


Article

RAFFIA: Short-term Forest Fire Danger Rating Prediction via Multiclass Logistic Regression

Lei Wang ^{1,2,*} , Qingjian Zhao ¹, Zuomin Wen ^{1,*} and Jiaming Qu ³

¹ School of Economics and Management, and Intelligent Big Service Laboratory (InBSLab), Nanjing Forestry University, Nanjing 210037, China; zhao5190@126.com

² Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin HSB 101, Hong Kong, China

³ School of Information and Library Science, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3175, USA; jiaming@ad.unc.edu

* Correspondence: leiwangchn@163.com (L.W.); zmwen@njfu.edu.cn (Z.W.); Tel.: +86-25-8542-7377 (L.W.)

Received: 22 October 2018; Accepted: 1 December 2018; Published: 5 December 2018



Abstract: Forest fire prevention is important because of human communities near forests or in the wildland-urban interfaces. Short-term forest fire danger rating prediction is an effective way to provide early guidance for forest fire managers. It can therefore effectively protect the forest resources and enhance the sustainability of the forest ecosystem. However, relevant existing forest fire danger rating prediction models operate well only when applied to distinct climates and fuel types separately. There are desires for an effective methodology, which can construct a specific short-term prediction model according to an evaluation of the data from that specific region. Moreover, a suitable method for prediction model construction needs to deal with some big data related computing challenges (i.e., data diversity coupled with complexity of solution space, and the requirement of real-time forest fire prevention application) when massively observed heterogeneous parameters are available for prediction (e.g., meteorology factor, the amount of litter in the area, soil moisture, etc.). To capture the influences of multiple prediction factors on the prediction results and effectively learn from fast cumulative historical big data, artificial intelligence methods are investigated in this paper, yielding a short-term Ratings of Forest Fire Danger Prediction via Multiclass Logistic Regression (or RAFFIA) model for forest fire danger rating online prediction. Experimental evaluations conducted on a sensor-based forest fire prevention experimental station show that RAFFIA (with 98.71% precision and 0.081 root mean square error) is more effective than the Least Square Fitting Regression (LSFR) and Random Forests (RF) prediction models.

Keywords: forest fire; danger rating online prediction; multiclass logistic regression; artificial intelligence; big data

1. Introduction

Forests are a critical component in terms of protecting environmental sustainability (e.g., air, ecological diversity, soil, and hydrological cycle, etc.). Nevertheless, forest fire is a common and natural part of forest ecosystems. It destroys forests and woodlands, emits greenhouse gases on a massive scale, and has been recognized as a critical disaster for humans [1]. The negative impacts of large-scale, uncontrolled forest fires have been increasing worldwide over the past three decades [2]. Just as an example, as a result of the growing frequency of droughts and/or warmer winters, occurrences of fires in Canadian forests have been observed to be on the rise; resulting in a decrease in atmospheric carbon at this source [3]. Some works also suggest that forest fires have a negative effects on water yield around the post-fire area in Melbourne, Australia [4]. Much evidence points out that forest fire management plays a significant role in guaranteeing the ecological environment and sustainable development.

Reactive and proactive methods have already been used to provide some early warnings of forest fires for forest managers [5]. In particular, the reactive methods, monitoring devices (e.g., sensor-networks, optical-, or infrared-based screen monitoring devices, unmanned aerial vehicles, remote sensing, etc.) are widely used for forest fire prevention. By analyzing text-, image-, or video-based data obtained from the above monitoring devices, one can detect forest fire occurrences at an early stage. Alarms for fighting the forest fire will be send out when fire is identified [6]. Proactive methods, by some prediction mechanisms, can provide some warning information to help to mitigate the dangers of forest fires before their occurrence. Therefore, following the proactive principle, an effective prediction in forest fires is more effective for mitigating or preventing loss.

1.1. Forest Fire Danger Rating Prediction Models

To date, several well-known forest fire danger rating prediction models have been proposed and used over the world, such as Fosberg Fire Weather Index (FFWI), McArthur Mark 5 Forest Fire Danger Meter, McArthur Mark 4 and Mark 5 Grassland Fire Danger Meters, to name a few [7].

Different prediction models fit well for different climates and fuel types in a specific region [7]. In particular, FFWI was employed to supplement the U.S. National Fire Danger Rating System (NFDRS [8]), e.g., it was used for Alaska wildfire danger assessment. Another well-known system is the Canadian Forest Fire Danger Rating System (CFFDRS [9]), which is based on the Canadian Forest Fire Weather Index (FWI) model [5]. Daily weather data are used in both NFDRS and CFFDRS to determine fuel moisture.

The Fire Susceptibility Index (FSI) proposed in Reference [10] is another model for identifying fire susceptibility in different spatial variations. A simplified metric (namely F) for determining the rating of forest fire is proposed in Reference [11]. Following McArthur Mark 4 Grassland Fire Danger Meter, McArthur Mark 5 Grassland Fire Danger Meter, and Fosberg Fire Weather Index, a simpler function is used in F to measure the forest fire danger rating.

1.2. AI for Forest Fire Danger Prediction

In recent studies, artificial intelligence (AI) methodologies have already been applied in the prediction of forest fire occurrence risk.

Elmas and Sönmez [12] present a Forest Fire Decision Support System (FOFDESS) based on multi-agent technology. Artificial Neural Networks (ANN), Naive Bayes Classifier (NBC), and Fuzzy Switching (FS) are used in the system to predict fire danger rating, estimate fire spread speed, and quickly detect started fires. ANN was also adopted in Reference [13] to predict the danger of forest fire occurrence. Support Vector Machines (SVM) were applied in Reference [14] for forest fire danger prediction.

In Reference [1], data mining technology (concerning association analysis) is also used to investigate the problems of global forest monitoring, relationship mining, and carbon danger scoring based on remote sensing image data. Interesting results concerting dependence rules across different spatio-temporal scale were discovered in the paper. For example, the anomalous warming of the eastern tropical region in the area of Pacific (i.e., El Niño phenomenon) may increase fire danger for Indonesian forests. Geographic information systems and remote sensing technologies are used in Reference [15,16] to assess spatial and temporal danger conditions of forest fires.

An advanced Gaussian-Process (GP)-emulator for wildland fire emission estimation was proposed in Reference [17]. Cluster analysis results in the simulation show that fire emission has a physical relationship with fuel and environmental conditions.

1.3. Factors Used for Forest Fire Danger Prediction

Since fuel accumulation, moisture, and forest structure are important influencing factors for forest fires, and these factors are mainly affected by mountainous topography, Kane et al. [18] adopted a random forest model to predict the site water balance and topography (slope position, slope,

and insolation) using them as environmental predictors. The prediction results show substantial portions of the variations in fire and forest structure.

Intelligent methodologies in terms of data mining and machine learning, including boosted regression tree (BRT), generalized additive model (GAM), and random forest (RF), have been adopted in Reference [19] for the features selection from 15 prior determined condition factors on forest fires. In particular, the condition factors were identified by expert opinion and a comprehensive literature review. As a result, three main driven factors (i.e., annual rainfall, distance to roads, and a land use factors) were selected. Two environmental factors (i.e., temperature and atmospheric moisture) were selected in Reference [20] to predict forest fire occurrence danger.

As a case study, Tian et al. [21] investigated the distribution characteristics and influence factors of forest fires in China. In sum, forest fires are more likely to occur in humid regions of the medium temperate zone. Pacheco et al. [22] also pointed out that terrain, fuels, as well as weather factors are the environmental effects for forest fires. Shang et al. [23] also demonstrated the positive relationship between accumulation of fuels and forest fire dangers.

In sum, the applications of forest fire danger rating prediction systems are extensive. Some typical examples include NFDRS and CFFDRS [24]. They perform predictions using daily weather indicators and the estimation of moisture content of the fuel [11]. Nevertheless, state of the art systems remain largely in the realm of research and application. First, different models are used for a specific region; they cannot scale for other areas because the model parameters are determined off-line by some statistical analysis tools based on observations of the corresponding region. There are desires for an intelligent methodology for model construction for forest fire danger rating prediction. Second, the models perform well in predicting long-term (e.g., on a daily basis) forest fire danger, but cannot estimate forest fires in the near future (referred to as online prediction, c.f., [25–27]) in view of real time environmental changes [22].

Long-term prediction results can help the development of rational and sensible forest fire prevention and protection policy for forest managers [28]. However, for some specific regions, such as economic forest areas, forest areas near scenic spots, and forests with varied topography and climate conditions, short-term forest fire danger rating prediction is essentially important. If the short-term forest fire danger rating prediction problem is not attended, real time forest fire risk may be dismissed in these regions. The social, ecological, economics, legal/institutional, and environmental policy costs for dealing with the forest fire danger would be increased. An effective short-term forest fire danger rating prediction can provide some early warnings for forest managers before forest fires start in these specific regions. Proactive decisions based on the prediction results to mitigate the fire danger can make the forest ecosystem more stable. The short-term forest fire danger rating prediction can therefore contribute to the sustainability of forests as well as the social, ecological, economics systems.

Wireless sensor networks (WSNs) are an emerging technology that can collect environment parameters almost in real time. Analyzing the collected data can therefore identify forest fires or even help to predict them before they start [29,30]. The online prediction of forest fire danger rating via WSNs-based monitoring data faces the following big data challenges.

1. *Variety of Observed Data:* Different forest fire danger influencing parameters have different measurement units (e.g., °C, m/s, etc.) and data ranges. Some parameters have positive effect on forest fire danger; others may be negative. There are desires for an unified metric to aggregate the diverse influences.
2. *Accuracy Guarantee for Online Prediction:* To provide effective early guidance for forest fire prevention, the prediction requires a high accuracy performance. To guarantee the prediction accuracy, WSNs-based real-time parameters should be computed in real time and the prediction model construction should be capable of dealing with a large-volume of cumulative sample data.
3. *Complexity of Solution Space:* It is very difficult to recognize complex patterns from the diverse, large-volume of data and guarantee the effectiveness of online prediction results. To provide a generalized model construction methodology for forest fire danger rating prediction which

can scale for different regions, the decision-making process cannot be handled by traditional models. The complex reasoning regularities and the high-accuracy requirements will result in a high computational overhead.

None of the relevant existing forest fire danger rating prediction models and approaches can systematically address these challenges and provide real-time short-term forest fire danger rating prediction results. In particular, the popular prediction models (e.g., NFDRS, CFDRS, FSI, etc.) in existing applications can often only provide long-term region-oriented forest fire danger rating prediction [5,8–10]. Some artificial intelligence (AI)-based prediction models [12–14,17,19] cannot directly scale for large-scale observation data and deal with the big data related challenges.

We propose a short-term forest fire danger rating prediction approach in this paper based on artificial intelligence technology to deal with the challenges highlighted above. In particular, the observation parameters are first pre-processed dimensionless by min-max transformation. Henceforth, each value is normalized into the range [0,1]: the larger value leading to a higher forest fire danger rating. To reduce the computational complexity of model construction, a lightweight machine learning approach is adopted, namely the multiclass logistic regression model. This model includes less weight values and can therefore speed up the learning process. We evaluated the proposed approach via a sensor-based forest fire prevention experimental station. The results demonstrated the high prediction performance of our model. The remainder of the paper is organized as follows. We describe the study approaches in Section 2. We present the effectiveness and efficiency evaluation results in Section 3. We give some discussions regarding the main contributions of this paper in Section 4. We conclude by identifying some important future works in Section 5.

2. Methods

This section provides a step-by-step study approach which includes information regarding the study area, the proposed prediction model and approach, the deployed forest fire prevention experimental station, and the data set and approaches used for comparison.

2.1. Study Area

In this paper, Xiashu Forest was selected as the study area. As demonstrated in Figure 1, Xiashu Forest, with a total area of 314.4 ha, is located in 119°14' east longitude and 31°59' north latitude, at the Jurong County, Jiangsu Province, China. Surrounded by low hills, the forest is divided into 11 compartments.

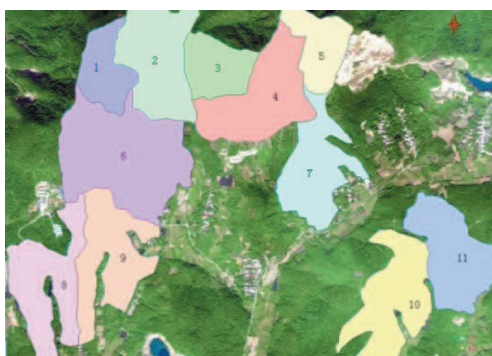


Figure 1. Study area—divided into 11 forest compartments.

The highest peak in the forest, WuQi Mountain, is 377.8 m above sea level, followed by KongQing Mountain with an altitude of 322.6 m above sea level. The surrounding area is generally about 100 m above sea level; the valley is 75 m. The relative height difference in the forest is 302.8 m. The overall situation of the forest farm is that the terrain is not undulating and relatively gentle, and it belongs to the low hilly area.

The soil in the forest farm is dominated by yellow brown soil and mountain yellow brown soil. The surface humus content is generally 2.5%. The humus layer is not thick—about 10 cm to 20 cm. The soil is acidic to strongly acidic; the soil texture is not sticky.

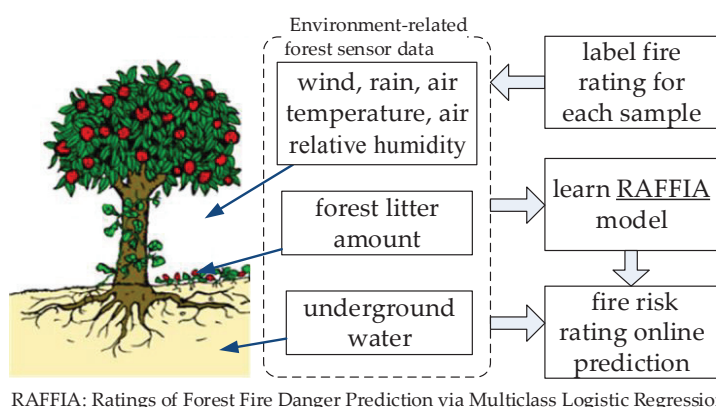
The local zonal vegetation is a deciduous mixed forest with evergreen components. Xiashu Forest belongs to the north subtropical monsoon climate zone. The climate is characterized by four distinct seasons of dry, wet and hot weather, sufficient sunshine and abundant water and heat resources. According to observations of the forest for many years, the annual average temperature is 15.2 °C, the annual average sunshine is 2157 h, and the sunshine percentage is 49%. The annual average rainfall amount is 1055.6 mm, which varies greatly from year to year. The maximum amount is 1408.3 mm (year 1962) and the minimum is 425.2 mm (year 1978). Most rainfall is in the summer, followed by spring and autumn. The annual average relative humidity of the air is 79%.

Xiashu Forest is bordered by Jurong County Forest. Xiashu Town is in the north, and the west and south bordering Tingzi Town. There are some residential areas in and around the forest farm where the villagers live. Human activity is common in the forest, which makes fire prevention a very critical issue in Xiashu Forest.

2.2. Forest Fire Danger Ratings Online Prediction

As stated in Section 1, prevention of forest fires is an extremely important issue to support the sustainable development of forest ecosystems, economic systems, and social systems. On the basis of the idea of proactive fault management (PFM) [27,31], predicting the occurrence of forest fires at an early stage can help to mitigate the fire danger ahead of the fire, thus ensuring the stable operation of the forest ecosystem. For some special forest fire prevention areas (such as Xiashu Forest), short-term fire danger rating predictions and the subsequent proactive management can effectively reduce the danger of fires, and lower the cost of ensuring the sustainability of forest ecosystems, economic/social systems. We propose a Ratings of Forest Fire Danger Prediction model via Multiclass Logistic Regression (namely RAFFIA) in this paper and validated the model in Xiashu Forest.

As demonstrated in Figure 2, the following three types of sensor data are included in RAFFIA model. (1) Above ground environment-related forest sensor data (such as wind, rain, air temperature, air relative humidity, etc.); (2) earth's surface sensor data (i.e., forest litter amount); and (3) under ground soil water content sensor data. The RAFFIA model learns the patterns from labeled historical data.



RAFFIA: Ratings of Forest Fire Danger Prediction via Multiclass Logistic Regression

Figure 2. Framework of forest fire danger rating online prediction.

The online prediction is conducted based on a trained RAFFIA model and the real-time collected environment-related forest sensor data. Firstly, the sufficient volume of training data set is collected based on the historical data. After non-dimensional normalization processing of the training set data, each record in the training data set is marked by a fire rank according to historical experience. The training data set is used as an example and is provided for subsequent model learning. Secondly,

the RAFFIA prediction model is trained via Multiclass Logistic Regression and the training data set. Finally, on the basis of the forest sensor data collected in real time, the forest fire danger rating short-term prediction is executed using the learned RAFFIA model.

2.2.1. Data Preparation

As for the WSNs-based forest fire danger rating prediction, we mainly consider six environment-related parameters for the RAFFIA model in this paper. The parameters can be categorized as follows (c.f., Figure 2). Category 1 is above ground indicators, i.e., air temperature (v_1), maximal horizontal wind speed (v_2), air relative humidity (v_3), and rainfall (v_4). Category 2 is the earth's surface related indicator. We mainly consider forest litter amount (v_5). Category 3 reflects the underground water amount, namely soil relative humidity (v_6).

Parameters for all the indicators above can be collected by forest environment monitoring WSNs. The collected data is first be cleaned. For example, the missing values are supplemented. Wrong data is also be corrected.

Each of the cleaned parameters has a different dimension due to the different measuring units, e.g., parameter v_4 is always measured by mm/h, while the metric for parameter v_5 is m by the popular forest litter amount sensor (e.g., SR50A). Moreover, the forest fire danger rating increases as some parameters (namely benefit-oriented indexes) become larger, e.g., v_1 , v_2 , and v_5 ; while as some parameters (namely cost-oriented indexes) become lower, e.g., v_3 , v_4 , and v_6 . The diversity of collected observation parameters make it difficult to discover patterns from the data. As with Reference [32,33], we employ the min-max transformation method [34] and define dimensionless additive utility functions to handle the parameters. Therefore, each value of the observed parameter will be normalized into a real number with the range [0,1].

Assume that v_{ij} represents an observed value of parameter v_i , for benefit-oriented indexes (i.e., $i \in \{1, 2, 5\}$), we normalized v_{ij} by

$$v_{ij} = \begin{cases} \frac{v_{ij} - v_i^{\min}}{v_i^{\max} - v_i^{\min}}, & v_i^{\max} \neq v_i^{\min} \\ 1, & v_i^{\max} = v_i^{\min} \end{cases} \quad (1)$$

For cost-oriented indexes v_{ij} , $i \in \{3, 4, 6\}$, we will normalized v_{ij} by

$$v_{ij} = \begin{cases} \frac{v_i^{\max} - v_{ij}}{v_i^{\max} - v_i^{\min}}, & v_i^{\max} \neq v_i^{\min} \\ 1, & v_i^{\max} = v_i^{\min} \end{cases} \quad (2)$$

in Equations (1) and (2), v_i^{\max} and v_i^{\min} represent the maximal and minimal value of v_i , respectively.

2.2.2. RAFFIA Model Construction

We first collect historical environment-related forest parameters to train the model (i.e., RAFFIA) for the online prediction of forest fire danger rating. After normalization, each group of the six parameter values, formally as $V_j = \{v_{1j}, v_{2j}, \dots, v_{6j}\}$, are labeled by a forest fire danger rating number. We then train a multiclass logistic regression [35] model to solve the following regression problem:

$$F_j^{rating} = f(v_{1j}, v_{2j}, \dots, v_{6j}), \quad (3)$$

where F_j^{rating} represents the forest fire danger rating for sample V_j .

As can be seen from Table 1, we follow the meteorological industry standard (i.e., QX/T77-2007) of the people's Republic of China, and divide the forest fire danger into five ratings.

Table 1. The forest fire weather ratings in reference to QX/T77-2007.

Rating	F^{rating}	Danger Degree	Flammable Degree	Warning Color
1	0.2	Very Low	Difficult	Green
2	0.4	Low	Very Difficult	Blue
3	0.6	High	Easy	Yellow
4	0.8	Very High	Very Easy	Orange
5	1.0	Extremely High	Extremely Easy	Red

Since the values in each sample V_j are normalized into the range $[0,1]$, we train a Sigmoid function to solve the regression problem presented in Equation (3). We have

$$F_{rating}^j = \frac{1}{1 + e^{-\theta^T V_j}}, \quad (4)$$

where $\theta^T V_j$ is a dot product operation, in which $\theta^T V_j = \sum_i \theta_i v_{ij}$, the vector $\theta = \{\theta_i\}_{i=1 \text{ to } 6}$ represents the weight parameters for each valuable.

Hence, the prediction of F_{rating}^j mainly depends on the weights θ . As each sample vector V_j of observed parameter is labeled by a rating number, we first convert the rating number into a real number F^{rating} according to Table 1. To determine the weight values and construct RAFFIA model, the gradient descent method [36] is used.

Let F_{rating}^j be the converted labeled forest fire danger rating value for a sample vector V_j (representing the actual forest fire danger value), let \hat{F}_{rating}^j be the output value by Equation (4) using the current weight values. We define the following Log Loss function L to measure the cost from the output value to the actual value.

$$L = -\frac{1}{S} \sum_{j=1}^S [F_{rating}^j \log(\hat{F}_{rating}^j) + (1 - F_{rating}^j) \log(1 - \hat{F}_{rating}^j)], \quad (5)$$

where S represents the total number of sample vectors.

The gradient descent for training the RAFFIA model is conducted by the following procedures.

- Step 1: we initialize each value of $\theta_i = 1$.
- Step 2: We randomly select S sample vectors from the training set, and calculate the L value of cost function according to the selected samples.
- Step 3: We update θ_i , for $i \in \{1, 2, \dots, 6\}$, using a grad function. Specifically, the purpose of updating weight values is reducing the value of cost function. For each $\theta_i \in \theta$, we assign a new value θ'_i using the following equation:

$$\theta_i \rightarrow \theta'_i = \theta_i - \eta \frac{\partial L}{\partial \theta_i}, \quad (6)$$

in which $\eta \in [0, 1]$ is defined as learning rate, which is used to control the speed of updating θ_i .

As for Step 3, when η is too large, the process of a gradient descent may cause the cost function to cross the bottom (i.e., the minimum value), in this situation, θ approximates the optimal solution. On the other hand, too small values of the η setting will lead to a higher computational complexity; it will therefore cause a low convergence speed for the learning algorithm. As regards the data, with the improved performance of computer hardware, especially the use of GPUs, the time complexity of the gradient descent algorithm in the process of solving the large-scale machine learning problems has gradually gained acceptance. In general, the optimal value of η should be determined via sufficient experimental evaluations.

Repeating the gradient descent procedures, Step 2 and Step 3, as stated above, we constantly update the weight values in the RAFFIA model. When $L = \arg \min_L (L \approx 0)$, the RAFFIA model training process achieves convergence. In this way, we use the trained RAFFIA model for forest fire danger rating online prediction via the real-time collected sensor-based forest environment observation parameters.

2.2.3. Online Prediction

Let $V_o = \{v_{10}, v_{20}, \dots, v_{60}\}$ be the real-time observed forest environment parameter values, vector V_o is substituted into Equation (4) to obtain the numbered forest fire danger value F_{rating}^o . We use an interval function to determine the danger rating (formally as R^o , where $R^o \in \{1, 2, \dots, 5\}$) for an online forest fire danger rating prediction.

$$R^o = \begin{cases} 1, & F_{rating}^o \leq \varepsilon_1 \\ 2 & \varepsilon_1 < F_{rating}^o \leq \varepsilon_2 \\ 3, & \varepsilon_2 < F_{rating}^o \leq \varepsilon_3, \\ 4, & \varepsilon_3 < F_{rating}^o \leq \varepsilon_4 \\ 5, & F_{rating}^o > \varepsilon_4. \end{cases} \quad (7)$$

Specifically, we set $\varepsilon_1 = 0.2$, $\varepsilon_2 = 0.4$, $\varepsilon_3 = 0.6$, $\varepsilon_4 = 0.8$, respectively, in this paper.

2.3. Framework of Xiashu Forest Fire Prevention Experimental Station

A forest fire prevention experimental station was deployed in Xiashu Forest. Located at $118^\circ 79'$ east longitude and $32^\circ 06'$ north latitude, the experimental station was behind the forest management office, at the 9th compartment (see Figure 1). As can be seen in Figure 3, a flux tower (with 50 m high) and some sensors continuously collected data to store at the data center servers. There were also some optical and infrared dual lens fire monitoring equipment (the monitoring distance was 10 km) in the flux tower. The data center servers were deployed in a container, which were used to store the data collected from the sensors.

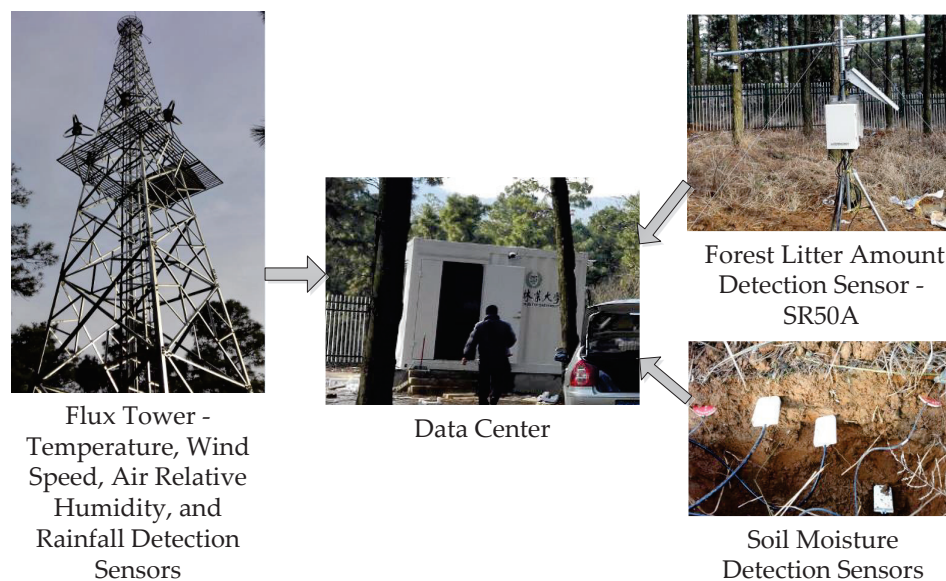


Figure 3. Data collection sensors and the data center—deployed in 9th compartment.

The experimental station collected Meteorological Data (including Temperature- v_1 , Maximal Horizontal Wind Speed- $v_2 (=Max(U_x, U_y))$, where U_x and U_y represent the wind speed in x and y

axis, respectively.), Air Relative Humidity (RH)- v_3 , and Rainfall- v_4), Forest Litter Amount- v_5 , and Soil Relative Humidity- v_6 . The hardware platform for the data centers included two Langchao servers—NF8460, equipped with two Xeon E7 4820 CPUs and 32GB RAM for each server, and a Langchao optical storage server-AS500E, with a 10TB HDD. A 100M The Chinese education network connected the experimental station to the university laboratory. This system provided a basic platform for the evaluation of the prediction results. The data collected by the sensor networks is summarized in Table 2.

Table 2. Forest environment-related sensor-monitored parameters

Variables	Sensors	Units	Data Ranges
v_1	Model 107 Temperature Probe	°C	[−35, +100]
v_2	IRGASON Integrated CO ² /H ₂ O Open-Path Gas Analyzer and 3D Sonic Anemometer	m/s	65.553
v_3	Model HMP 155A Temperature and Relative Humidity Probe	%RH	[0,100]
v_4	TE525 Tipping Bucket Rain Gage	mm/h	[0,30]
v_5	SR50A	m	[0.5,10]
v_6	Model HFP01 Soil Heat Flux Plate	%(m ³ /m ³)	[0,100]

Taking 16:30 real-time data on 14 November 2018 as an example, the collected sensor data related to this paper mainly include: temperature 7.7481 °C, maximal horizontal wind speed 3.0710 m/s, Air relative humidity 55.6597%, rainfall 0.00 mm, forest litter amount 16.20 cm, the soil relative humidity 0.2871%. These parameters were used to train the model and conduct fire danger rating predictions.

2.4. Data Set

We collected the forest environment parameters using the sensors stated in Table 2 every 30 min, continuously for 12 months. As a result, there are altogether $8760 = 365 \times 24$ samples. As for each sample, we labeled the forest fire danger rating based on the daily forecasting result (http://www.slhf.gov.cn/slfnw/Category_81/Index.aspx) published by China Meteorological Administration and the State Forestry Administration of the People’s Republic of China. We then invited a forest fire prevention expert to artificially correct the labeled rating for each sample. We randomly selected 6000 samples for the training set, and used the remaining 2760 samples as the testing set. Specifically, each sample in training set and testing set was labeled by a danger rating number with $\{1, 2, \dots, 5\}$ via the above artificially correction.

2.5. Metrics

We employ two popular metrics including precision and Root Mean Square Error (RMSE) to evaluate the prediction accuracy.

Particularly, precision for forest fire danger rating online prediction is defined as

$$precision = (N_h/N) \times 100\%, \quad (8)$$

where N represents total number of prediction times, N_h represents the time of prediction result hits the labeled danger rating.

RMSE is defined as

$$RMSE = \sqrt{\frac{\sum_{n=1}^N (R_n^o - R_n^c)^2}{N}}, \quad (9)$$

where R_n^o is the danger rating by n th time of prediction, R_n^c represents the labeled ranking, N is the total number of predictions.

From the above metrics, higher precision and lower RMSE indicate higher prediction accuracy.

2.6. Approaches Subjected to Comparison

Two relevant existing regression prediction approaches including Least Square Fitting Regression (LSFR) and Random Forests (RF) were also conducted. These approaches were compared with the proposed RAFFIA approach for forest fire danger ranking prediction to evaluate the effectiveness of the proposed approach.

For LSFR, we defined a fitting function as follows

$$F_j^{rating} = \sum_{k=0}^4 a_k (V_j)^k, \quad (10)$$

where a_k is a regression coefficient, $(V_j)^k$ represents the k-power of sample vector V_j .

Coefficients a_k were solved using MATLAB 7.0 toolkit based on the training set. The prediction using the LSFR model with trained coefficients were conducted in Java.

The RF model is constructed in Java using Spark 1.4.0. Classes

`org.apache.spark.mllib.tree.RandomFores` for `trainClassifier` method, and `org.apache.spark.mllib.tree.model.RandomForestModel` for `predict` method

in Spark were used for model training and prediction, respectively.

The proposed RAFFIA approach was conducted using Java. All the experiments were implemented on a PC with Intel(R) Core(TM) i7 2600 CPU, 4GB RAM, Seagate 1TB HDD. The results were obtained by averaging over 50 runs of all the predictions under the testing set, and the models training under the training set, respectively.

3. Results

This section presents the effectiveness and efficiency evaluation results, and demonstrates the application of the proposed RAFFIA model based on the study approach.

3.1. Impact of η and S

We first investigated the effect of learning rate η and the number of sample vector S for each step of gradient descent on the effectiveness and efficiency of the RAFFIA model. We fixed the value of s as 20 and varied the value of η from 0.1 to 1, with a step value of 0.1. We compared the precision and RMSE, respectively, under different parameter settings for the predictions based on the RAFFIA model. We then fixed the value of $\eta = 0.5$, and varied the value of S for from 10 to 80, with a step value of 10, and studied the execution time for the RAFFIA model training under different parameter settings.

As can be seen from Figures 4–6, the precision, RMSE, and execution time all exhibit an obvious impact on the prediction accuracy. In sum, a smaller value of η and larger value of S result in lower prediction accuracy and higher computation complexity for the RAFFIA model construction.

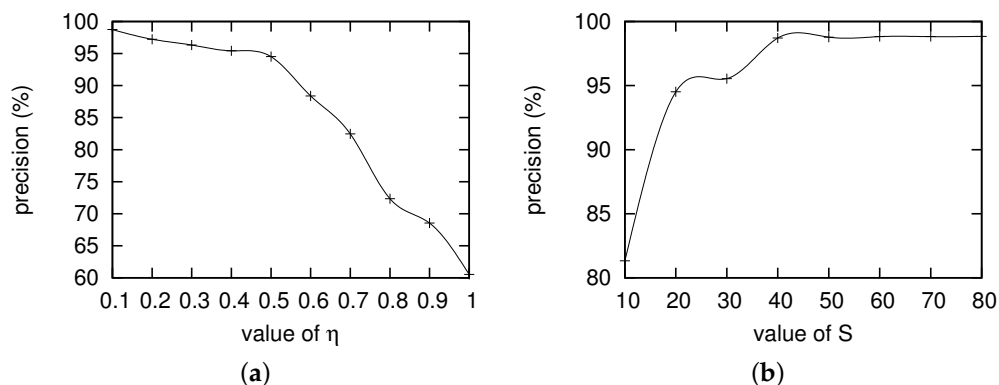


Figure 4. Precision. (a) Impact of η ($S = 20$); (b) Impact of S ($\eta = 0.5$).

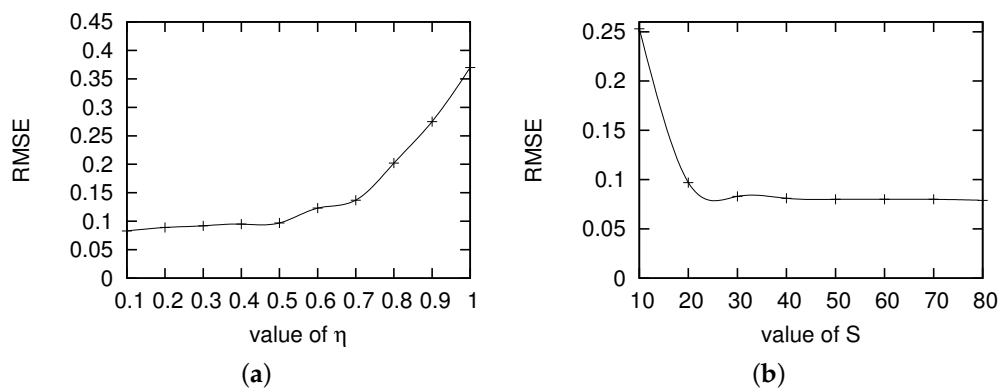


Figure 5. Root Mean Square Error (RMSE). (a) Impact of η ($S = 20$); (b) Impact of S ($\eta = 0.5$).

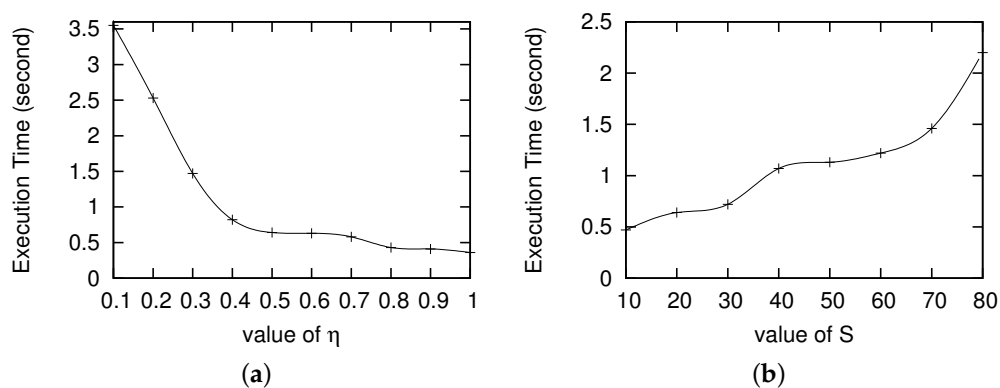


Figure 6. Execution Time for Prediction Model Construction. (a) Impact of η ($S = 20$); (b) Impact of S ($\eta = 0.5$).

First, when S is fixed, a smaller value of η makes the stride length for each step's grad function become smaller. We are more easily close to the optimal solution for the RAFFIA model construction via gradient descent. The enlargement of η gradually makes the minimum value of cost function L enlarge (away from 0), in this situation, the algorithm for the RAFFIA model construction converges. Especially, when η reaches 0.5, the prediction accuracy declines more obviously as η further increases. The results suggest that, in this case, η should be set less than 0.5 to effectively predict the forest fire danger rating.

Second, when η is fixed at 0.5, the prediction accuracy obviously improves when S increases. The enlargement trend gradually reduces when S reaches 20. The changes in prediction accuracy gradually level off when S is 40. This finding indicates that larger amount of randomly selected sample vectors for each step of the gradient descent make the RAFFIA model more effective. But when S reaches 40, any further increment of S becomes irrelevant.

Third, when S is fixed, the execution time for the RAFFIA model construction declines significantly as η increases. The growth trend slows down until η reaches 0.5. It is because when $\eta \leq 0.5$, the enlargement of η makes the step length for gradient become larger, it results in a quicker convergence by finding the optimal values of θ for the RAFFIA model training. On the other hand, when $\eta > 0.5$, the step value for the gradient is larger, and the convergence is determined by identifying the minimal value of L ; it, however, makes the efficiency of the RAFFIA model poorer. Moreover, when η is fixed, the enlargement of S makes the computational complexity for each step's gradient descent become large. Particularly, the results show a liner growth for the time complexity.

To sum up, reasonable values for S and η should be set as $S \geq 40$, and $\eta \leq 0.5$. Nevertheless, trade-off decisions should also be made for optimal values of S and η regarding the prediction accuracy and time complexity, in the situation of different applications.

3.2. Performance Comparison

According to the above findings, we set $S = 40$ and $\eta = 0.5$ for RAFFIA, and compared the precision and RMSE of RAFFIA with LSFR and RF methods, respectively. Because the process of the LSFR model construction needs an off-line Matlab toolkit calculation, we could not compare the execution time for LSFR with other methods with a uniform metric. In this paper, we only compare the time complexity for the model construction of RAFFIA and RF.

As can be seen from Figure 7, RAFFIA outperforms the other methods in both prediction accuracy and computational complexity. RF is better than LSFR as regards prediction accuracy.

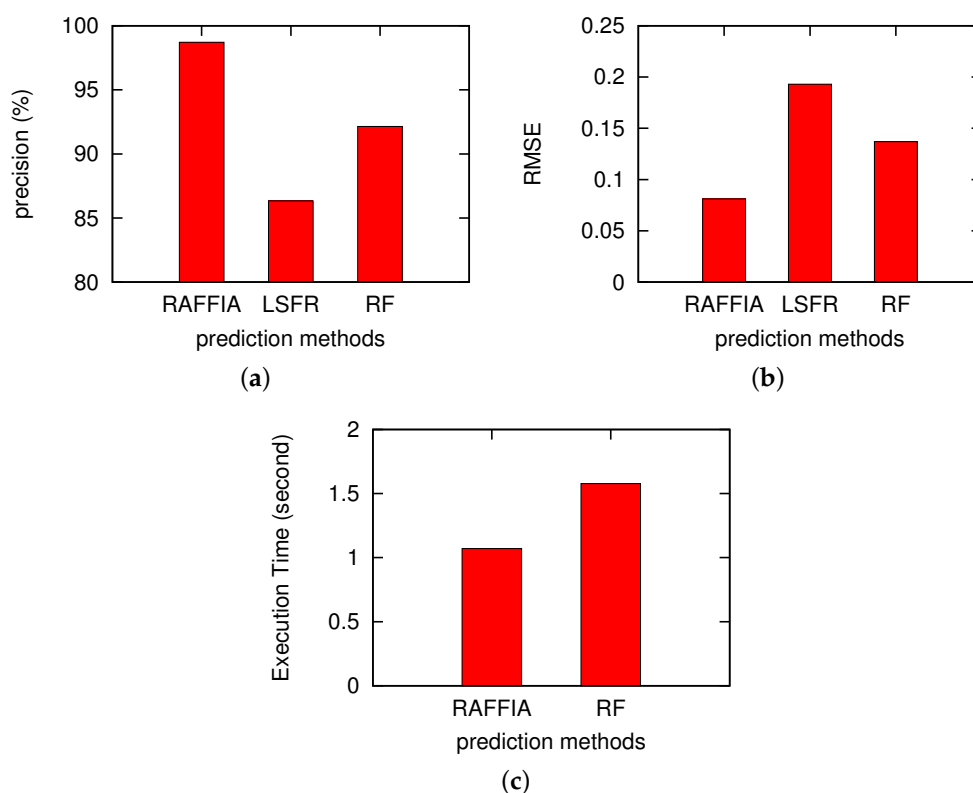


Figure 7. Performance Comparison for Different Approaches ($\eta = 0.5, S = 40$). (a) precision; (b) RMSE; (c) Execution Time for Model Construction.

Moreover, we evaluated the efficiency of the proposed RAFFIA approach for online prediction. We repeated runs of the prediction from 30 to 100 times, with a step value of 10 times, and compared the results with the LSFR and RF methods.

As can be seen from Figure 8, the execution time for the same amount of runs of the RAFFIA approach is very similar to that of the LSFR method. It is much better than the RF method as regards of running times. The results also show a liner increment as for the prediction times increase. As for each execution of prediction, RAFFIA takes around 250 ms, showing a good real-time efficiency.

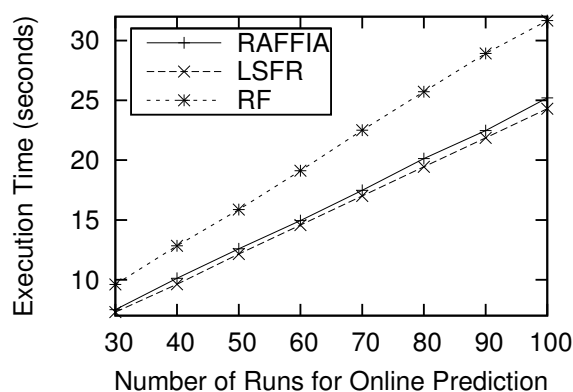


Figure 8. Efficiency Comparison.

In sum, the proposed RAFFIA model is more fit for short-term forest fire danger rating prediction applications in the situation with big data related computing challenges, i.e., a big volume and large variety of observed data, and the complexity of solving the short-term prediction problem under the big data environment.

4. Discussion

Different from the traditional long-term forest fire danger rating prediction methods (such as providing monthly or daily prediction results) [5,8,9,13], the proposed RAFFIA model can perform short-term prediction based on real-time data collected by sensor networks. Long-term prediction methods usually focus on meteorological factors [7]. However, the distribution of soil and tree species in forests often vary widely from one area to another. This shows that the soil relative humidity, the forest litter amount, etc., are largely different in different regions [17]. For some key fire prevention areas in the forest, such as the areas with intensive human activities, the areas with important social, economic, and ecological values, it is especially important to predict small-scale, short-term fire danger ratings to guarantee the sustainable and stable development of these areas. More real-time parameters including meteorological and environmental parameters of the specific area should be collected based on the sensor networks to enhance the prediction accuracy.

The experimental evaluation results in this paper show that the prediction accuracy and computational complexity of the RAFFIA model are better than other competitive methods (i.e., LSFR and RF [18]). In terms of dealing with the big data challenges, the RAFFIA model only spent 1.07 s average for training a RAFFIA model under a 6000 samples' training set. The average online prediction time is only 250 ms, which is better than LSFR and RF methods. It can be seen that the proposed method is more suitable for small-scale short-term forest fire danger rating prediction applications.

It is worth noting that, for the application of RAFFIA, more fire risk parameters can be easily added to the RAFFIA model, such as terrain, population density, etc. In particular, for some non-quantitative parameters, such as slope degree, slope direction, etc., we can first qualitatively describe the fire danger, and then convert the qualitative value into a quantitative value according to Saaty's 1-to-9 scale for analytic hierarchy processes (AHP) and analytic network processes (ANP) [37]. By considering more fire risk factors, the effectiveness of the proposed method in small-scale and short-term forest fire danger rating prediction for specific areas can be further improved.

The application of the RAFFIA-based prediction method requires the support of sensor networks. These devices increase the cost of forest fire prevention. However, compared to remote sensing-based, aircraft, drone, and optical-based monitoring methods [15,16], the financial investment is still relatively low. In addition, the monitoring methods above can only find burning fires, and cannot predict the fire danger in the near future. From the perspective of forest fire emergency management, prediction results can provide earlier information for decision making than monitoring results [27,31]. Early warning and intervention mechanism can easily be set up by forest managers and emergency response

professionals based on the RAFFIA prediction results. This will greatly reduce the fire danger in key fire prevention areas, and reduce the investment of fire inspectors and equipment. The prediction method proposed in this paper does not damage the environment. It can facilitate the environment and social management.

In sum, with the deployment and application of the RAFFIA model in the fields of fire emergency management, environmental emergency management, and social system emergency management, the proactive emergency management ability of managers to respond to the danger of forest fires will be improved. The ecosystem, economic system, and social system will be able to respond to the dynamic changes of the external environment and respond in advance to forest fires to make the systems more sustainable.

5. Conclusions

Short-term forest fire danger rating prediction has become a critical problem for avoiding the danger of forest fires and supporting the sustainability of forest ecosystem. We propose a RAFFIA model in this paper to cope with the big data related challenges of short-term forest fire danger rating prediction (retrieve Section 1). On the basis of Multiclass Logistic Regression, the RAFFIA-based prediction approach learns from the samples of labeled forest fire danger ratings for environment-related forest sensor data, including wind, rain, air temperature, air relative humidity, forest litter amount, and underground water amount. Experiments based on an experimental station deployed in Xiashu Forest demonstrate the effectiveness and efficiency of the proposed approach. Under a refined setting of RAFFIA model parameters, the precision can reach 98.74, with RMSE being 0.083. To make a trade-off with time complexity for prediction construction, the precision is 98.71, with RMSE being 0.081, and execution time for the model construction being 1.07 s. Moreover, each RAFFIA prediction needs around 250 ms on average.

It is worth noting that, the short-term forest fire danger rating prediction is extensively needed for forest fire prevention in key areas. These areas are often intensive. Occurrence of fires in these key areas can always lead to significant loss of life and property. The RAFFIA model can be deployed in these key areas to enhance the sustainability of ecosystems as well as economic and social systems.

With the RAFFIA model and the sensor networks, a short-term forest fire warning system can be easily constructed. Using the real-time prediction results, if the fire possibility is high, an early warning could be activated to send to the forest fire manager. The forest fire prevention plan to reduce the fire danger would be activated and performed. In the long run, the application of this model will help to reduce the cost of forest fire prevention for the key forest fire prevention areas, and contribute to the sustainable development.

This paper can be extended in the following future directions:

First, as terrain and human activities are also very important factors influencing the occurrence and spread of forest fires, these factors can also be integrated in the RAFFIA model. Since these factors should always be identified through remote sensing and video monitoring, we plan to extend our model to support the prediction based on multiple types of data sources.

Second, we also plan to investigate other artificial intelligence methods, such as convolutional neural networks (CNNs), deep belief networks (DBNs), etc., to further enhance the effectiveness of forest fire danger rating online prediction.

Author Contributions: L.W. and Z.W. conceived and designed the models; L.W. and Q.Z. performed the experiments; L.W. wrote the paper; L.W., Q.Z., Z.W. and J.Q. revised the paper.

Funding: This work was particularly supported by Humanity and Social Science Youth Fund of Ministry of Education of China (No. 18YJCZH170), Key projects of the National Social Science Fund of China (No. 18AGL017), NSFC projects (Nos. 31361130342, 71373125, 71403122), Philosophy and Social Science Fund of Higher Education in Jiangsu Prov. (No. 2016SJB630009), and Youth Innovation Fund of Science and Technology of NJFU (No. CX2016031).

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

References

- Mithal, V.; Garg, A.; Boriah, S.; Steinbach, M.; Kumar, V.; Potter, C.; Klooster, S.; Castilla-Rubio, J.C. Monitoring global forest cover using data mining. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 36. [[CrossRef](#)]
- Jurvélius, M. HEALTH AND PROTECTION | Forest Fires (Prediction, Prevention, Preparedness and Suppression). In *Encyclopedia of Forest Sciences*; Burley, J., Ed.; Elsevier: Oxford, UK, 2004; pp. 334–339. doi:10.1016/B0-12-145160-7/00277-5.
- Kurz, W.; Apps, M. Developing Canada’s national forest carbon monitoring, accounting and reporting system to meet the reporting requirements of the Kyoto Protocol. *Mitig. Adapt. Strateg. Glob. Chang.* **2006**, *11*, 33–43. [[CrossRef](#)]
- Feikema, P.M.; Sherwin, C.B.; Lane, P.N. Influence of climate, fire severity and forest mortality on predictions of long term streamflow: Potential effect of the 2009 wildfire on Melbournes water supply catchments. *J. Hydrol.* **2013**, *488*, 1–16. doi:10.1016/j.jhydrol.2013.02.001. [[CrossRef](#)]
- de Groot, W.J.; Wotton, B.M.; Flannigan, M.D. Wildland Fire Danger Rating and Early Warning Systems. *Wildfire Hazards Risks Disasters* **2014**, 207–228. [[CrossRef](#)]
- Saoudi, M.; Bounceur, A.; Euler, R.; Kechadi, T. Data Mining Techniques Applied to Wireless Sensor Networks for Early Forest Fire Detection. In *Proceedings of the International Conference on Internet of Things and Cloud Computing*; ACM: New York, NY, USA, 2016; p. 71.
- Beck, J. Equations for the forest fire behaviour tables for Western Australia. *CALM Sci.* **1995**, *1*, 325–348.
- Deeming, J.E.; Burgan, R.E.; Cohen, J.D. The National Fire-Danger Rating System–1978. In *USDA Forest Service General Technical Report INTUS (USA), INT-39*; Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station: Ogden, UT, USA, 1977; p. 63.
- Van Wagner, C.; Forest, P. *Development and Structure of the Canadian Forest Fire Weather Index System*; The Print Shoppe LTO; Canadian Forestry Service: Ottawa, ON, Canada, 1987.
- Beverly, J.L.; Herd, E.P.; Conner, J.R. Modeling fire susceptibility in west central Alberta, Canada. *For. Ecol. Manag.* **2009**, *258*, 1465–1478. [[CrossRef](#)]
- Sharples, J.; McRae, R.; Weber, R.; Gill, A.M. A simple index for assessing fire danger rating. *Environ. Modell. Softw.* **2009**, *24*, 764–774. [[CrossRef](#)]
- Elmas, Ç.; Sönmez, Y. A data fusion framework with novel hybrid algorithm for multi-agent Decision Support System for Forest Fire. *Expert Syst. Appl.* **2011**, *38*, 9225–9236. [[CrossRef](#)]
- Hamadeh, N.; Hilal, A.; Daya, B.; Chauvet, P. Studying the factors affecting the risk of forest fire occurrence and applying neural networks for prediction. In *Proceedings of the SAI Intelligent Systems Conference*, London, UK, 10–11 November 2015; pp. 522–526.
- Sakr, G.E.; Elhajj, I.H.; Mitri, G.; Wejinya, U.C. Artificial intelligence for forest fire prediction. In *Proceedings of the 2010 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, Montreal, ON, Canada, 6–9 July 2010; pp. 1311–1316. doi:10.1109/AIM.2010.5695809. [[CrossRef](#)]
- Chuvieco, E.; Aguado, I.; Yebra, M.; Nieto, H.; Salas, J.; Martín, M.P.; Vilar, L.; Martínez, J.; Martín, S.; Ibarra, P.; et al. Development of a framework for fire risk assessment using remote sensing and geographic information system technologies. *Ecol. Model.* **2010**, *221*, 46–58. [[CrossRef](#)]
- Ichoku, C.; Kaufman, Y.J. A method to derive smoke emission rates from MODIS fire radiative energy measurements. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 2636–2649. [[CrossRef](#)]
- Katurji, M.; Nikolic, J.; Zhong, S.; Pratt, S.; Yu, L.; Heilman, W.E. Application of a statistical emulator to fire emission modeling. *Environ. Model. Softw.* **2015**, *73*, 254–259. doi:10.1016/j.envsoft.2015.08.016. [[CrossRef](#)]
- Kane, V.R.; Lutz, J.A.; Cansler, C.A.; Povak, N.A.; Churchill, D.J.; Smith, D.F.; Kane, J.T.; North, M.P. Water balance and topography predict fire and forest structure patterns. *For. Ecol. Manag.* **2015**, *338*, 1–13. [[CrossRef](#)]
- Pourtaghi, Z.S.; Pourghasemi, H.R.; Aretano, R.; Semeraro, T. Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques. *Ecol. Indic.* **2016**, *64*, 72–84. [[CrossRef](#)]
- Sakr, G.E.; Elhajj, I.H.; Mitri, G. Efficient forest fire occurrence prediction for developing countries using two weather parameters. *Eng. Appl. Artif. Intell.* **2011**, *24*, 888–894. [[CrossRef](#)]
- Tian, X.; Zhao, F.; Shu, L.; Wang, M. Distribution characteristics and the influence factors of forest fires in China. *For. Ecol. Manag.* **2013**, *310*, 460–467. [[CrossRef](#)]

22. Pacheco, A.P.; Claro, J.; Fernandes, P.M.; de Neufville, R.; Oliveira, T.M.; Borges, J.G.; Rodrigues, J.C. Cohesive fire management within an uncertain environment: A review of risk handling and decision support systems. *For. Ecol. Manag.* **2015**, *347*, 1–17. [[CrossRef](#)]
23. Shang, Z.; He, H.S.; Lytle, D.E.; Shifley, S.R.; Crow, T.R. Modeling the long-term effects of fire suppression on central hardwood forests in Missouri Ozarks, using LANDIS. *For. Ecol. Manag.* **2007**, *242*, 776–790. doi:10.1016/j.foreco.2007.02.026. [[CrossRef](#)]
24. Mölders, N. Comparison of Canadian Forest Fire Danger Rating System and National Fire Danger Rating System fire indices derived from Weather Research and Forecasting (WRF) model data for the June 2005 Interior Alaska wildfires. *Atmos. Res.* **2010**, *95*, 290–306. [[CrossRef](#)]
25. Wang, L.; Wang, H.; Yu, Q.; Sun, H.; Bouguettaya, A. Online reliability time series prediction for service-oriented system of systems. In *Service-Oriented Computing*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 421–428.
26. Wang, H.; Wang, L.; Yu, Q.; Zheng, Z.; Bouguettaya, A.; Lyu, M.R. Online Reliability Prediction via Motifs-Based Dynamic Bayesian Networks for Service-Oriented Systems. *IEEE Trans. Softw. Eng.* **2017**, *43*, 556–579. [[CrossRef](#)]
27. Wang, H.; Wang, L.; Yu, Q.; Zheng, Z.; Yang, Z. A proactive approach based on online reliability prediction for adaptation of service-oriented systems. *J. Parallel Distrib. Comput.* **2018**, *114*, 70–84. [[CrossRef](#)]
28. Iliadis, L. A decision support system applying an integrated fuzzy model for long-term forest fire risk estimation. *Environ. Model. Softw.* **2005**, *20*, 613–621. [[CrossRef](#)]
29. Mahmood, A.; Shi, K.; Khatoon, S.; Xiao, M. Data mining techniques for wireless sensor networks: A survey. *Int. J. Distrib. Sens. Netw.* **2013**, *2013*, 1–24. [[CrossRef](#)]
30. Sabri, Y.; Kamoun, N.E. A prototype for wireless sensor networks to the detection of forest fires in large-scale. *Next Gener. Netw. Serv.* **2012**, 116–122. doi:10.1109/NGNS.2012.6656065. [[CrossRef](#)]
31. Salfner, F.; Lenk, M.; Malek, M. A survey of online failure prediction methods. *ACM Comput. Surv.* **2010**, *42*, 10. [[CrossRef](#)]
32. Wang, H.; Wang, L.; Yu, C. Integrating trust with qualitative and quantitative preference for service selection. In Proceedings of the IEEE International Conference on Services Computing, Anchorage, AK, USA, 27 June–2 July 2014; pp. 299–306.
33. Wang, H.; Yu, C.; Wang, L.; Yu, Q. Effective BigData-Space Service Selection over Trust and Heterogeneous QoS Preferences. *IEEE Trans. Serv. Comput.* **2018**, *11*, 644–657. [[CrossRef](#)]
34. Zeng, L.; Benatallah, B.; Ngu, A.H.; Dumas, M.; Kalagnanam, J.; Chang, H. Qos-aware middleware for web services composition. *IEEE Trans. Softw. Eng.* **2004**, *30*, 311–327. [[CrossRef](#)]
35. Bishop, C. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006.
36. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
37. Saaty, T.L. Decision making with the analytic hierarchy process. *Int. J. Serv. Sci.* **2008**, *1*, 83–98. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).