of features selected for SVM and KLR.

The performance of the five models was evaluated and compared for both the training and validation datasets of 2009. Table 6 shows that the CNN had the higher specificity (93.77%), sensitivity (97.84%), PPV (94.02%), NPV (97.75%), and overall accuracy (95.81%) for the training dataset than the four benchmark methods. The proposed CNN model had the highest overall accuracy (87.92%) for

the validation dataset, followed by the RF (84.36%), KLR (81.23%), SVM (80.04%), and MLP (78.47%).

Figure 8 illustrates the ignition susceptibility map constructed by the benchmark models using the same test datasets. The values denote the probabilities of each pixel having an ignition occurrence. To enhance the comparability of the prediction results for the CNN and benchmark models, the probability map was divided into five classes with the same natural breaks method as CNN. It can be observed that in the prediction maps of SVM, MLP, and KLR, forest fire susceptibility of most areas in the northwest of Yunnan Province was categorized as very low and low. In contrast, CNN and RF can well-identify the ignition probabilities in the northwest. For the southern region of Yunnan Province, compared with the predicted results of the CNN model, the RF model divided most regions of southern Yunnan into very high susceptible zones, while the results of the SVM, MLP, and KLR models predicted high susceptibility only in the southwest region of Yunnan Province. The high fire occurrence region in the southeast was not clearly identified in the predicted results.

Figure 9 shows that the very high and very low susceptibility classes in the CNN model account for 77.51% of the total area, which was the highest proportion among all the models; whereas the remaining three classes of high, moderate, and low have the lowest proportion of all models, accounting for 7.15%, 6.33%, and 9.01% of the total area, respectively. The results show that the proposed CNN model can effectively divide the very high and very low susceptible zones. However, the probability prediction results of the benchmark models had many areas within medium and high susceptibility zones, and there was no clear determination regarding zones with high forest fire susceptibility; thus, what threshold segmentation method should be used to divide the fire warning areas needs to be considered. Most traditional ML methods will face this problem after obtaining the probability prediction results. In contrast, the probability results obtained by the CNN are more distinct in their spatial pattern, and the statistical values reflect a high and low bipolar distribution, which is more advantageous in the division of the areas with fire warnings and those without fire warnings and can reduce the influence of the threshold segmentation of the delineation of fire warning areas. Thus, the result of the CNN model is better for forest fire predictions.

To confirm the statistical significance of the prediction performance between the proposed CNN model and the benchmark models, the nonparametric WSRT (Wilcoxon 1945) was employed for paired comparisons. The null-hypothesis (H_0) was that there was no significant difference at 95% confidence intervals of two prediction models. If the p value was less than the significance level (0.05), H_0 was rejected, and a significant difference exists in the

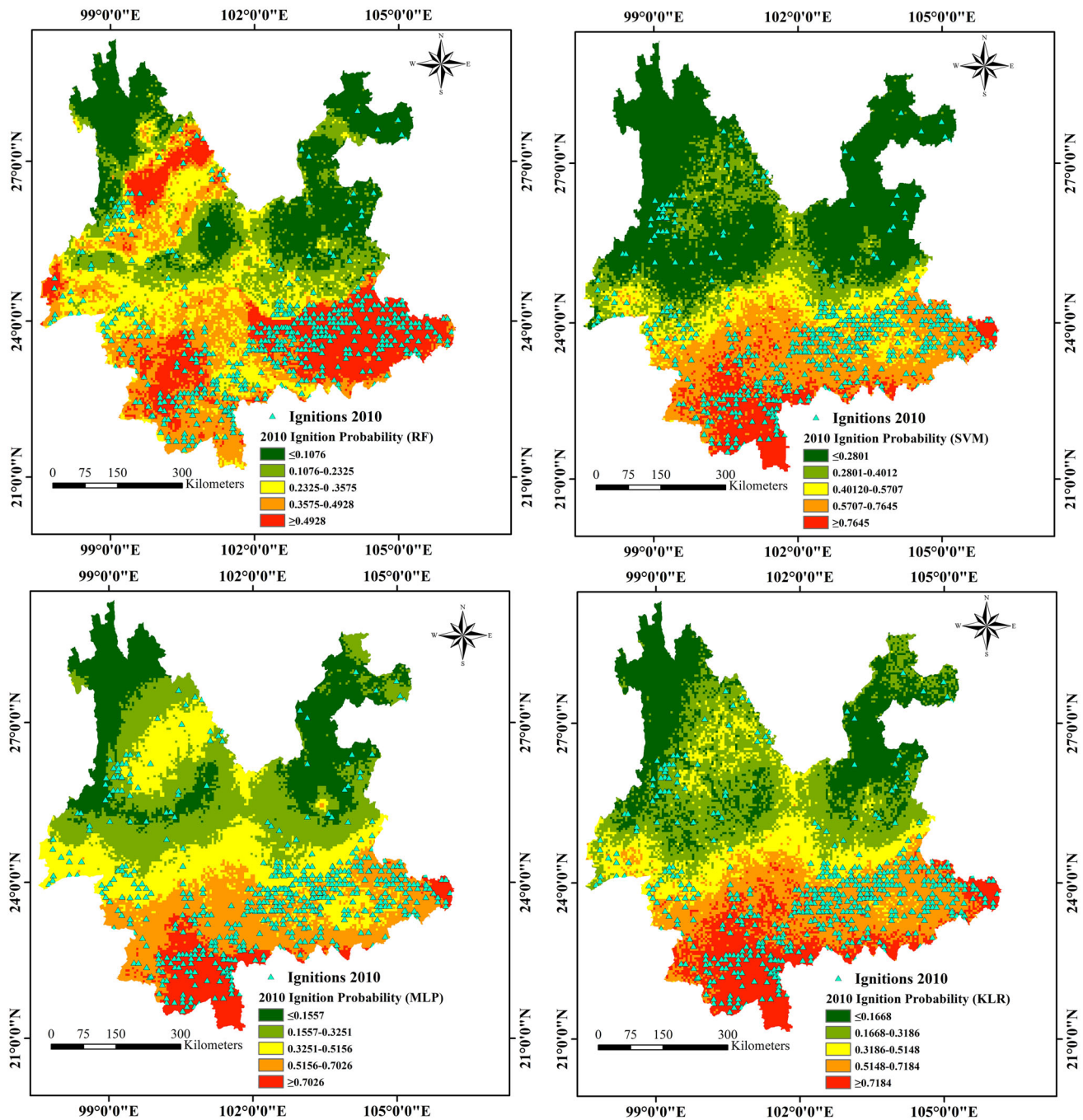


Fig. 8 Prediction maps of the ignition probabilities for the benchmark models for Yunnan Province, China. *KLR* kernel logistic regression; *MLP* multilayer perceptron neural network; *RF* random forests; *SVM* support vector machine

models (Tien Bui et al. 2019). The analysis was performed using SPSS (Statistical Package for the Social Sciences), and the Wilcoxon signed-rank test (WSRT) results are reported in Table 7. The p value of all pairwise comparisons was less than 0.05, confirming that the classification performances of the proposed model and benchmark models are significantly different.

The ROC curves for the five models are depicted in Fig. 10. They were drawn using the 2010 prediction maps of the ignition probabilities and the corresponding IRD. The AUC provides a single measure of a classifier's performance to evaluate which model is better on average. The AUC of the CNN model is 0.86 (Fig. 10), indicating that the global fit of the model with the testing dataset is 86%, followed by RF (0.82), SVM (0.79), MLP (0.78), and

Table 5 A list of the main hyperparameters utilized in the benchmark methods

Benchmark methods	Hyperparameters
RF	Number of trees: 160; max_features: "sqrt"; bootstrap: True; max_depth=20
SVM	Penalty factor (C): 100; kernel function="RBF"; gamma=1
MLP	Number of hidden layers: 1; momentum: 0.2; learning rate: 0.001; iteration=300; alpha=0.01; solver="Adam"; activation="ReLU"
KLR	Kernel function="RBF"; tuning parameter (δ): 0.02; regularize parameter (C): 0.025

RF random forests; SVM support vector machine; MLP multilayer perceptron neural network; KLR kernel logistic regression; RBF radial basis function; ReLU rectified linear unit

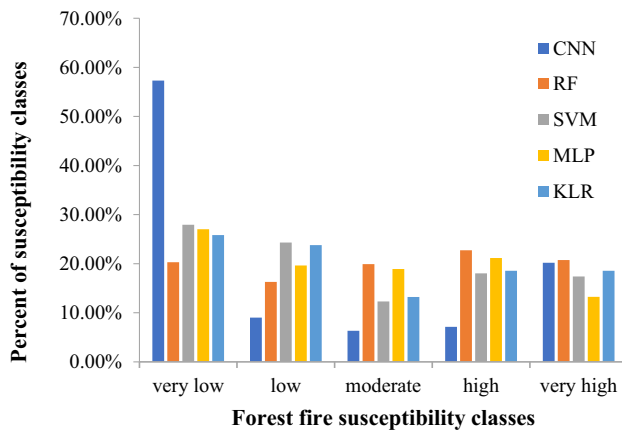


Fig. 9 Percentages of different forest fire susceptibility classes. CNN convolutional neural network; RF random forests; SVM support vector machine; MLP multilayer perceptron neural network; KLR kernel logistic regression

KLR (0.78). The ROC curves clearly show that the CNN model has the highest prediction performance.

5 Discussion

This section first describes the advantages of CNN compared with benchmark methods, then discusses the selection of the model architecture and window size, and finally discusses the differences between the CNN and traditional ML algorithms.

5.1 Advantage of the Convolutional Neural Network Method

Compared with the benchmark methods, the CNN has the following advantages. First, because the CNN can consider the correlation of adjacent spatial information, it has advantages in the study of problems with spatial and geographical correlation characteristics. Second, the CNN preserves the spatial relationships between pixels by learning the internal feature representations from factor

vectors. The process of DL reveals the deep features and can distinguish the differences between different geographical units. The CNN was used to conduct multiple convolution and pooling operations to extract the characteristics. As the convolutions and pooling increased, these features became more advanced and more abstract. These abstract features depicted the degree of forest fire susceptibility, which was the decisive factor for determining forest fire susceptibility. Third, the CNN reduces the number of weights that need to be trained and the computational complexity of the network through weight sharing.

5.2 Model Sensitivity

The architecture of the CNN model should be selected in accordance with the quantity of sample data and the problem complexity. For those with small quantities of sample data and simple classification problems, the complex structure was prone to model overfitting. For those with large quantities of sample data and complex classification problems, the simple structure was prone to model nonconvergence. Both of these problems should be avoided as much as possible in the training of the CNN model.

The selection of the window size should be consistent with the maximum geospatial area that affects the forest fire susceptibility of the center window pixel. A larger window size means that the pixels of a larger geographical unit impact the center window pixel. In fact, the pixels that affect the fire susceptibility of the center window pixel have a certain geographical spatial range. Therefore, the selection of the window size must be appropriate. After the experiments, the 25×25 window size was most reasonable. First, this size corresponds to the maximum geographical spatial range that affects the fire susceptibility of a pixel. Second, the size of this window is smaller than that of the commonly used windows (for example, 224×224) in image processing because forest fire probability prediction and image classification are completely different

Table 6 Training and validation performance comparison

Phase	Performance	Prediction model				
		CNN	RF	SVM	MLP	KLR
Training	TP	3126	3053	2860	2800	2940
	TN	2996	2796	2322	2349	2372
	FP	199	399	873	846	823
	FN	69	142	335	395	345
	Sensitivity (%)	97.84	95.56	89.51	87.64	89.5
	Specificity (%)	93.77	87.51	72.68	73.52	74.24
	PPV (%)	94.02	88.44	76.61	76.8	78.13
	NPV (%)	97.75	95.17	87.39	85.6	87.3
	Overall accuracy (%)	95.81	91.53	81.1	80.58	81.98
Validation	TP	739	580	532	491	577
	TN	666	768	747	763	721
	FP	133	31	52	36	78
	FN	60	219	267	308	222
	Sensitivity (%)	92.49	72.59	66.58	61.45	72.22
	Specificity (%)	83.35	96.12	93.49	95.49	90.24
	PPV (%)	84.75	94.93	91.1	93.17	88.09
	NPV (%)	91.74	77.81	73.67	71.24	76.46
	Overall accuracy (%)	87.92	84.36	80.04	78.47	81.23

CNN convolutional neural network; *FP* false positive; *FN* false negative; *KLR* kernel logistic regression; *MLP* multilayer perceptron neural network; *NPV* negative predictive value; *PPV* positive predictive value; *RF* random forests; *SVM* support vector machine; *TN* true negative; *TP* true positive

Table 7 Wilcoxon signed-rank test (two-tailed)

No.	Pairwise comparison	<i>p</i> value	Significance
1	CNN versus RF	< 0.0001	Yes
2	CNN versus SLM	< 0.0001	Yes
3	CNN versus MLP	< 0.0001	Yes
4	CNN versus KLR	< 0.0001	Yes

CNN convolutional neural network; *RF* random forests; *SVM* support vector machine; *MLP* multilayer perceptron neural network; *KLR* kernel logistic regression

applications, and the CNN model that is used for image processing cannot be completely duplicated.

It is necessary to discuss the differences between the CNN and traditional ML algorithms. The first is the dependency of the CNN model on the data and hardware. The CNN requires a large number of training samples, and its performance increases as the scale of data increases. Because the CNN inherently performs a large number of matrix multiplication operations, they rely heavily on high-end machines compared to traditional ML algorithms that can run on low-end machines. The second is that CNN can automatically explore high-level features from raw data.

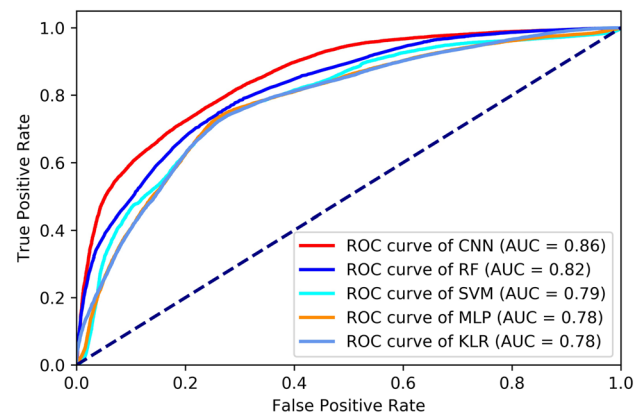


Fig. 10 Receiver operating characteristic (ROC) curves and area under the curve (AUCs) of the five models. *CNN* convolutional neural network; *RF* random forests; *SVM* support vector machine; *MLP* multilayer perceptron neural network; *KLR* kernel logistic regression

This is a very distinctive part of DL and a major step ahead of traditional ML. Moreover, the CNN model shows a strong generalization ability compared with the benchmark methods. In terms of time efficiency, because the CNN has a large number of parameters to learn, the training time of the CNN is longer than those of traditional ML models. Because of the reusability of the CNN neural network after

initial training, there is still considerable room for improving the efficiency in the later training time. The prediction time of the CNN is relatively short when using GPU-accelerated computing technology. Finally, although CNN has shown excellent performance, the interpretability is its deficiency, which needs to be further studied.

6 Conclusion

In this article, we investigated a CNN with deep architectures for the spatial prediction of forest fire susceptibility in Yunnan Province, China. Past forest fire locations from 2002 to 2010 were extracted and a set of 14 forest fire influencing factors were optimized using multicollinearity analysis and the IGR technique. We explored the preprocessing methods for forest fire influencing factors and the methods for generating effective training/validation sample libraries. The CNN architecture suitable for the prediction of forest fire susceptibility in the study area was designed, and hyperparameters were optimized to improve the prediction accuracy. Several common methods, such as more training samples, regularization (dropout), batch normalization, and reduced architecture complexity, were used in the CNN model to mitigate overfitting. Then, the test dataset was fed into the trained model and the prediction map of ignition probabilities was constructed by the CNN model. Finally, the performance of the proposed model was compared with traditional ML methods using several statistical measures, including WSRT, ROC, and AUC.

Through this research, we found that the CNN model performs better than the benchmark methods. The CNN model (AUC=0.86) has higher predictive power than the benchmark methods according to the ROC–AUC. The probability result obtained by the CNN can clearly distinguish the very high and very low susceptible zones, and the susceptibility spatial pattern was very distinct. The CNN model shows a strong generalization ability and the prediction time of the CNN was relatively short when using GPU-accelerated computing technology. In conclusion, the CNN has the advantages of considering neighborhood information, extracting deep features, sharing weights, and pooling operations, which allow the CNN to obtain better prediction results. The CNN model will have important practical application value for forest fire prevention planning and forest management.

There are still some limitations in the research. For example, the influence of different CNN architectures—such as VGG-net (Visual Geometry Group Network), RES-net Residential Energy Services Network), and GoogLeNet—on forest fire prediction results have not been studied in depth. In addition, more actual data are needed for the experimental verification of the method. In recent years,

the application of CNNs has become increasingly extensive. Many different variants of the architecture have been derived and many ensemble classifiers have been proposed. Comparing various classifiers and exploring the most suitable models to improve forest fire prediction should be investigated in the future.

Acknowledgements This research was supported by the National Key Research and Development Plan (2017YFC1502902) and National Natural Science Foundation of China (41621601). The financial support is highly appreciated. We thank Yinxue Cao for her help in getting the data from Climate Forecast System Reanalysis. We are also grateful to the anonymous reviewers and the editors for their constructive comments.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Adab, H., K.D. Kanniah, and K. Solaimani. 2013. Modeling forest fire risk in the northeast of Iran using remote sensing and GIS techniques. *Natural Hazards* 65(3): 1723–1743.
- Arpaci, A., B. Malowerschnig, O. Sass, and H. Vacik. 2014. Using multi variate data mining techniques for estimating fire susceptibility of Tyrolean forests. *Applied Geography* 53: 258–270.
- Bajocco, S., E. Dragoz, I. Gitas, D. Smiraglia, L. Salvati, and C. Ricotta. 2015. Mapping forest fuels through vegetation phenology: The role of coarse-resolution satellite time-series. *PLoS ONE* 10(3): 1–14.
- Bar Massada, A., A.D. Syphard, S., I. Stewart, and V.C. Radeloff. 2013. Wildfire ignition-distribution modelling: a comparative study in the Huron–Manistee National Forest, Michigan, USA. *International Journal of Wildland Fire* 22(2): 174–183.
- Bergstra, J., and Y. Bengio. 2012. Random search for hyperparameter optimization. *Journal of Machine Learning Research* 13(1): 281–305.
- Bisquert, M., E. Caselles, J.M. Sánchez, and V. Caselles. 2012. 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* 21(8): 1025–1029.
- Buda, M., A. Maki, and M.A. Mazurowski. 2018. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks* 106: 249–259.
- Bui, D.T., K.T.T. Le, V.C. Nguyen, H.D. Le, and I. Revhaug. 2016. Tropical forest fire susceptibility mapping at the Cat Ba National Park area, Hai Phong City, Vietnam, using GIS-based Kernel logistic regression. *Remote Sensing* 8(4): 1–15.
- Bui, D.T., T.A. Tuan, H. Klempe, B. Pradhan, and I. Revhaug. 2016. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* 13(2): 361–378.
- Cao, Y., M. Wang, and K. Liu. 2017. Wildfire susceptibility assessment in Southern China: A comparison of multiple methods. *International Journal of Disaster Risk Science* 8(2): 164–181.

- Colkesen, I., E.K. Sahin, and T. Kavzoglu. 2016. Susceptibility mapping of shallow landslides using kernel-based Gaussian process, support vector machines and logistic regression. *Journal of African Earth Sciences* 118: 53–64.
- Crimmins, M.A. 2006. Synoptic climatology of extreme fire-weather conditions across the southwest United States. *International Journal of Climatology* 26(8): 1001–1016.
- Dash, M., and H. Liu. 1997. Feature selection for classification. *Intelligent Data Analysis* 1(1–4): 131–156.
- Dimuccio, L.A., R. Ferreira, L. Cunha, and A. Campar de Almeida. 2011. Regional forest-fire susceptibility analysis in central Portugal using a probabilistic ratings procedure and artificial neural network weights assignment. *International Journal of Wildland Fire* 20(6): 776–791.
- Elmas, Ç., and Y. Sönmez. 2011. A data fusion framework with novel hybrid algorithm for multi-agent Decision Support System for Forest Fire. *Expert Systems with Applications* 38(8): 9225–9236.
- Freeman, E.A., and G.G. Moisen. 2008. A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecological Modelling* 217(1): 48–58.
- Guo, F., Z. Su, G. Wang, L. Sun, F. Lin, and A. Liu. 2016. Wildfire ignition in the forests of southeast China: Identifying drivers and spatial distribution to predict wildfire likelihood. *Applied Geography* 66: 12–21.
- Hantson, S., S. Pueyo, and E. Chuvieco. 2015. Global fire size distribution is driven by human impact and climate. *Global Ecology and Biogeography* 24(1): 77–86.
- Hinton, G.E., and R.R. Salakhutdinov. 2006. Reducing the dimensionality of data with neural networks. *Science* 313(5786): 504–507.
- Hong, H., A. Jaafari, and E.K. Zenner. 2019. Predicting spatial patterns of wildfire susceptibility in the Huichang County, China: An integrated model to analysis of landscape indicators. *Ecological Indicators* 101: 878–891.
- Hong, H., B. Pradhan, C. Xu, and D. Tien Bui. 2015. Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. *Catena* 133: 266–281.
- Hong, H., P. Tsangaratos, I. Ilia, J. Liu, A.X. Zhu, and C. Xu. 2018. 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* 630: 1044–1056.
- Hu, F., G.S. Xia, J. Hu, and L. Zhang. 2015. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sensing* 7(11): 14680–14707.
- Jaafari, A., E.K. Zenner, M. Panahi, and H. Shahabi. 2019. Hybrid artificial intelligence models based on a neuro-fuzzy system and metaheuristic optimization algorithms for spatial prediction of wildfire probability. *Agricultural and Forest Meteorology* 266–267: 198–207.
- Jaafari, A., E.K. Zenner, and B.T. Pham. 2018. Wildfire spatial pattern analysis in the Zagros Mountains, Iran: A comparative study of decision tree based classifiers. *Ecological Informatics* 43: 200–211.
- Kim, S.J., C.H. Lim, G.S. Kim, J. Lee, T. Geiger, O. Rahmati, Y. Son, and W.K. Lee. 2019. Multi-temporal analysis of forest fire probability using socio-economic and environmental variables. *Remote Sensing* 11(1): Article 86.
- Kingma, D.P., and J. Ba. 2014. Adam: A method for stochastic optimization. Presented as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015. arXiv preprint abs/1412.6980. Ithaca, NY: Cornell University.
- Krizhevsky, A., I. Sutskever, G.E. Hinton. 2013. ImageNet classification with deep convolutional neural networks. In *Proceedings of the 26th Annual Conference on Neural Information Processing Systems (NIPS)*, 3–6 December 2012, Lake Tahoe, Nevada, USA, ed. F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, Vol. 2, 1097–1105.
- Lecun, Y., Y. Bengio, and G. Hinton. 2015. Deep learning. *Nature* 521(7553): 436–444.
- Lecun, Y., L. Bottou, Y. Bengio, and P. Haffner. 1998. Gradient-based learning applied to document recognition. *Proceedings of IEEE* 86(11): 2278–2324.
- Leuenberger, M., J. Parente, M. Tonini, M.G. Pereira, and M. Kanevski. 2018. Wildfire susceptibility mapping: Deterministic vs. stochastic approaches. *Environmental Modelling and Software* 101: 194–203.
- Liu, T., and A. Abd-Elrahman. 2018. Deep convolutional neural network training enrichment using multi-view object-based analysis of Unmanned Aerial systems imagery for wetlands classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 139: 154–170.
- López, V., A. Fernández, S. García, V. Palade, and F. Herrera. 2013. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences* 250: 113–141.
- Moritz, M. A., M.-A. Parisien, E. Batllori, M. A. Krawchuk, J. Van Dorn, D.J. Ganz, and K. Hayhoe. 2012. Climate change and disruptions to global fire activity. *Ecosphere* 3(6): Article 49.
- Muhammad, K., J. Ahmad, and S.W. Baik. 2018. Early fire detection using convolutional neural networks during surveillance for effective disaster management. *Neurocomputing* 288(C): 30–42.
- Ngoc Thach, N., D. Bao-Toan Ngo, P. Xuan-Canh, N. Hong-Thi, B. Hang Thi, H. Nhat-Duc, and T.B. Dieu. 2018. Spatial pattern assessment of tropical forest fire danger at Thuan Chau area (Vietnam) using GIS-based advanced machine learning algorithms: A comparative study. *Ecological Informatics* 46: 74–85.
- O'Brien, R.M. 2007. A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity* 41(5): 673–690.
- Oliveira, S., F. Oehler, J. San-Miguel-Ayaz, A. Camia, and J.M.C. Pereira. 2012. Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *Forest Ecology and Management* 275: 117–129.
- Pew, K.L., and C.P.S. Larsen. 2001. GIS analysis of spatial and temporal patterns of human-caused wildfires in the temperate rain forest of Vancouver Island, Canada. *Forest Ecology and Management* 140(1): 1–18.
- Pourtaghi, Z.S., H.R. Pourghasemi, R. Aretano, and T. Semeraro. 2016. Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques. *Ecological Indicators* 64: 72–84.
- Renard, Q., R. Plissier, B.R. Ramesh, and N. Kodandapani. 2012. Environmental susceptibility model for predicting forest fire occurrence in the Western Ghats of India. *International Journal of Wildland Fire* 21(4): 368–379.
- Running, S.W. 2006. Is global warming causing more, larger wildfires? *Science* 313(5789): 927–928.
- Sachdeva, S., T. Bhatia, and A.K. Verma. 2018. GIS-based evolutionary optimized Gradient Boosted Decision Trees for forest fire susceptibility mapping. *Natural Hazards* 92(3): 1399–1418.
- Satir, O., S. Berberoglu, and C. Donmez. 2016. Mapping regional forest fire probability using artificial neural network model in a Mediterranean forest ecosystem. *Geomatics, Natural Hazards and Risk* 7(5): 1645–1658.
- Schmidhuber, J. 2015. Deep Learning in neural networks: An overview. *Neural Networks* 61: 85–117.

- Sokolova, M., and G. Lapalme. 2009. A systematic analysis of performance measures for classification tasks. *Information Processing and Management* 45(4): 427–437.
- Tien Bui, D., Q.T. Bui, Q.P. Nguyen, B. Pradhan, H. Nampak, and P. T. Trinh. 2017. 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* 233: 32–44.
- Tien Bui, D., N.D. Hoang, and P. Samui. 2019. Spatial pattern analysis and prediction of forest fire using new machine learning approach of Multivariate Adaptive Regression Splines and Differential Flower Pollination optimization: A case study at Lao Cai province (Viet Nam). *Journal of Environmental Management* 237: 476–487.
- Tobler, W.R. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46(sup1): 234–240.
- Vasconcelos, M.J. P. de, S. Silva, M. Tomé, M. Alvim, and J. M. C. Perelra. 2001. Spatial Prediction of Fire Ignition Probabilities: Comparing Logistic Regression and Neural Networks. *Photogrammetric Engineering & Remote Sensing* 67(1): 73–81.
- Vetrivel, A., M. Gerke, N. Kerle, F. Nex, and G. Vosselman. 2018. Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning. *ISPRS Journal of Photogrammetry and Remote Sensing* 140: 45–59.
- Wang, Y., Z. Fang, and H. Hong. 2019. Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China. *Science of the Total Environment* 666: 975–993.
- Wilcoxon, F. 1945. Individual comparisons by ranking methods. *Biometrics Bulletin* 1(6):80–83.
- Yamashita, R., M. Nishio, R. K.G. Do, and K. Togashi. 2018. Convolutional neural networks: An overview and application in radiology. *Insights into Imaging* 9(4): 611–629.
- Yi, K., H. Tani, J. Zhang, M. Guo, X. Wang, and G. Zhong. 2013. Long-term satellite detection of post-fire vegetation trends in boreal forests of China. *Remote Sensing* 5(12): 6938–6957.
- Ying, L., J. Han, Y. Du, and Z. Shen. 2018. Forest fire characteristics in China: Spatial patterns and determinants with thresholds. *Forest Ecology and Management* 424: 345–354.
- Zhang, X. 2007. Vegetation map of the People's Republic of China (1:1000000). Beijing: Geology Press (**in Chinese**).
- Zhang, C., X. Pan, H. Li, A. Gardiner, I. Sargent, J. Hare, and P.M. Atkinson. 2018. A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 140: 133–144.