IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 21, 2019, accepted December 2, 2019, date of publication December 5, 2019,
date of current version December 23, 2019.

*Digital Object Identifier 10.1109/ACCESS.2019.2957837*

# A Neural Network Model for Wildfire Scale Prediction Using Meteorological Factors

**HAO LIANG, MENG ZHANG, AND HAILAN WANG**
School of Technology, Beijing Forestry University, Beijing 100083, China
Beijing Laboratory of Urban and Rural Ecological Environment, Beijing 100083, China
Key Lab of State Forestry Administration for Forestry Equipment and Automation, Beijing 10083, China

Corresponding author: Hailan Wang (wanghailan@bjfu.edu.cn)

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant BLX201717 and in part by the National Key Research and Development Program of China under Grant 2017YFD0600901.

**ABSTRACT** A forest fire is a natural disaster that destroys forest resources, thus having a severe impact on humans and on the animals and plants that depend on the forest environment. This paper presents a model for predicting the scale of forest wildfires of Alberta, Canada. A fire's scale is determined by the combination of the fire's duration and the size of the area it burns. Our prediction model enables fire rescuers to take appropriate measures to minimize damage caused by a wildfire based on its predicted scale in the fire's early stages. The modeling data were collected from the Canada National Fire Database (CNFDB) published by Natural Resources Canada, which includes wildfire and meteorological data for Alberta, Canada. The size of the burned area and the fire's duration were used to estimate the scale of a wildfire. After multicollinearity testing and feature normalization, the data were divided into training and testing sets. Taking the meteorological factors as input values, a backpropagation neural network (BPNN), a recurrent neural network (RNN), and long short-term memory (LSTM) were implemented to establish prediction models. Of these classification methods, LSTM exhibited the highest accuracy, 90.9%. The results indicate that it is feasible to predict the scale of a forest wildfire at the beginning of its occurrence using meteorological information.

**INDEX TERMS** Forest wildfire, LSTM model, meteorological factors, prediction of wildfire scale.

## I. INTRODUCTION

Forest resources are some of the most important natural resources on Earth. Not only are they the main component of the Earth's ecosystem, they also provide many resources for human life and production [1]. Forest fires are one of the primary natural disasters that destroy forest resources. The losses caused by forest fires are enormous [2] and the challenges of fighting forest fires cannot be underestimated. The causes of forest fires are complex. Whereas most forest fires with human-based causes such as smoking, hunting, and logging can be prevented or reduced by various interventions [3], it is difficult to monitor or control forest wildfires that occur naturally under certain meteorological conditions. Wildfires are destructive and spread rapidly [4]. Based on data from Canada's National Forestry Database, over 8,000 forest fires each year burn an average of over 2.1 million ha.

The associate editor coordinating the review of this manuscript and approving it for publication was Jenny Mahoney.

Forest wildfires are a key environmental issue. Therefore, it is important to take effective preventive measures in forest management to reduce the losses of human life and property caused by forest fires. In this context, a wildfire scale prediction model would be useful in guiding forest fire rescue work and understanding their risk.

Because forest fires are a global problem, researchers in various countries have long been committed to their study. Recent years have seen numerous technical developments in the field that are aimed at improving fire prediction system design. Wotton *et al.* used Poisson regression analysis methods to predict the number of human-caused forest fires in the forested area of Ontario, Canada, finding that the total number of forest fires caused by human activities in Ontario could increase by approximately 18% by 2020–2040 and by 50% by the end of the 21st century [5]. Marchal *et al.* collected all forest wildfire data for broad-leaved forests in southern Quebec, Canada, from 2000 to 2010, analyzed the relationship between weather conditions, road density, land cover, and fire occurrence, and established a Poisson

regression model to predict the occurrence of forest wildfires in this area. Their results showed that road density exerted the strongest influence on the frequency of human-caused fires [6]. Plucinski *et al.* studied wildfires that occurred in ten managed forest areas in Western Australia. They developed a model using negative binomial regression with a dataset covering three years and evaluated it using data from an independent year. The methods they demonstrated can be used to develop predictive tools for day-to-day fire planning operations in other regions with suitable fire incident and meteorological records [7].

The study of forest fire prediction has focused mainly on the prediction of forest fire frequency, meaning that these models simply predict how many fires will occur [8], [9]. Information about the scale of fires is not included. If the scale of forest fires can be predicted in their early stages, relevant rescue and firefighting plans can be made in time to be most effective, thereby reducing the damage caused [10]. Therefore, the ability to predict the scale of forest fires is crucial. Regression models have been widely used in wildfire probability modelling since the 1990s [11]. Traditional models implemented for predicting forest fires include generalized linear models based on logistic, Poisson, and negative binomial distributions [12], [13]. However, these models cannot process multi-dimensional big data, which restricts their generalizability. Very recent comparative studies have shown that traditional regression models fail to accurately estimate the patterns of wildfire probabilities because they assume a linear relationship between a fire's occurrence and its causative factors [14], [15] when, in fact, the drivers of wildfires act nonlinearly within a broad range of spatiotemporal scales. Therefore, nonlinear models are needed to handle the complexities of the underlying processes.

In recent years, artificial intelligence (AI) methods have proven to be very effective for predicting natural hazards [16], [17]. Furthermore, AI methods have frequently been used in the context of wildfire modeling and have outperformed conventional statistical methods in many cases [14], [18]–[20]. In our study, which used time series data to predict forest wildfires, we implemented a long short-term memory (LSTM) model to predict the scale of forest wildfires. The LSTM model is a special type of recurrent neural network (RNN) that preserves historical information in data using a selective internal memory unit [21]. LSTM has shown to be more effective for analyzing time series prediction problems than other AI methods [22]–[24]. Since fire occurrences have obvious rules and trends and LSTM has advantages in predicting occurrence trends, an LSTM model was employed to predict the scale of wildfires.

The main objectives of this study are (1) to propose a new definition of wildfire scale that defined by the combination of a fire's duration and the size of the area it burns, (2) to investigate the capability of three neural network models (i.e., backpropagation neural network (BPNN), RNN, and LSTM models) to predict the scale of wildfires in forests using meteorological data, and (3) to provide an efficient estimate of future wildfires' scales. The application of these models is underpinned by real-world data from a fire-prone landscape in Alberta, Canada. Our model predicts the scale of forest wildfires in their early stages; our findings can provide new insights to guide authorities in taking effective and prompt fire management measures with a high level of accuracy in reasonable computation time.

**FIGURE 1.** Location of the study area, Alberta, Canada.

## II. STUDY AREA AND METHODOLOGY

### A. STUDY AREA

The study region, Alberta, located in western Canada, is illustrated in Fig. 1. Alberta has a humid continental climate with warm summers and cold winters. Regional average temperatures range from −15°C in winter to 24.5°C in summer and the annual average temperature is 2.5°C. Average seasonal precipitation ranges from 200 mm to 325 mm in winter and from 150 mm to 275 mm in summer and the annual precipitation is 300 mm to 600 mm. Alberta is a sunny province. Annual bright sunshine totals range between 1,900 h and 2,600 h per year. Most of the northern half of the province is boreal forest and the Rocky Mountains along its southwestern border are largely forested [25].

### B. DATA PREPARATION

The fire occurrence records available for the study area were obtained from the Canada National Fire Database (CNFDB). This database contains the coordinates of the locations and the ignition dates of all fires that occurred in the Alberta region between 1990 and 2018, inclusive. The database also records specific information about the forest fires, including their latitude and longitude, date of ignition, date of extinction, burned area, and cause. We also retrieved weather data from the CNFDB for the period 1990 to 2018; these were collected from 6,970 weather stations. The data contained 11 meteorological elements: maximum temperature (°C),

minimum temperature (°C), mean temperature (°C), heating degree days, cooling degree days, total rain (mm), total snow (mm), total precipitation (mm), snow on ground (cm), direction of maximum wind gust (tens of degrees), and speed of maximum wind gust (km/h). Because of the large quantity of meteorological data and the difficulty of obtaining them from web pages, web crawler technology was used to extract the meteorological data.

### C. DATA PRETREATMENT

#### 1) DATA FILTERING AND MATCHING

A total of 377,719 fire data records were obtained. However, the data records did not cover all 11 meteorological elements; therefore, we eliminated the records that were lacking any specific information relevant to the forest fire. Then, using the forest fire type, human-caused fires (such as those caused by smoking or hunting) were deleted. The remaining wildfire data consisted of 35,685 records available for modeling. Next, using their latitude and longitude, we matched the ignition points with the weather stations that had recorded meteorological data for the date of each wildfire. In all, 394,366 meteorological data records were selected for the years 1990 to 2018.

A full record for modeling needed to include information about the wildfire and corresponding meteorological information. Therefore, the fire data records were matched with the meteorological data records using the latitude and longitude of the ignition points. The final dataset for modelling consisted of 24,108 full records (or samples). In subsequent experiments, we used these records as samples for model training.

#### 2) MULTI-COLLINEARITY TEST

Multi-collinearity refers to strong correlations in the relationships among the explanatory variables in the regression model, which distort the estimations made by the model and can cause deviations from the ground truth [26]. To avoid having variables with significant collinearity affect the accuracy of our forest wildfire scale prediction model, we calculated the variance inflation factor (VIF) [27] to evaluate the collinearity between variables. It is generally believed that a variable whose VIF is greater than 10 should be eliminated because such a values indicates that significant collinearity between independent variables exists [28].

#### 3) FEATURE NORMALIZATION

Feature normalization independently normalizes each variable to a given range by a selected normalization method. In this study, the values of the input variables in the dataset varied widely, which led the gradient descent process used to find optimal solutions to be complex and time consuming and, further, affected prediction accuracy. Therefore, we introduced a feature normalization method to transform the variables into a spatial category that gave the input variables equal weight. The output variables with wildfire information were also normalized. The Min-Max Scaling algorithm [29] was used to quantify variables into the [0,1] range. The formula for feature quantization is given by Eq. (1):

$$z = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)}, \tag{1}$$

where $z$ is the output value after feature normalization, $x_i$ is the value of the variable, $\max(x_i)$ is the maximum value of the variable, and $\min(x_i)$ is the minimum value of the variable. Normalizing the area and duration of fire gave the two normalization results equal weight when calculating the average value, which was used as the final normalization result.

### D. WILDFIRE SCALE PREDICTION MODELING

#### 1) BACKPROPAGATION NEURAL NETWORK

A BPNN is a multilayer feed-forward neural network based on backpropagation [30]. Typical BPNNs consist of one input layer, one or more hidden layers, and one output layer. Each layer consists of several neurons (nodes). The output value of each node is determined by its input value, action function, and threshold. The network's learning process includes two processes: forward propagation of information and backpropagation of error. In forward propagation, input information is transmitted from the input layer to the output layer through the hidden layer, which is obtained by operation of the action function. If the output contains errors when compared with the desired value, reverse propagation of the error signal is conducted. Errors are reduced by modifying the weights of each layer of neurons so that the output meets accuracy requirements. By alternating the two processes, in the right vector space execution error function gradient descent strategy, dynamic iterative search weight vector, the network error function is minimized, completing the process of information extraction and memory. The topological structure of BPNN is shown in Fig. 2.

#### 2) RECURRENT NEURAL NETWORK

An RNN is a neural network model for time series data. The special network structure of an RNN enables the output of a neuron to act as input directly to itself at the succeeding time point [31], [32]. The result of each hidden layer in the network is determined by the output of the current input and the output of the previous hidden layer; that is, an RNN can record the results of previous calculations [33]. However, with an RNN, the problems of gradient disappearance or explosion can easily occur [34], [35]. As shown in Fig. 3, given the input sequence $(x_1, x_2, \ldots, x_t)$ and the hidden layer state $(h_1, h_2, \ldots, h_t)$, at time $t$, RNN units are updated as shown in Eqs. (2)–(3).

$$h_t = \sigma(Ux_t + Wh_{t-1} + b), \tag{2}$$

$$o_t = \sigma(Vh_t + b). \tag{3}$$

#### 3) LONG SHORT-TERM MEMORY

LSTM is a variant RNN; it can learn long-term dependency information and avoid gradient disappearance [36], [37].

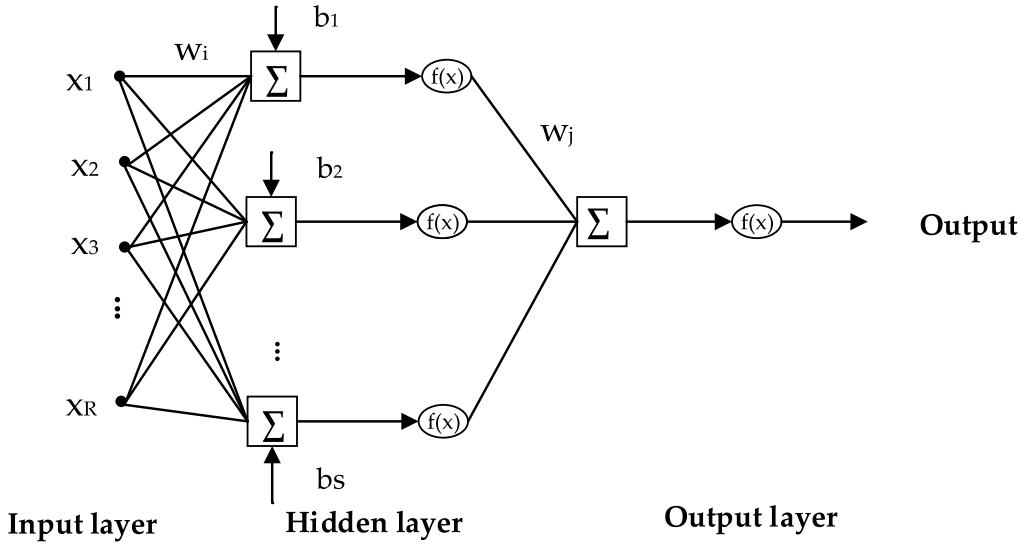


**FIGURE 2.** Structure of BP neural network. Symbols: X, input gate; $w_i$ = weight between input layer and hidden layer; $w_j$ = weight between hidden layer and output layer; b = bais; f(x) = transfer function.

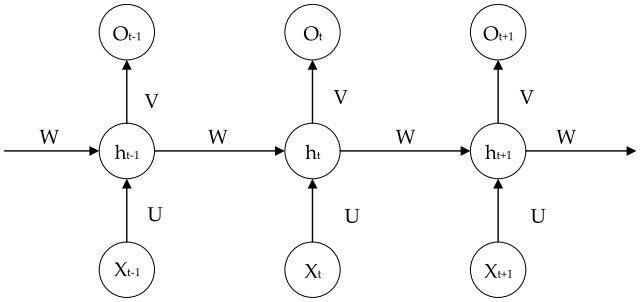


**FIGURE 3.** Structure of recurrent neural network. Symbols: o = output gate; h = hidden layer; x = input gate; t = time; W = weight matrix of hidden layer; U = weight matrix of input layer to hidden layer; V = weight matrix of hidden layer to output layer.

LSTM improves long-term dependencies by increasing input, output, and "forget" gates in neurons. The structure of an LSTM neuron is shown in Fig. 4.

Each LSTM neuron has three gates—the input, output, and forget gates—to control the state of the cell. Given the input sequence $(x_1, x_2, \ldots, x_t)$ and hidden layer state $(h_1, h_2, \ldots, h_t)$, at time $t$, the LSTM cells are updated as shown in Eqs. (4)–(8) [38].

$$f_t = \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_f\right) \tag{4}$$

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{5}$$

$$o_t = \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right) \tag{6}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh\left(W_c \cdot [h_{t-1}, x_t] + b_c\right) \tag{7}$$

$$h_t = o_t \odot \left(\tanh\left(c_t\right)\right) \tag{8}$$

where $\sigma$ is the sigmoid activation function, $f$ is the forget gate, $i$ is the input gate, $o$ is the output gate, $c$ is the cell, $h$ is the status of the hidden layers, $W$ is the weight matrix, and $b$ is the deviation vector.

#### 4) TRAINING THE MODELS

In our study, the objective function for wildfire scale prediction modeling was the root mean square error (RMSE) measuring the magnitude of the error between the observations and predictions, which should be minimized:

$$e = t - y \tag{9}$$

$$RMSE = \sqrt{Mean(e^2)} \tag{10}$$

$$\text{Objectivefunction} = min(RMSE) \tag{11}$$

where $t$ is the target data, $y$ is a function of the input data and the models whose values are optimized, and $e$ is the error function value to be minimized.

### E. PERFORMANCE ASSESSMENT AND COMPARISON

In our study, forest wildfire scale was graded into five ratings. In addition to calculating the prediction accuracy of the BPNN, RNN, and LSTM models on each rating, the receiver operating characteristic (ROC) was used to evaluate the overall performance of the models. The ROC of the training dataset shows the model's success rate and indicates how well the modeling results fit the training dataset. In contrast, the ROC of the test dataset yields the prediction rate of the model and measures how well the model predicts the general probability of fire scale across the study area.

### F. SOFTWARE

Data collection and preprocessing, and the establishment of the classification models, were developed and implemented in the TensorFlow framework of the Anaconda 3 software (Anaconda, Inc., Austin, TX).

## III. RESULTS AND DISCUSSION

### A. DATA PROCESSING

After filtering the original fire data, we obtained the ignition points, corresponding to latitudes and longitudes in Alberta.

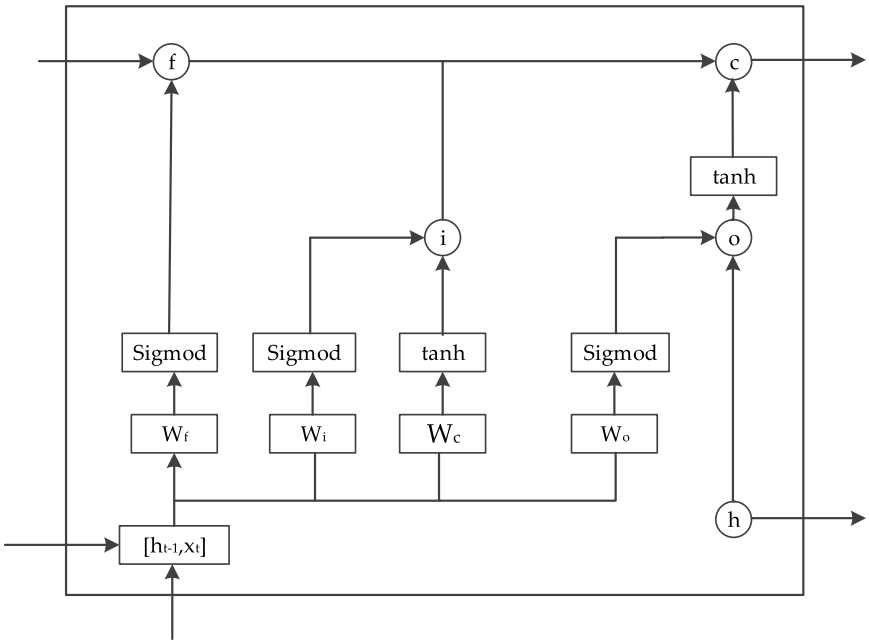


**FIGURE 4.** Structure of LSTM cell. Symbols: c = cell; f = forget gate; h = hidden layer; i = input gate; o = output gate; Sigmod = sigmoid activation function; t = time; x = input; W = weight matrix; b = deviation vector.

**TABLE 1.** Wildfire data.

| Variable | Example Values | | | | | |
|---|---|---|---|---|---|---|
| Latitude | 54.7167 | 58.46403 | 51.53933 | 58.09718 | … | 58.30206 |
| Longitude | −111.261 | −114.485 | −115.179 | −116.148 | … | −116.201 |
| Ignition Date | 1990/11/5 | 1998/4/2 | 2009/4/11 | 2013/5/7 | … | 2016/5/9 |
| Out Date | 1990/11/13 | 1998/4/3 | 2009/4/13 | 2013/5/13 | … | 2016/5/14 |
| Duration of Fire (d) | 9 | 2 | 3 | 7 | … | 6 |
| Burned Area (ha) | 25.6 | 35 | 5.2 | 140 | … | 9.7 |

**TABLE 2.** Post-processed weather data.

| Variable | Example Values | | | | | |
|---|---|---|---|---|---|---|
| Maximum Temperature (°C) | 10.1 | 16.5 | 11 | 12.7 | … | 23.5 |
| Minimum Temperature (°C) | −4.9 | 2 | −1.6 | 0.9 | … | 9.6 |
| Mean Temperature (°C) | 2.6 | 9.3 | 4.7 | 6.8 | … | 16.6 |
| Cooling Degree Days [a] | 0 | 0 | 0 | 0 | … | 0 |
| Heating Degree Days [b] | 15.4 | 8.7 | 13.3 | 11.2 | … | 1.4 |
| Total Rain (mm) | 0 | 0 | 0 | 0 | … | 0 |
| Total Snow (mm) | 0 | 0 | 0 | 0 | … | 0 |
| Total Precipitation (mm) | 0 | 0 | 0 | 0 | … | 3.2 |
| Snow on Ground (cm) | 0 | 0 | 0 | 0 | … | 0 |
| Direction of Maximum Gust (tens of degrees) | 22 | 16 | 5 | 21 | … | 31 |
| Speed of Maximum Gust (km/h) | 37 | 38 | 43 | 33 | … | 46 |

[a] The number of days before the wildfire occurrence on which the daily mean temperature was below 18°C.
[b] The number of days before the wildfire occurrence on which the daily mean temperature was above 18°C.

The format of the fire data is shown in Table 1. These data included the latitude and longitude of the ignition point, the ignition and extinction dates of the fire, and the size of the burned area. As the study focused on wildfires, fires with human causes were eliminated; the types that remained included crown fires, grass fires, ground fires, and forest fires. Filtering for the meteorological data preserved those meteorological records that contained all 11 variables; their format is presented in Table 2.

### B. MULTI-COLLINEARITY TEST AND DATA FEATURE NORMALIZATION

A multi-collinearity test was used to test the collinearity of the meteorological variables affecting forest fire occurrence. The VIFs of the meteorological variables are shown in Table 3. The minimum temperature, mean temperature, and total precipitation variables should be excluded as their VIF values were greater than 10. After eliminating the meteorological variables with multiple collinearities, the fire

**TABLE 3.** The variance inflation factors (VIFs) of the meteorological variables.

| Variable | VIF |
|---|---|
| Maximum Temperature (°C) | 4.88 |
| Minimum Temperature (°C) | 21.44 |
| Mean Temperature (°C) | 21.60 |
| Cooling Degree Days | 3.63 |
| Heating Degree Days | 1.55 |
| Total Rain (mm) | 1.11 |
| Total Snow (mm) | 1.13 |
| Total Precipitation (mm) | 121.15 |
| Snow on Ground (cm) | 1.42 |
| Direction of Maximum Gust (tens of degrees) | 1.08 |
| Speed of Maximum Gust (km/h) | 1.09 |

**TABLE 4.** Statistics of the dataset for modelling.

| Variable | Maximum | Minimum | Mean | S.D. |
|---|---|---|---|---|
| Maximum Temperature (°C) | 35.6 | −22.2 | 11.54 | 11.60 |
| Cooling Degree Days | 10 | 0 | 0.41 | 1.25 |
| Heating Degree Days | 44 | 0 | 12.59 | 10.38 |
| Total Rain (mm) | 124 | 0 | 2.47 | 6.68 |
| Total Snow (mm) | 43.5 | 0 | 0.51 | 2.51 |
| Snow on Ground (cm) | 63 | 0 | 4.34 | 10.65 |
| Direction of Maximum Gust (tens of degrees) | 36 | 0 | 20.67 | 10.10 |
| Speed of Maximum Gust (km/h) | 109 | 0 | 32.09 | 17.95 |
| Burned Area (ha) | 335 | 0 | 0.31 | 5.10 |
| Duration of Fire (d) | 48 | 0 | 0.14 | 0.95 |

**TABLE 5.** Variable values after feature normalization.

| Variable | Example Values after Feature Normalization | | | | | |
|---|---|---|---|---|---|---|
| Maximum Temperature (°C) | 0.26 | 0.45 | 0.29 | 0.32 | … | 0.65 |
| Cooling Degree Days | 0 | 0 | 0 | 0 | … | 0 |
| Heating Degree Days | 0.68 | 0.38 | 0.58 | 0.57 | … | 0.06 |
| Total Rain (mm) | 0 | 0 | 0 | 0 | … | 0 |
| Total Snow (mm) | 0 | 0 | 0 | 0 | … | 0 |
| Snow on Ground (cm) | 0 | 0 | 0 | 0 | … | 0 |
| Direction of Maximum Gust (tens of degrees) | 0.61 | 0.44 | 0.14 | 0.58 | … | 0.86 |
| Speed of Maximum Gust (km/h) | 0.45 | 0.46 | 0.52 | 0.46 | … | 0.56 |
| Burned Area (ha) | 0.68 | 0.15 | 0.44 | 0.26 | … | 0.13 |
| Duration of Fire (d) | 0.42 | 0.27 | 0.37 | 0.31 | … | 0.16 |

and meteorological data were matched based on the latitude and longitude of the fire ignition points. Then, the modeling dataset was prepared. Each sample in the dataset was constructed of eight meteorological variables and two wildfire variables (i.e., the burned area and fire duration). The meteorological variables were the maximum temperature (MT), cooling degree days (CDD), heating degree days (HDD), total rain (TR), total snow (TS), snow on ground (SG), direction of maximum wind gust (DMG), and speed of maximum wind gust (SMG). The traditional method for rating wildfire scale is based on the size of the burned area [39]. However, we propose a more meaningful definition of scale that encompasses not only the size of the burned area but also the fire's duration [40]. The basic statistical descriptions of the dataset variables are given in Table 4. Additionally, we implemented Min-Max Scaling to normalize the variables into the [0,1] range. The results of feature normalization are shown in Table 5.

### C. CORRELATION BETWEEN FIRE SCALE AND METEOROLOGICAL VARIABLES

After multi-collinearity testing and feature normalization, we carried out correlation analysis of the meteorological factors and fire scale. The results are shown in Table 6. MT, TR, and SMG have the greatest impact on a fire's scale. Moreover, MT, HDD, DMG, and SMG were significantly positively correlated with fire scale, whereas CDD, TR, SG, and TS were significantly negatively correlated with fire scale. Furthermore, the relative importance of meteorological factors to wildfire scale prediction is shown in Fig. 5.

### D. FOREST WILDFIRE SCALE PREDICTION

#### 1) WILDFIRE SCALE RATINGS

After feature normalization, we used the mean $z$ of the normalized duration of the fire and the normalized size of the

**TABLE 6.** Correlation between fire scales and environmental factors.

| Variable | MT | CDD | HDD | TR | TS | SG | DMG | SMG | Fire Scales |
|---|---|---|---|---|---|---|---|---|---|
| MT | 1 | | | | | | | | |
| CDD | -0.421** | 1 | | | | | | | |
| HDD | 0.523** | -0.462** | 1 | | | | | | |
| TR | -0.033** | -0.013** | -0.032** | 1 | | | | | |
| TS | -0.421** | 0.365** | -0.324** | 0.241** | 1 | | | | |
| SN | -0.442** | 0.214** | -0.448** | -0.314** | 0.531** | 1 | | | |
| DMG | 0.323** | 0.412** | 0.021** | 0.338** | 0.132** | -0.012** | 1 | | |
| SMG | 0.481** | 0.423** | 0.032** | 0.347** | 0.122** | -0.023** | 0.622** | 1 | |
| Fire Scales | 0.732** | -0.422** | 0.543** | -0.688** | -0.312** | -0.283** | 0.214** | 0.617** | 1 |

MT: Maximum Temperature (° C), CDD: Cooling Degree Days, HDD: Heating Degree Days, TR: Total Rain (mm), TS: Total Snow (mm), SG: Snow on Ground (cm), DMG: Direction of Maximum Gust (tens of degrees), SMG: Speed of Maximum Gust (km/h), **: Significant correlation at 0.01 level.

**TABLE 7.** Results of dataset division by improved Kennard–Stone (K-S) method.

| Scale of Wildfire | Training Set (n[a] = 50,627) | Testing Set (n = 21,697) | Total (n = 72,324) |
|---|---|---|---|
| Level 5 | 1,239 | 531 | 1,770 |
| Level 4 | 2,154 | 923 | 3,077 |
| Level 3 | 4,919 | 2,108 | 7,027 |
| Level 2 | 5,178 | 2,219 | 7,397 |
| Level 1 | 3,385 | 1,451 | 4,836 |
| No fire | 33,752 | 14,465 | 48,217 |

[a]n represents the number of records.

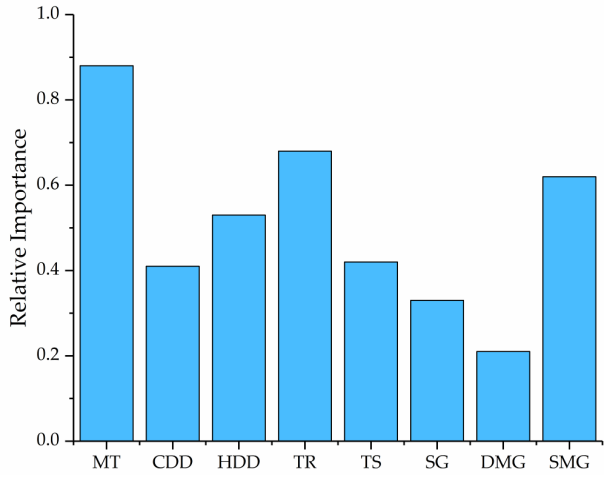


**FIGURE 5.** The relative importance of meteorological factors. MT: Maximum Temperature (°C), CDD: Cooling Degree Days, HDD: Heating Degree Days, TR: Total Rain (mm), TS: Total Snow (mm), SG: Snow on Ground (cm), DMG: Direction of Maximum Gust (tens of degrees), SMG: Speed of Maximum Gust (km/h).

burned area to classify the wildfires by their scale. The wildfires were classified into five scale ratings according to their $z$ values: $z$ values in the ranges of [0,0.2), [0.2,0.4), [0.4,0.6), [0.6,0.8), and [0.8,1) were classified as Levels 1, 2, 3, 4, and 5, respectively. Level 5 represents a very large-scale wildfire, whereas Level 1 indicates a wildfire that was very small in scale [41].

### 2) ANALYSIS OF THE PREDICTIVE MODELS

The dataset for the establishment of the predictive models included 24,108 samples, representing 24,108 occurrences of wildfires between 1990 and 2018. To reveal the patterns in forest wildfire scales more objectively, 48,216 random samples of records without fire occurrences were selected (twice the number of wildfire occurrence records). Thus, the full dataset for wildfire prediction modeling contained 72,324 samples. The samples were divided into training and testing sets in a ratio of 7:3; that is, 50,627 records were put into the training set and 21,697 records were put into the testing set [42]. To guarantee the applicability and stability of the prediction model, we applied the improved Kennard–Stone (K–S) method [43] to put the most diverse samples into the training set. The number of records of each level placed into the training set and the testing set are shown in Table 7.

Setting the eight meteorological variables as input and wildfire scale rating $z$ as output, the study used BPNN, RNN, and LSTM models to analyze the relationship between meteorological factors and wildfire scale. A comparison of the prediction accuracy of the BPNN, RNN, and LSTM on the testing set is shown in Table 8. The LSTM model obtained the best predictive performance of the three models, having the highest average accuracy of 90.9%; that is, 19,721 of the 21,697 records in the testing set were predicted correctly. Details of the prediction distribution obtained using LSTM are presented in Table 9. For example, there were 531 records with a rating of Level 5 in the testing set, of which the LSTM predictive model discriminated 487 records to the correct level; therefore, the accuracy of LSTM for Level 5 was 91.7%.

As a further evaluation, Fig. 6 displays the ROC curves of the test dataset and the full dataset using the LSTM model. The diagonal of Table 9 shows training samples with correct predictions and other cells show training samples with wrong

**TABLE 8.** Comparison of results by prediction models using back propagation neural network (BPNN), recurrent neural network (RNN), and long short-term memory (LSTM).

| Model Type | Accuracy (%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | Level 5 | Level 4 | Level 3 | Level 2 | Level 1 | No fire | Overall |
| BPNN | 71.9% | 69.3% | 70.5% | 70.1% | 68.2% | 72.4% | 70.4% |
| RNN | 84.2% | 83.3% | 83.9% | 82.7% | 83.1% | 83.8% | 83.5% |
| LSTM | 91.7% | 89.9% | 90.5% | 90.3% | 90.6% | 91.1% | 90.9% |

**TABLE 9.** Prediction results for the five levels of forest wildfire scales using LSTM.

| True Wildfire Scale Rating | Predicted Wildfire Scale Rating | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Level 5 | Level 4 | Level 3 | Level 2 | Level 1 | No fire | Total | Accuracy |
| Level 5 | 487 | 20 | 13 | 9 | 2 | 0 | 531 | 91.7% |
| Level 4 | 47 | 830 | 23 | 13 | 7 | 3 | 923 | 89.9% |
| Level 3 | 12 | 65 | 1,907 | 97 | 17 | 10 | 2,108 | 90.5% |
| Level 2 | 11 | 30 | 57 | 2,004 | 79 | 38 | 2,219 | 90.3% |
| Level 1 | 9 | 14 | 22 | 39 | 1,315 | 52 | 1,451 | 90.6% |
| No fire | 68 | 102 | 214 | 307 | 596 | 13,178 | 14,465 | 91.1% |
| Total | 634 | 1,061 | 2,236 | 2,469 | 2,016 | 13,281 | 21,697 | 90.9% |

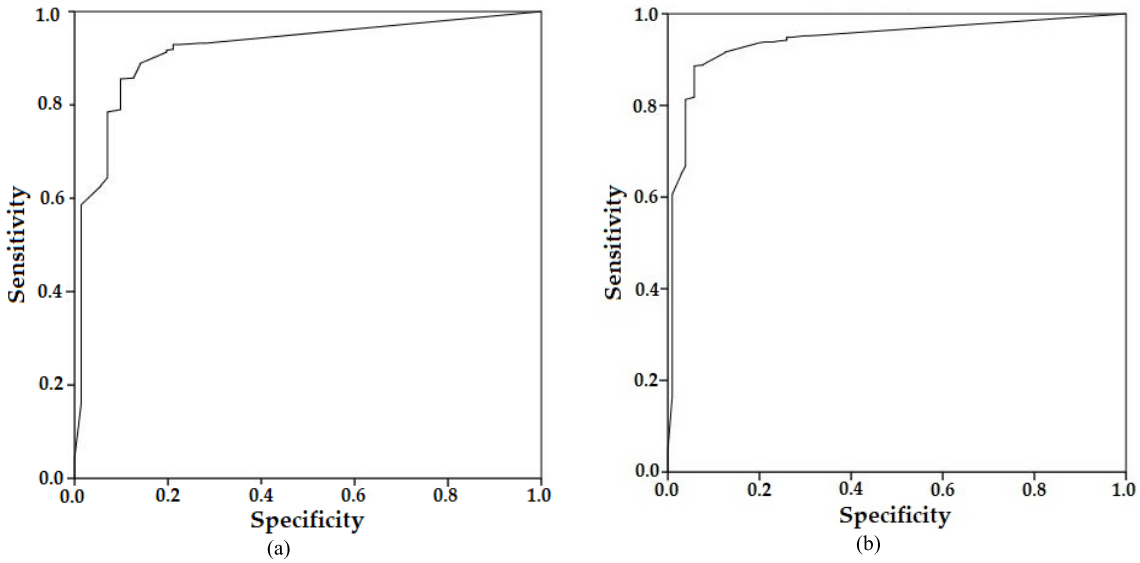


**FIGURE 6.** Receiver operating characteristic (ROC) curves of LSTM model. (a) Test dataset; (b) full dataset.

predictions. The ROC curves of the training and test sets are drawn according to the samples with correct training and wrong training. The area under the ROC curve (AUC) values reflect the accuracy of the model's prediction. When $0.5 < \text{AUC} < 1$, the larger the AUC value, the better the model's fit. (Some scholars have pointed out that $\text{AUC} = 0.5$ indicates that the regression equation has no meaning for the interpretation of dependent variables; when $\text{AUC} > 0.5$, the independent variables are better able to explain the dependent variables [44].) The AUC values of the test dataset and the full dataset were 0.918 and 0.942, respectively.

The experimental results indicate that LSTM can predict the scale of wildfires by using meteorological variables, which can provide a scientific basis for forest wildfire scale prediction in Alberta.

## IV. CONCLUSION AND FUTURE WORK

This study revealed the relationship between meteorological factors and wildfire scale in the forests of Alberta, Canada. The main contribution of this work is twofold:

(1) The study proposed a more meaningful definition of wildfire scale that is estimated by the size of the burned area and the duration of the fire. These scales are classified into five levels.

(2) Of the three neural network models examined, the LSTM model exhibited the best ability to predict the scales of forest wildfires, with an overall predictive accuracy of 90.9%. Additionally, the ROC curve indicated that the LSTM model fit the data well (having an AUC of 0.942 for the full dataset).

The results of this study demonstrate that it is feasible to predict the scale of forest wildfires using meteorological data, which will be helpful in forest fire prevention and rescue, especially for wildfires occurring in forests. Fire rescuers and firefighters will be able to take effective and appropriate measures in accordance with a fire's scale as predicted at its initial stages, thereby reducing the losses caused by forest wildfires.

Because the data used for modeling came from a single region, the types of forest and the topography throughout the dataset were similar, and the predictive model has some limitations. This study thus constitutes a step forward in the domain of forest wildfire prediction by considering other factors, including terrain topography, altitude, type of forest, population density, and the manual intervention in fire extinguishing that, along with meteorological factors, may influence the scale of forest wildfires. As more factors are considered, the model will have the potential to predict the scales of a greater range of wildfire occurrences.

## ACKNOWLEDGMENT

The authors would like to thank the Canadian Forest Service for providing the National Fire Database Agency Fire Data.

## REFERENCES

[1] R. Baker, "Forest history: International studies on socioeconomic and forest ecosystem change," *Forest Ecol. Manage.*, vol. 159, no. 3, p. 293, 2002, doi: 10.1016/S0378-1127(01)00441-8.

[2] L. Shu, X. Tian, and X. Kou, "The focus and progress on forest fire research," *World Forestry Res.*, vol. 16, no. 3, pp. 37–40, 2003, doi: 10.3969/j.issn.1001-4241.2003.04.007.

[3] D. J. Rasbash, "Effects of fire on items which may have helped cause the fire," *Fire Saf. J.*, vol. 7, no. 3, pp. 293–294, 1984, doi: 10.1016/0379-7112(84)90027-4.

[4] A. Sekizawa, "Fire risk analysis: Its validity and potential for application in fire safety," presented at the 8th Int. Symp. Fire Saf. Sci., Beijing, China, Sep. 2005, pp. 85–100. [Online]. Available: http://www.iafss.org/publications/fss/8/85/view/fss_8-85.pdf

[5] B. M. Wotton, D. L. Martell, and K. A. Logan, "Climate change and people-caused forest fire occurrence in Ontario," *Climatic Change*, vol. 60, pp. 275–295, Oct. 2003, doi: 10.1023/A.1026075919710.

[6] J. Marchal, S. G. Cumming, and E. J. B. Mcintire, "Exploiting Poisson additivity to predict fire frequency from maps of fire weather and land cover in boreal forests of Québec, Canada," *Ecography*, vol. 40, no. 1, pp. 200–209, 2017, doi: 10.1111/ecog.01849.

[7] M. P. Plucinski, W. L. Mccaw, J. S. Gould, and B. M. Wotton, "Predicting the number of daily human-caused bushfires to assist suppression planning in south-west Western Australia," *Int. J. Wildland Fire*, vol. 23, pp. 520–531, Jul. 2014, doi: 10.1071/WF13090.

[8] A. C. Alencar, L. A. Solórzano, and D. C. Nepstad, "Modeling forest understory fires in an eastern Amazonian landscape," *Ecol. Appl.*, vol. 14, pp. 139–149, Aug. 2014, doi: 10.1890/01-6029.

[9] M. G. Rollins, P. Morgan, and T. Swetnam, "Landscape-scale controls over $20^{th}$ century fire occurrence in two large Rocky Mountain (USA) wilderness areas," *Landscape Ecol.*, vol. 17, pp. 537–557, Aug. 2002, doi: 10.1023/A:1021584519109.

[10] J. P. Prestemon, M. L. Chas-Amil, J. M. Touza, and S. L. Goodrick, "Forecasting intentional wildfires using temporal and spatiotemporal autocorrelations," *Int. J. Wildland Fire*, vol. 21, pp. 743–754, Oct. 2012, doi: 10.1071/WF11049.

[11] M. H. Nami, A. Jaafari, M. Fallah, and S. Nabiuni, "Spatial prediction of wildfire probability in the Hyrcanian ecoregion using evidential belief function model and GIS," *Int. J. Environ. Sci. Technol.*, vol. 15, pp. 373–384, Feb. 2017, doi: 10.1007/s13762-017-1371-6.

[12] B. M. Wotton and D. L. Martell, "A lightning fire occurrence model for Ontario," *Can. J. Forest Res.*, vol. 35, pp. 1389–1401, Jun. 2005, doi:10.1139/x05-071.

[13] L. Vilar, D. G. Woolford, D. L. Martell, and M. P. Martín, "A model for predicting human-caused wildfire occurrence in the region of Madrid, Spain," *Int. J. Wildland Fire*, vol. 19, pp. 325–337, Jun. 2010, doi: 10.1071/WF09030.

[14] F. Guo, L. Zhang, S. Jin, M. Tigabu, Z. Su, and W. Wang, "Modeling anthropogenic fire occurrence in the boreal forest of China using logistic regression and random forests," *Forests*, vol. 7, no. 11, p. 250, 2016, doi: 10.3390/f7110250.

[15] Y. J. Goldarag, A. Mohammadzadeh, and A. S. Ardakani, "Fire risk assessment using neural network and logistic regression," *J. Indian Soc. Remote Sens.*, vol. 44, pp. 885–894, Dec. 2016, doi: 10.1007/s12524-016-0557-6.

[16] D. T. Bui, B. Pradhan, H. Nampak, Q.-T. Bui, Q.-A. Tran, and Q.-P. Nguyen, "Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibilitgy modeling in a high-frequency tropical cyclone area using GIS," *J. Hydrol.*, vol. 540, pp. 317–330, Sep. 2016, doi: 10.1016/j.jhydrol.2016.06.027.

[17] H. Hong, M. Panahi, A. Shirzadi, T. Ma, J. Liu, A.-X. Zhu, W. Chen, I. Kougias, and N. Kazakis, "Flood susceptibility assessment in Hengfeng area coupling adaptive neuro-fuzzy inference system with genetic algorithm and differential evolution," *Sci. Total Environ.*, vol. 621, pp. 1124–1141, Apr. 2018, doi: 10.1016/j.scitotenv.2017.10.114.

[18] S. Oliveira, F. Oehler, J. San-Miguel-Ayanz, A. Camia, and J. M. C. Pereira, "Modeling spatial patterns of fire occurrence in Mediterranean Europe using multiple regression and random forest," *Forest Ecol. Manage.*, vol. 275, pp. 117–129, Jul. 2012, doi: 10.1016/j.foreco.2012.03.003.

[19] M. Rodrigues and J. de la Riva, "An insight into machine-learning algorithms to model human-caused wildfire occurrence," *Environ. Model. Softw.*, vol. 57, pp. 192–201, 2014, doi: 10.1016/j.envsoft.2014.03.003.

[20] A. Jaafari, E. K. Zenner, M. Panahi, and H. Shahabi, "Hybrid artificial intelligence models based on a neuro-fuzzy system and metaheuristic optimization algorithms for spatial prediction of wildfire probability," *Agricult. Forest Meteorol.*, vols. 266–267, pp. 198–207, 2019, doi: 10.1016/j.agrformet.2018.12.015.

[21] F. A. Gers and E. Schmidhuber, "LSTM recurrent networks learn simple context-free and context-sensitive languages," *IEEE Trans. Neural Netw.*, vol. 12, no. 6, pp. 1333–1340, Nov. 2001, doi: 10.1109/72.963769.

[22] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, 2000, doi: 10.1162/089976600300015015.

[23] X. Qing and Y. Niu, "Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM," *Energy*, vol. 148, pp. 461–468, Apr. 2018, doi: 10.1016/j.energy.2018.01.177.

[24] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "LSTM fully convolutional networks for time series classification," *IEEE Access*, vol. 6, pp. 1662–1669, 2017, doi: 10.1109/ACCESS.2017.2779939.

[25] S. Mcallister, I. Grenfell, A. Hadlow, W. M. Jolly, M. Finney, and J. Cohen, "Piloted ignition of live forest fuels," *Fire Saf. J.*, vol. 51, pp. 133–142, Jul. 2012, doi: 10.1016/j.firesaf.2012.04.001.

[26] L. Huiling, L. Yurui, and Y. Guang, "Application of random forest algorithm on the forest fire prediction in Tahe area based on meteorological factors," *Sci. Silvae Sinicae*, vol. 52, pp. 90–98, Jan. 2016, doi: 10.13360/j.issn.2096-1359.2019.03.020.

[27] J. S. Littell, D. McKenzie, D. L. Peterson, and A. L. Westerling, "Climate and wildfire area burned in western US ecoprovinces, 1916–2003," *Ecol. Appl.*, vol. 19, no. 4, pp. 1003–1021, 2009, doi: 10.1080/10807031003670469.

[28] L. Vilar, H. Nieto, and M. P. Martín, "Integration of lightning- and human-caused wildfire occurrence models," *Hum. Ecol. Risk Assessment, Int. J.*, vol. 16, pp. 340–364, Apr. 2010, doi: 10.1080/10807031003670469.

[29] Y. Li, N. Wang, J. Shi, X. Hou, and J. Liu, "Adaptive batch normalization for practical domain adaptation," *Pattern Recognit.*, vol. 80, pp. 109–117, Aug. 2018, doi: 10.1016/j.patcog.2018.03.005.

[30] R. Zhang, Y. Wang, K. Wang, H. Zhao, S. Xu, L. Mu, and G. Zhou, "An evaluating model for smart growth plan based on BP neural network and set pair analysis," *J. Cleaner Prod.*, vol. 226, pp. 928–939, Jul. 2019, doi: 10.1016/j.jclepro.2019.03.053.

[31] T. Zhang, W. Dong, M. Keeter-Brewer, S. Konar, R. N. Njabon, and Z. R. Tian, "Site-specific nucleation and growth kinetics in hierarchical nanosyntheses of branched zno crystallites," *J. Amer. Chem. Soc.*, vol. 128, no. 33, pp. 10960–10968, 2006, doi: 10.1021/ja0631596.

[32] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," presented at the 30th Int. Conf. Mach. Learn., Atlanta, GA, USA, vol. 28, 2013. [Online]. Available: http://proceedings.mlr.press/v28/pascanu13.pdf
[33] E. D. Übeyli, "Combining recurrent neural networks with eigenvector methods for classification of ECG beats," *Digit. Signal Process.*, vol. 19, no. 2, pp. 320–329, 2009, doi: 10.1016/j.dsp.2008.09.002.
[34] G. Hinton, L. Deng, and D. Yu, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012, doi: 10.1109/MSP.2012.2205597.
[35] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015, doi: 10.1038/nature14539.
[36] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
[37] M. Sundermeyer, H. Ney, and R. Schlüter, "From feedforward to recurrent LSTM neural networks for language modeling," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 3, pp. 517–529, Mar. 2015, doi: 10.1109/TASLP.2015.2400218.
[38] D. Monner and J. A. Reggia, "A generalized LSTM-like training algorithm for second-order recurrent neural networks," *Neural Netw.*, vol. 25, pp. 70–83, Jan. 2012, doi: 10.1016/j.neunet.2011.07.003.
[39] A. Krasovskii, N. Khabarov, and J. Pirker, "Modeling burned areas in Indonesia: The FLAM approach," *Forests*, vol. 9, p. 437, Jul. 2018, doi: 10.3390/f9070437.
[40] J. Rodríguez-Veiga, M. Ginzo-Villamayor, and B. Casas-Méndez, "An integer linear programming model to select and temporally allocate resources for fighting forest fires," *Forests*, vol. 9, p. 583, Sep. 2018, doi: 10.3390/f9100583.
[41] H. K. Preisler, D. R. Brillinger, R. E. Burgan, and J. W. Benoit, "Probability based models for estimation of wildfire risk," *Int. J. Wildland Fire*, vol. 13, pp. 133–142, Jul. 2014, doi: 10.1071/WF02061.
[42] H. Li, J.-X. Wang, Z.-N. Xing, and G. Shen, "Influence of improved Kennard/Stone algorithm on the calibration transfer in near-infrared spectroscopy," *Spectrosc. Spectral Anal.*, vol. 31, pp. 362–365, Feb. 2011, doi: 10.3964/j.issn.1000-0593(2011)02-0362-04.
[43] H. Liang, M. Zhang, C. Gao, and Y. Zhao, "Non-destructive methodology to determine modulus of elasticity in static bending of quercus mongolica using near-infrared spectroscopy," *Sensors*, vol. 18, p. 1963, Jun. 2018, doi: 10.3390/s18061963.
[44] M. Jesús, C. Vega-Garcia, and E. Chuvieco, "Human-caused wildfire risk rating for prevention planning in Spain," *J. Environ. Manage.*, vol. 90, pp. 1241–1252, Feb. 2009, doi: 10.1016/j.jenvman.2008.07.005.

**HAO LIANG** was born in Jinzhong, Shanxi, China, in 1989. He received the B.S. and M.S. degrees in computer software and theory and the Ph.D. degree in forestry engineering automation from Northeast Forestry University, China, in 2012, 2014, and 2017, respectively. Since 2017, he has been a Lecturer with the School of Technology, Beijing Forestry University. His research interests include artificial intelligent, pattern recognition, smart forestry, the Internet of Things, and autonomous systems. He was a recipient of the Fundamental Research Funds for the central universities, in 2017.

**MENG ZHANG** was born in Harbin, Heilongjiang, China, in 1996. She received the B.S. degree in automation from Beijing Forestry University, China, in 2018, where she is currently pursuing the M.S. degree in control engineering. Her research interests include signal processing, big data, artificial intelligence, and smart forestry.

**HAILAN WANG** was born in Shannxi, China. She received the B.S. degree in electrical engineering and the M.S. degree in control engineering and control theory from Liaoning Technical University, in 1996 and 1999, respectively, and the Ph.D. degree in forest engineering from Beijing Forestry University, in 2011. She is an Associate Professor with the School of Technology, Beijing Forestry University. Her research interests include smart forestry, sensors, the Internet of Things, and control engineering.