Journal of Environmental Management 237 (2019) 476–487

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: www.elsevier.com/locate/jenvman

Research article

Spatial pattern analysis and prediction of forest fire using new machine learning approach of Multivariate Adaptive Regression Splines and Differential Flower Pollination optimization: A case study at Lao Cai province (Viet Nam)

Dieu Tien Bui[a,b], Nhat-Duc Hoang[c], Pijush Samui[d,e,*]

[a] Institute of Research and Development, Duy Tan University, Da Nang, 550000, Viet Nam
[b] Geographic Information System Group, Department of Business and IT, University of South-Eastern Norway, N-3800, Bø i Telemark, Norway
[c] Faculty of Civil Engineering, Institute of Research and Development, Duy Tan University, Da Nang, 550000, Viet Nam
[d] Geographic Information Science Research Group, Ton Duc Thang University, Ho Chi Minh City, Viet Nam
[e] Faculty of Environment and Labour Safety, Ton Duc Thang University, Ho Chi Minh City, Viet Nam

ARTICLE INFO

Keywords:
Forest fire susceptibility
GIS
Multivariate adaptive regression splines
Differential Flower Pollination
Hybrid computational intelligence

ABSTRACT

Understanding spatial patterns of forest fire is of key important for fire danger management and ecological implication. This aim of this study was to propose a new machine learning methodology for analyzing and predicting spatial patterns of forest fire danger with a case study of tropical forest fire at Lao Cai province (Vietnam). For this purpose, a Geographical Information System (GIS) database for the study area was established, including ten influencing factors (slope, aspect, elevation, land use, distance to road, normalized difference vegetation index, rainfall, temperature, wind speed, and humidity) and 257 fire locations. The relevance level of these factors with the forest fire was analyzed and assessed using the Mutual Information algorithm. Then, a new hybrid artificial intelligence model named as MARS-DFP, which was Multivariate Adaptive Regression Splines (MARS) optimized by Differential Flower Pollination (DFP), was proposed and used construct forest fire model for generating spatial patterns of forest fire. MARS is employed to build the forest fire model for generalizing a classification boundary that distinguishes fire and non-fire areas, whereas DFP, a metaheuristic approach, was utilized to optimize the model. Finally, global prediction performance of the model was assessed using Area Under the curve (AUC), Classification Accuracy Rate (CAR), Wilcoxon signed-rank test, and various statistical indices. The result demonstrated that the predictive performance of the MARS-DFP model was high (AUC = 0.91 and CAR = 86.57%) and better to those of other benchmark methods, Backpropagation Artificial Neural Network, Adaptive neuro fuzzy inference system, Radial Basis Function Neural Network. This fact confirms that the newly constructed MARS-DFP model is a promising alternative for spatial prediction of forest fire susceptibility.

1. Introduction

In the natural world, forest fires seem to be inevitable and they play an important role in vegetation succession and landscape transformation (Chuvieco et al., 2010; Wang et al., 2017). Nevertheless, uncontrolled forest fires may bring about negative impacts on the environment and the local communities. It is because these fires do not only damage human life and properties but also threaten the stability of ecosystems (Rajan and Shanmugam, 2018). In the past decade, there has been an increasing trend in both number and severity of forest fires occurred across the globe (Catry et al., 2010; Jaafari et al., 2017; Robinne et al., 2016). This notable trend is spurring public concerns about the ecological and socioeconomic impacts of forest fires (Molina et al., 2018; Valdez et al., 2017).

According to a report of the Ministry of Agriculture and Rural Development in 2016, Vietnam has a forest cover of about 40% of its geographical area. In recent years, due to climate change and expanded economic activities of human, forest fires in this country have become a serious natural hazard that destroyed vast amounts of natural resources, degraded the soil, and caused air pollution. Besides the activities of human in land use altering, prolonged dry weather with exceptionally high temperature increases the number of fires in a large number of

* Corresponding author. Geographic Information Science Research Group, Ton Duc Thang University, Ho Chi Minh City, Viet Nam.
E-mail addresses: Dieu.T.Bui@usn.no (D. Tien Bui), pijush.samui@tdtu.edu.vn (P. Samui).

https://doi.org/10.1016/j.jenvman.2019.01.108
Received 12 August 2018; Received in revised form 21 January 2019; Accepted 27 January 2019
Available online 27 February 2019
0301-4797/

provinces of Vietnam. Moreover, these phenomena are also observed in many other counties (Herawati and Santoso, 2011; Marchal et al., 2017; Schweizer and Cisneros, 2017; Wu et al., 2015; Yao et al., 2018).

Due to the significant impact of forest fires on the ecosystems and the socio-economic conditions, the prevention and suppression of forest fires have become a common interest of governments and researchers around the world (Nami et al., 2018). To establish an effective fire prevention and suppression plans, it is necessary to construct the fire susceptibility maps at the regional scale. The reason is that such maps do not only facilitate the reasonable allocation of resources needed for fire prevention and suppression but also tremendously support the tasks of land use planning (Bax and Francesconi, 2018).

Recent advancements in geographic information system (GIS) immensely have supported the task of constructing forest fire susceptibility maps. GIS allows the tasks of capturing, analyzing, managing, and presenting geographic data to be performed conveniently. Since multiple layers of information can be displayed on a single GIS-based map, the forest fire danger in a region can be assessed with the consideration of various influencing factors including climate, vegetation, topography, and human activities (Nami et al., 2018). The spatial relationships between those factors and the regional historical fire inventory can be used to construct data-driven models which perform accurate fire susceptibility predictions for all areas in the region (Tien Bui et al., 2017c).

Given the complex and the multivariate nature of forest fire susceptibility prediction, various authors have resorted to computational intelligence as a means of data analysis. The integration of GIS and computational intelligence technique has been proven to be an effective method for identifying patterns and discovering knowledge from GIS databases (Correia et al., 2017; Duarte et al., 2017; Jaafari et al., 2018; Nami et al., 2018; Teodoro et al., 2015). The spatial prediction of forest fire susceptibility is often modeled as a two-class pattern recognition problem. A machine learning or a data mining algorithm is employed by various scholars to generalize a classification boundary that separates the pixels in a map into two categories: fire and non-fire (Hong et al., 2017). Based on performance comparison, it has been found that computational intelligence methods are capable of delivering much more accurate prediction results than those of the conventional statistical approaches (Pham et al., 2018; Razavi Termeh et al., 2018; Tien Bui et al., 2017b).

Decision tree (Jaafari et al., 2018; Lozano et al., 2008) and its ensemble strategy of Random Forest (RF) (Arpaci et al., 2014; Leuenberger et al., 2018; Oliveira et al., 2017) have been applied to predict forest fire susceptibility as well as forest fire occurrence probability with good performance. Silva et al. (2015) investigated the applicability of Bayesian modeling approach for modeling the proportion of burned forest area; Markov chain Monte Carlo methods were used to identify the model parameters.

Neural network (Satir et al., 2016) and fuzzy neural network (Tien Bui et al., 2017c), with their capabilities of nonlinear modeling, have been employed to construct classification model for spatial forest fire susceptibility prediction. A new variant of the neural network named Extreme Learning Machine has also been used for the task of interest (Leuenberger et al., 2018). Pourghasemi et al. (2016) established Mamdani fuzzy logic models and compared their predictive inference with a modified analytical hierarchy process. Tien Bui et al. (2016a) integrated GIS and Kernel Logistic Regression to perform forest fire susceptibility mapping in a tropical region. A random forest classifier and a Genetic Algorithm optimized Support Vector Machine (SVM) model has recently been utilized to establish forest fire susceptibility maps (Hong et al., 2018); this research found that random forest classifier outperformed SVM.

From the literature reviewed, it can be observed that there is an increasing trend of applying computational intelligence approaches for analyzing GIS data sets at a regional scale. These approaches have demonstrated their capabilities in establishing accurate classification models used for forest fire mapping. Following this research trend, the current study explores the performance of Multivariate Adaptive Regression Splines (MARS) (Friedman, 1991) in spatial forest fire susceptibility modeling.

Spatial modeling of forest fire susceptibility is inherently complex that requires the analysis of various conditioning factors (Tien Bui et al., 2017c). Therefore, MARS can be a suitable approach to approximate the decision boundary that separates non-fire and fire data samples. It is because the objective of forest fire susceptibility modeling is to determine a classification model that both attains good prediction accuracy and can avoid overfitting. MARS deals with complex modeling tasks by dividing the learning space of forest fire conditioning factors into sub-domains; the data in each sub-domain is fitted by a linear model. In generally, this classifier constructs a piecewise linear function to represent a global model with an adaptive manner. Moreover, MARS employs an automatic approach of overfitting prevention by a backward operation for casting out redundant model's components (Yilmaz et al., 2018).

This computational intelligence algorithm relies on a divide-and-conquer strategy within which the data set are divided into distinctive regions; each is handled by a linear model (García-Nieto et al., 2016; Goh et al., 2016; Hoang et al., 2017; Lokuge et al., 2018). Therefore, MARS can be an appropriate method for capturing knowledge from a highly complex GIS database. Nevertheless, the capability of this machine learning algorithm has rarely been explored in spatial forest fire danger mapping. Our current research is an attempt to fill this gap in the literature.

It is noted that the learning and predictive performances of MARS depends on the setting of the two tuning parameters, namely the maximum number of basis functions (*BFs*) and the penalty coefficient (*c*). Based on previous works (Li et al., 2017; Liao et al., 2018; Moayed et al., 2017), the task of parameter setting can be modeled as an optimization problem in continuous domains. Thus, this study proposes a hybridization of Differential Flower Pollination (DFP) and MARS to construct a self-tuned intelligent model used for forest fire susceptibility mapping. The newly constructed model is trained and validated with a GIS database collected in Lao Cai province (Vietnam).

The rest of this study is organized as follows: The second section describes the study area and the GIS database, followed by the third section that reviews the employed computational intelligence approaches. The next section provides the description of the proposed hybrid model for forest fires susceptibility prediction. The fifth section reports the experimental results and comparison. Several conclusions of this study are stated in the final section.

2. The study area and the GIS database

2.1. General description of Lao Cai province

The study area of this study lies in Lao Cai province located in the mountainous Northwest region of Vietnam (see Fig. 1). Lao Cai is between the longitudes of 103° 32′ E − 104° 05′ E and the latitudes of 22° 08′ N - 22° 48′ N. The altitude ranges between 200 m at the Thao river valley and over 3000 m above sea level on the strip of the Hoang Lien Son mountain (Tien Bui et al., 2016c). This study area covers an area of 2253 km^2. Forest accounts for roughly 44% of the total area with roughly 279,000 ha. Lao Cai possesses a considerable area of fertile land which is suitable for medicinal plants, fruit trees, and vegetables. About 11% of the area is humus soil on the mountain which features rich canopies of mixed forest. Approximately 2% of the area is used for rice planting in terraced fields. The total annual rainfall of the study area ranges from 2000 to 3600 mm. The rainfall intensifies in June, July, and August accounting for 80–85% of the total annual rainfall. The temperature of the province ranges between 18 °C and 28 °C with the average value of 23 °C.

In recent years, due to radical climate change and human activities,

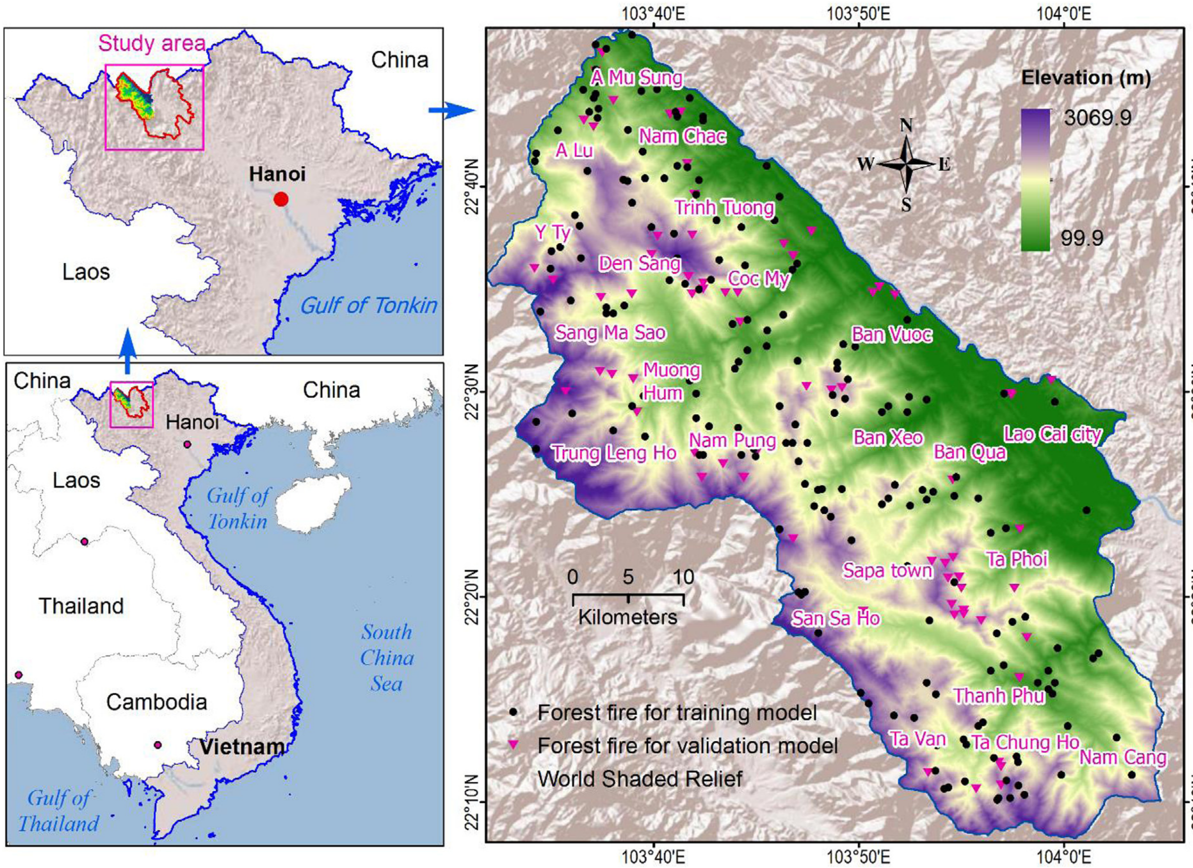

Fig. 1. Location of the study area and forest fire inventories.

there is an increasing trend of fire number. Accordingly, a large area of natural forest in the study area has been subject to fire. The heat wave and dry winds appearing in the dry season make fire extinguishment a challenging task. Based on the local hydrometeorology station, the lack of rainfall in the dry season caused a severe water shortage and brought about a high risk of forest fire (Vietnamplus, 2015). In 2016, based on a local government's report, about 8 ha of forest in the nature reserve area was destroyed in a single fire (Vietnamnews, 2016).

2.2. Historical forest fire events and ignition factors

2.2.1. Historical forest fire events

In this research, information on 257 fire locations that occurred from January 2008 to May 2016 is used as a database of historical forest fires. These fire locations have been retrieved from the national forest fire database produced by the Department of Forest Protection of Vietnam (Ministry of Agriculture and Rural Development of Vietnam, 2016). Most of the fires in the studied area occurred in months from February to May, September, and November (Fig. 2). Among the 257 historical fire locations, 187 locations were randomly selected (Bisquert et al., 2012b) used for model training and the remaining, 70 locations, were employed for model validation.

2.2.2. Forest fire ignition factors and the GIS database

After the inventory map has been established, the next step of forest fire susceptibility modeling is to select a set of fire influencing factors. Based on the findings of previous works, the likelihood of forest fire occurrence and its spread are highly influenced by many factors. These influencing factors can be divided into four main groups: climate, vegetation, topography, and human activities (Nami et al., 2018). Based on literature reviewed (Hong et al., 2018; Nami et al., 2018; Pourghasemi et al., 2016; Sahana and Ganaie, 2017; Tien Bui et al., 2016a, 2017c; Verde and Zêzere, 2010) and the availability of data sources, ten factors are selected to predict the forest fire susceptibility for this study.

The factors of slope, aspect, elevation, curvature, and distance to road were extracted from topographic maps provided by the Ministry of Natural Resources and Environment of Vietnam. The factors of normalized difference vegetation index (NDVI) was computed from the Landsat-8 OLI. In addition, the four climatic factors of rainfall (average

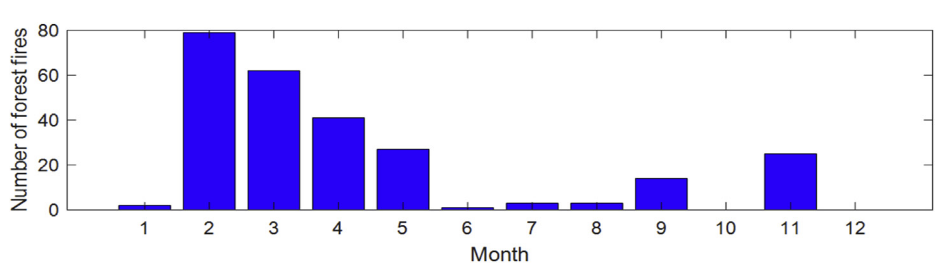

Fig. 2. Temporal distribution of the forest fires in the study area.

monthly rainfall), temperature (maximum monthly temperature), wind speed (average monthly wind speed), and humidity (average monthly relative humidity) (Tehrany et al., 2018), which were acquired by analyzing weather data from 2008 to 2016 collected from the Climate Forecast System Reanalysis (https://www.ncdc.noaa.gov/), were considered. It should be noted that only the weather data in months with the most forest fires (February to May, September, and November) were used.

Topographic factors of slope, aspect, elevation, and curvature influences the forest fire occurrence by determining local climate, vegetation characteristics, and human accessibility (Pourtaghi et al., 2016). Vegetation reflected by NDVI is a crucial factor since it strongly affects the fire ignition and the spread of fire through fuel characteristics such as types of plan and moisture content (Tien Bui et al., 2017c). The influence of human activities on forest fire can also be expressed through the factors of distance to road and land use (Parisien et al., 2016). The reason is that local people may alter the pattern and frequency of fires by creating ignition sources and changing the vegetation conditions; these facts may significantly enhance the likelihood of forest fire. In addition, it is obvious that climatic variables including temperature, precipitation, wind, and evapotranspiration have both direct and indirect influences on ignitability (Hong et al., 2018). The spatial distributions of the ten forest fire influencing variables are demonstrated in Fig. 3.

It is noted that a Digital Elevation Model (DEM) constructed for the study area has been created by the employment of topographic maps with the scale of 1:50,000. The topographic maps are provided by the Ministry of Natural Resources and Environment of Vietnam. It is also proper to note that the interval terrain contour of these maps is 5 m in areas with slopes less than 2°, 10 m in areas with slopes from 2° to 15°, and 20 m in other areas (Tien Bui et al., 2017a).

Furthermore, to compute the factor of NDVI, the following equation has been used (Tien Bui et al., 2017c):

$$\mathrm{NDVI} = (\mathrm{NIB} - \mathrm{REB})/(\mathrm{NIB} + \mathrm{REB}) \tag{1}$$

where NIB and REB are the near-infrared band (Band 4, 0.76–0.90 μm) and the red band (Band 3, 0.63–0.69 μm), respectively.

The factor of land use, reflected in the land use map at the scale of 1:50,000, has been constructed by retrieving information from the national land use map in 2015. This national land use map has been provided by the local authority. Moreover, the factor of distance to road was established by buffering the road network which is obtained from the national topographic map at the scale of 1:50,000. In addition, the climatic factors of temperature, wind, and rainfall in the study area were acquired from the Ministry of Natural Resources and Environment of Vietnam. Based on the records in the forest fire inventories, climatic data has been selected with the focus on the months featuring a high number of fire occurrences (as illustrated in Fig. 2 above).

As mentioned earlier, the data related to previous occurrences of forest fire are employed to generate a forest fire inventory map of the study area. This inventory map and the aforementioned ten factors are used by computational models to create a classification function that provides the status of fire/non-fire for each pixel in the map. Based on historical records of the Department of Forest Protection, a GIS database of the Lao Cai province (see Fig. 4) has been established with 257 pixels within which 187 pixels are utilized for the model construction phase and 70 pixels are employed for model testing phase.

3. Background of computational intelligence algorithms

3.1. Multivariate Adaptive Regression Splines (MARS)

MARS, put forward by Friedman (1991), is a non-parametric approach for learning a function that describes a mapping between a set of predictor X and an output Y. The model constructed by MARS requires no assumptions regarding the distribution of the involved variables. MARS is specifically useful for the circumstance in which the mapping function differs for sub-groups of the collected data (York et al., 2006). Previous works in various study regions (Hong et al., 2018; Satir et al., 2016; Tien Bui et al., 2017c) demonstrate that the classification boundaries used for spatial forest fire modeling are indeed complex. Hence, MARS can be a good alternative for learning such classification functions. This machine learning approach divides the learning space of forest fire influencing variables into smaller domains (see Fig. 5). In each divided domain, MARS employs a linear classification model to fit the data subset.

In overall, MARS constructs a piecewise linear function for characterizing the global model with an adaptive manner (Hoang et al., 2017). Good performance of this data modeling approach has been observed in various fields (Ferreira-Santiago et al., 2016; Haghiabi, 2017; Liao et al., 2018). The learning phase of the final prediction model discovered by MARS is accomplished through the identification of a set of basis functions (*BFs*). These *BFs* describe the relationship between forest fire affecting variables and target output of fire status. A typical form of a *BF* is described as follows:

$$b_m(x) = \max(0, C - x) \, or b_m(x) = \max(0, x - C) \tag{2}$$

where b_m denotes a *BF*; x represents a forest fire susceptibility predictor; C denotes a threshold parameter which governs the process that separates the original range of x into sub-ranges.

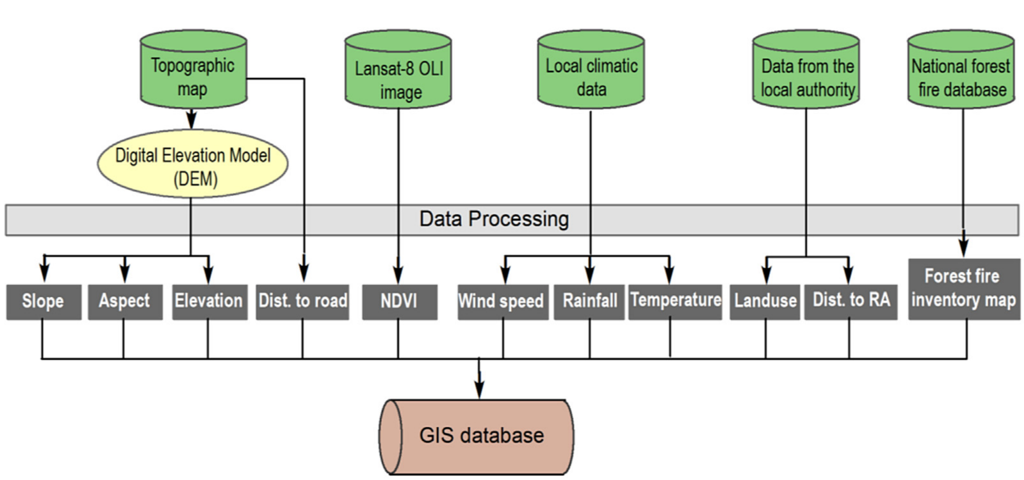

Fig. 3. Forest fire influencing factors: (a) Slope; (b) Elevation; (c) Aspect; (d) Land use; (e) Distance to residential areas; (f) NDVI; (g) Rainfall; (h) Temperature (i) Wind speed; (j) Humidity.

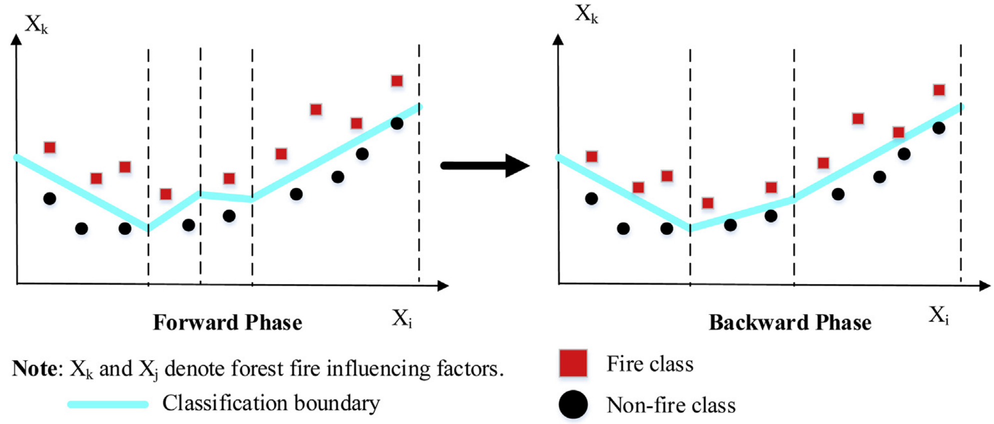

Fig. 4. The established GIS database for the study area.

The final functional form of the MARS model used in forest fire susceptibility prediction is given as follows:

$$f(x) = sign(\alpha_0 + \sum_{m=1}^{k} \alpha_m b_m(x)) \tag{3}$$

where $\alpha_0, \alpha_1, ..., \alpha_M$ denote weighting coefficients; $f(x)$ is the model output which is −1 for the case of 'non-fire' class and 1 for the case of 'fire' class. *k* represents the number of weighting coefficients.

It is noted that the model training process of MARS is accomplished in two phases: forward and backward phases (see Fig. 5). In the first phase, *BFs* are added into the model to minimize the prediction error; this phase stops when the maximum number of *BF* is reached. The second phase, aiming at diminishing the risk of overfitting, removes redundant *BFs*. The MARS model is therefore simplified after the second phase. In addition, the quality of each sub-model discovered by MARS is evaluated by the generalized cross-validation (GCV) index given as follows (Jekabsons, 2016; Suman et al., 2016):

$$\mathrm{GCV} = \mathrm{MSE}/(1 - \frac{k + 0.5c(k-1)}{n})^2 \tag{4}$$

where MSE is the mean square error of the model, *k* denotes the number of *BFs*. *n* represents the number of samples in the training data. *c* is the penalty coefficient; Friedman (1991) and Jekabsons (2016) suggested that the appropriate value of the parameter *c* should be in the range of [2, 4]. Besides *c*, the number of *BFs* is also a sensitive parameter of MARS. In this study, both c and *BFs* are adaptively determined by the Differential Flower Pollination metaheuristic algorithm.

3.2. Differential Flower Pollination (DFP) for global optimization

In recent years, metaheuristic has become a popular method for solving complex optimization problems. This method relies on the collective intelligence of a population of agents to identify a near optimal solution (Yang, 2014). In many applications, metaheuristic algorithms have demonstrated satisfactory optimization performances which are superior to those of conventional approaches (Chatterjee and Chowdhury, 2017; Crawford et al., 2017). The determination of tuning parameters for machine learning model, also called model selection, is a challenging task. One reason is that it is very difficult to obtain the landscape feature of the objective function of the optimization problem.

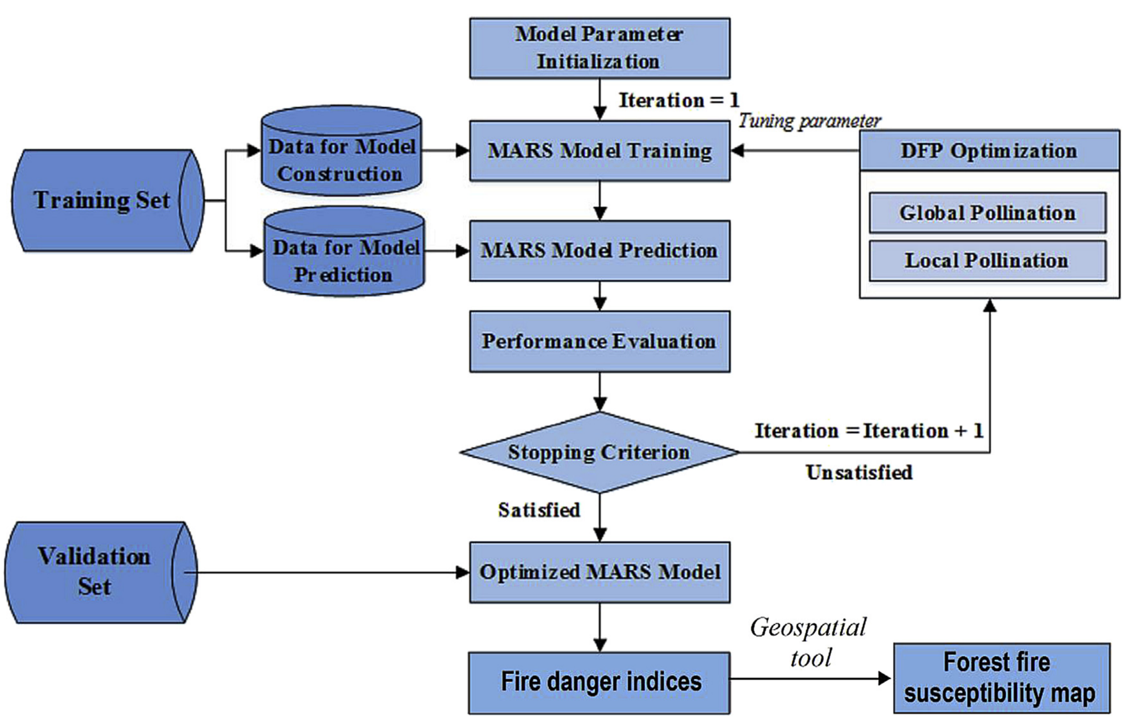

Fig. 5. Demonstration of the MARS learning process.

Another reason is that the tuning parameters are often searched in continuous domains; this means there is an infinite number of parameter sets. Accordingly, metaheuristic methods have been successfully employed in the task of model selection (Prayogo and Susanto, 2018; Tien Bui et al., 2016b; Xue, 2017).

Proposed by Hoang et al. (2016), Differential Flower Pollination (DFP) is a metaheuristic that combines two individual optimization algorithms of the Differential Evolution (DE) (Storn and Price, 1997) and the Flower Pollination Algorithm (FPA) (Yang, 2012). DFP possesses the exploitative search based on mutation-crossover operators of DE and the explorative operation based on the Levy flight of FPA. The mutation-crossover operators of DE replace the local pollination process of the standard FPA. Meanwhile, the Levy-flight based global pollination is preserved by DFP to facilitate the exploration stage with large searching step size. As shown by experiments, the DFP approach can achieve desired optimization results (Hoang et al., 2016).

The global pollination operator of DFP is given as follows:

$$X_i^{trial} = X_i^g + L.\,(X_i^g - X_{best}) \tag{5}$$

where g denotes the current number of generations. X_i^{trial} is a trial solution.

The mathematical formula of the local pollination process is provided as follows:

$$\text{Mutatedsolution:}X_{i,g}^{mutated} = x_{r1,g} + F.\,(x_{r2,g} - x_{r3,g}) \tag{6}$$

where $r1$, $r2$, and $r3$ are three random indices. F denotes the parameter of a mutation scale factor. As recommended in previous works (Hoang et al., 2016; Zhang and Sanderson, 2009), F is generated from a Normal distribution with the mean = 0.5 and the standard deviation = 0.15.

$$\text{Crossedflower:}X_{j,i,g+1}^{crossed} = \begin{cases} X_{j,i}^{mutated}, & if\ r\ and_j \le Cr\ or\ j = rnb(i) \\ X_{j,i,g}, & if\ r\ and_j > Cr\ and\ j \neq rnb(i) \end{cases} \tag{7}$$

where Cr is the crossover probability. This parameter is generally set to be 0.8 as suggested by Price et al. (2005).

4. The proposed MARS-DFP method for forest spatial pattern analysis and prediction

The overall structure of the proposed hybrid model of MARS and DFP (denoted as MARS-DFP) for spatial prediction of forest fires susceptibility is presented in this section of the study. MARS-DFP is a combination of MARS as a machine learning approach and DFP as a metaheuristic optimization algorithm. The hybrid model relies on MARS to construct a classification boundary that assigns the pixel in the map of the study area to either 'fire' or 'non-fire' class. It is noted that MARS is implemented by the toolbox developed by Jekabsons (2016). DFP optimization algorithm has been programmed by the author in MATLAB environment. The data used for model construction and verification was acquired, processed, and compiled by ArcGIS 10.4 and IDRISI Selva 16 packages. The model structure is displayed in Fig. 6.

4.1. Data establishment

Before the model construction phase, the whole data set is temporal separated into the training set (187 fire locations or 72.8%) and the validation set (70 fire locations or 27.2%), also called the testing set. The first set consists fire locations occurred from January 2008 to November 2012 and the second set contains fire locations occurred from February 2013 to May 2016. The training set is then further divided into two subsets: the data for model construction (80%) and the data for model prediction (20%).

Furthermore, it is noted that 257 pixels in the map are locations of fire events. To constitute the complete data set used for model training, it is necessary to randomly sample the equal number of non-fire pixels. We selected areas with NDVI lower than 0, i.e. water surfaces and bare lands to sample these non-fire pixels as suggested in (Tien Bui et al., 2017b). As a result, the total number of data points in the data set is 514. Ten aforementioned variables (Fig. 3, see the supplementary material) are used as input information. The model output is −1 for non-fire class and +1 for fire class. Moreover, a data normalization process should be carried out to negate the effect of unbalanced magnitude (Hoang and Tien Bui, 2016). In this study, the ten influencing factors were converted from categorical classes into continuous domains within the range of 0.01 and 0.99 by the method described in the previous work of Tien Bui et al. (2015).

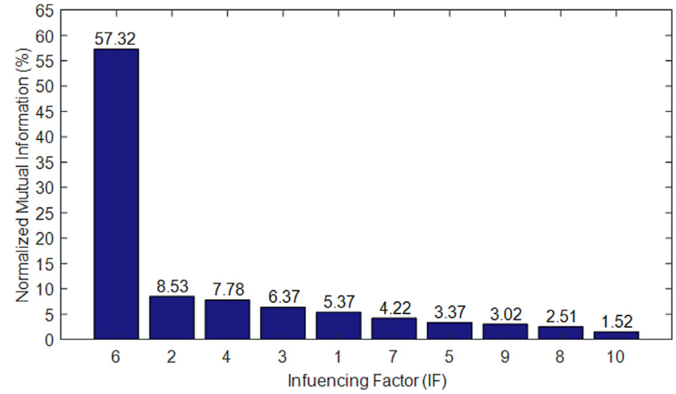

Fig. 6. The proposed MARS-DFP model for the forest fire spatial modeling.

In the next step, the relevancy of influencing factors is preliminarily investigated to confirm their usefulness in forest fire susceptibility mapping. Feature examination is also an important task since it can reveal potentially redundant variables (Emary et al., 2016). In this study, the mutual information (MI) approach, described in Peng et al. (2005), is used for feature examination. This feature investigation approach is selected due to its acceptable analysis result and low computational cost. With the MI method, the feature relevancy is expressed by the conditional mutual information between the feature and the class output (Cheng and Hoang, 2015). This analysis approach is based on probability and information theories and is capable of exhibiting conditional dependencies between influencing factors and yielding a unified view on the importance of factors used forest fire susceptibility assessment.

4.2. Prediction performance evaluation

It is noted that the problem of forest fire susceptibility prediction in this study is modeled as a two-class recognition task. Thus, to appraise the recognition capability, the true positive rate TPR (the percentage of positive instances correctly classified), the false positive rate FPR (the percentage of negative instances misclassified), the false negative rate FNR (the percentage of positive instances misclassified), and the true negative rate TNR (the percentage of negative instances correctly classified) are often employed (López et al., 2013). The four rates are computed in the following manner:

$$TPR = \frac{TP}{TP + FN};\ FPR = \frac{FP}{FP + TN};\ FNR = \frac{FN}{TP + FN};\ TNR = \frac{TN}{TN + FP} \tag{8}$$

where TP, TN, FP, and FN are the values of true positive, true negative, false positive, and false negative, respectively.

In addition, the overall predictive ability of the forest fire susceptibility model can be quantified by the Receiver operating characteristic (ROC) curve that is drawn by the results of sensitivity (true positivity rate) and specificity (false negative rate) (Bradley, 1997). The Area under the Curve (AUC) that varies between 0.5 and 1 is often calculated to express the classification result. It is noted that AUC values of 0.5–0.6 imply insufficient whereas values of 0.6–0.7 indicate poor performance.

Predictors with AUC values of 0.7–0.8 have moderate performance while predictors with AUC values 0.7–0.8 possess good performance. Classification models with AUC values of 0.8–0.9 demonstrate very good performance (Peterson et al., 2008). Besides AUC, the Classification Accuracy Rate (CAR) can also be used to express the proportion of correctly classified data instances.

4.3. The objective function of the model construction phase

As presented earlier, the prediction capability of the MARS model is affected by the maximum number of basis functions (*BFs*) and the penalty coefficient (*c*). Hence, these two parameters should be meticulously specified. In this study, the DFP optimization is employed to search for appropriate values of *BFs* and *c* in order to optimize the classification outcome of MARS. Accordingly, the fitness of the MARS model is measured using our proposed cost function (Eq. (9)). To do so, the training data set was separated into two groups, Group 1 (80%) was used for the construction of the MARS model and Group 2 (20%) that was employed to test the MARS model (see Fig. 6).

The aim of this separation of data set is to avoid the risk of the overfitting problem. In machine learning, overfitting easily occurs if the training model is fitted too well but the model performance in predicting novel data is too poor. This is because the classification result of the training set is not sufficiently good to express the model generalization; the employment of the novel data stored in the Group 2 is necessary to safeguard against overfitting (Hoang et al., 2016). Accordingly, the following cost function should be minimized to obtain a good set of *BFs* and *c*:

$$f = \frac{1}{AUC_{Group1} + AUC_{Group2}} \tag{9}$$

where AUC_{Group1} and AUC_{Group2} are AUC values obtained from the prediction results of the data in the Group 1 and Group 2, respectively.

4.4. The training and validating phases of the MARS-DFP model

With the defined the cost function, the DFP optimization can commence within the MARS training process. Since the number of searched variables = 2 which is not large, the total number of DFP generations G_{MAX} is determined to be 100. In the first generation, initial values of the tuning parameters (*BFs* and *c*) are randomly generated within the range of lower and upper bounds in the following manner:

$$X_{i,0} = LB + R_{(0,1)} \times (UB - LB) \tag{10}$$

where $X_{i,0}$ is the variables to be searched *i* at the first generation. $R_{(0,1)}$ represents a uniformly distributed random number between 0 and 1. *LB* and *UB* denote two vectors of lower and upper bounds for parameters. The lower boundaries of *BFs* and *c* are 5 and 0.1, respectively. The upper boundaries of *BFs* and *c* are 100 and 10, respectively.

During the optimization process, the DFP algorithm guides its population to examine a large number of the tuned parameter sets. At each generation, the cost function of each tuning parameter set is computed by Eq. (9). The metaheuristic approach removes inferior individuals and allows robust individuals to survive in the next generations. Eventually, better sets of *BFs* and *c* solution can be identified. With the optimized values of tuning parameters, the MARS-DFP model is re-trained with all data points in the training dataset to establish the final MARS-DFP prediction model and the forest fire susceptibility prediction results of the validation dataset can be attained. The final model is then employed to compute the forest fire index for all the pixels in the map. The forest fire indices are converted to the raster format by a geospatial tool developed by the authors and opened in the ArcGIS 10.4. Then, the forest fire map is generated with five categories using the standard deviation method available in ArcGIS 10.4.

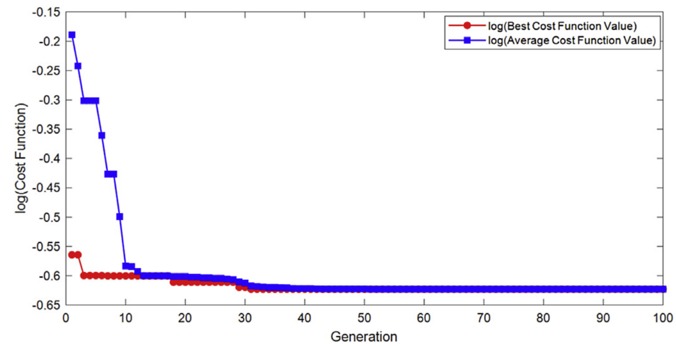

Fig. 7. Factor analysis results (IF_1: Slope, IF_2: Aspect, IF_3: Elevation, IF_4: Land use, IF_5: Distance to road, IF_6: NDVI, IF_7: Rainfall, IF_8: Temperature, IF_9: Wind speed, IF_{10}: Humidity).

4.5. Model comparison

To demonstrate the capability of the newly constructed MARS-DFP, the performance of this model is benchmarked with three other machine learning models: Backpropagation Artificial Neural Network (BPANN) (Hagan et al., 1996), Adaptive neuro fuzzy inference system (ANFIS) (Jang et al., 1997), Radial Basis Function Neural Network (RBFANN) (Chen et al., 1991), Random Forest (Breiman, 2001; Ho, 1995), and MARS (Friedman, 1991) without DFP optimization.

BPANN (Bisquert et al., 2012a; Satir et al., 2016) and ANFIS have been successfully employed for forest fire spatial modeling (Tien Bui et al., 2017c). RBFANN was also demonstrated to be a capable tool of nonlinear modeling (Hoang and Tien Bui, 2017; Lee and Kwak, 2016). In addition, RF and MARS have also been successfully applied in forest fire modeling in other study areas (Bar Massada et al., 2013; Boulanger et al., 2014; Rodrigues and de la Riva, 2014; West et al., 2016).

Notably, it is difficult to employ the DFP metaheuristic to optimize the selection of hyper-parameters for the machine learning approaches of BPANN, ANFIS, and RBFANN. It is because the BPANN and ANFIS models are trained by the gradient descent based backpropagation algorithms (Jang et al., 1997). RBFANN requires the selection the center vectors of the radial basis functions in the hidden layer (Bishop, 2011); the determination of the center vectors is achieved via unsupervised approach of clustering algorithms. Therefore, the performance of BPANN, ANFIS, and RBFANN corresponding to a specific set of hyper-parameters is not consistent due to the randomness in the model construction phase. This inconsistency causes difficulty for the optimization process of DFP because one solution may have different values of the cost function.

It is noted that the implementations of the machine learning algorithms require the appropriate setting of several tuning parameters. To select the model's free parameters, the training data set is also separated into the groups: data for model construction (Group 1) and data for model prediction (Group 2). This procedure of data separation of the benchmark models is similar to that of the MARS-DFP model. The tuning parameters of BPANN, ANFIS, and RBFANN are set based on the prediction performances (AUC) with the data in the Group 2.

In the case of the BPANN model, the log-sigmoid function is chosen as the activation function and the maximum number of training epochs is fixed to be 1000. The number of neurons in the hidden layer and the learning rate should be determined. The value of neurons in the hidden layer ranging from 5 to 20 is examined. The values of the learning rate (0.001, 0.01, 0.1, and 1) are tried to identify the proper one. Base on the experimental result with the data set at hand, the suitable number of neurons and the learning rate are 16 and 0.01, respectively.

The ANFIS model structure is initially created by the fuzzy c-means clustering algorithm (MathWorks, 2012). As suggested by Jang et al. (1997), the Gaussian fuzzy membership function is used in the fuzzification process. In addition, the subtractive clustering algorithm is

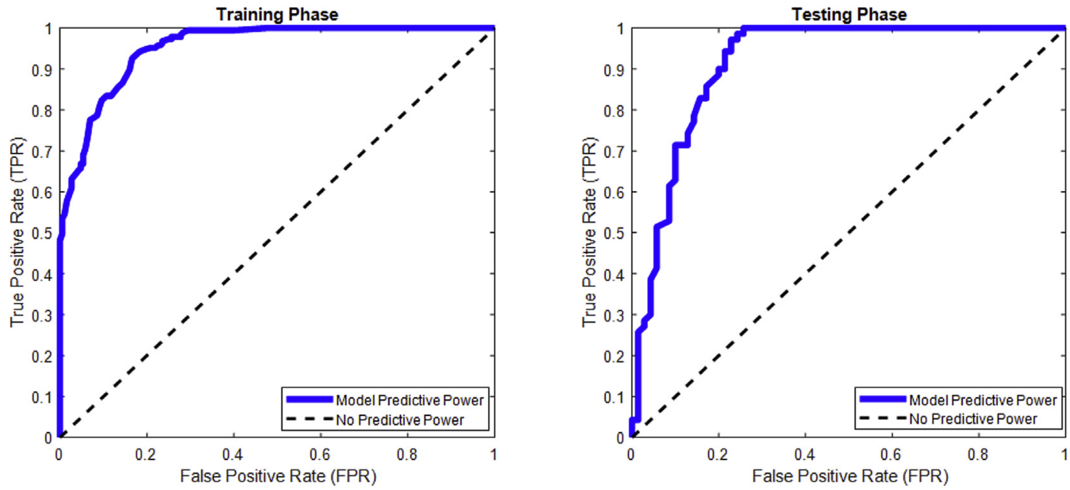

Fig. 8. Optimization process of the MARS-DFP model.

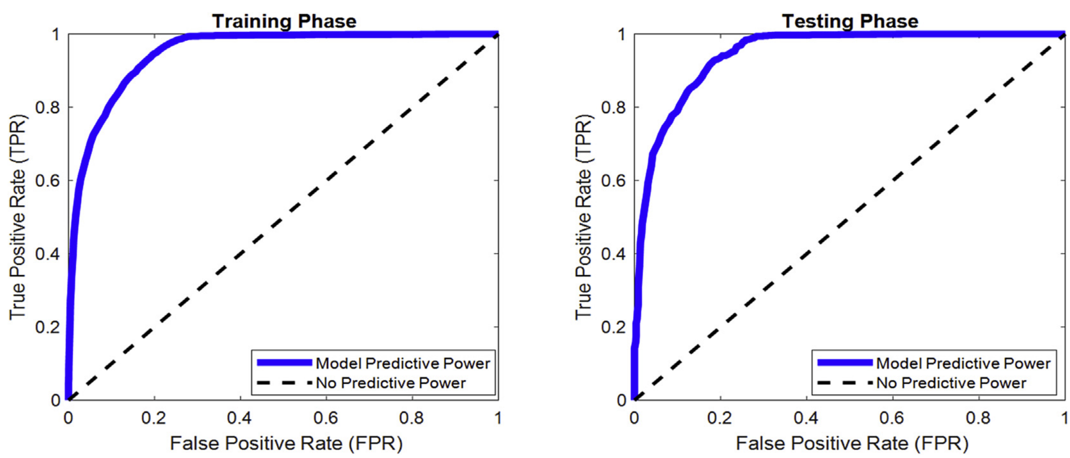

Fig. 9. ROCs of training and testing performances of MARS-DFP.

Table 1
Training and validation performance comparison.

Phase	Performance	Prediction Models					
		MARS-DFP	BPANN	ANFIS	RBFANN	RF	MARS
Training	AUC	0.95	0.96	0.96	0.94	0.86	0.96
	CAR (%)	87.7	92.25	90.91	87.17	85.56	89.72
	TPR	0.95	0.89	0.93	0.9	1.00	0.95
	FPR	0.19	0.05	0.11	0.16	0.29	0.16
	FNR	0.05	0.11	0.07	0.1	0.00	0.05
	TNR	0.81	0.95	0.89	0.84	0.71	0.84
Validation	AUC	0.91	0.84	0.84	0.87	0.83	0.88
	CAR (%)	84.29	83.57	80.00	80.00	83.12	82.47
	TPR	0.86	0.77	0.76	0.76	0.99	0.94
	FPR	0.17	0.1	0.16	0.16	0.32	0.29
	FNR	0.14	0.23	0.24	0.24	0.01	0.06
	TNR	0.83	0.9	0.84	0.84	0.68	0.71

utilized to specify the numbers of membership functions of input factors (Chiu, 1994). The ANFIS model used for forest fire susceptibility modeling is trained by the hybrid method of the least-squares estimation and back-propagation described in Jang et al. (1997). Moreover, the training epoch number is set to be 100. In the case of RBFANN, it is required to fine-tune the parameters of the maximum number of neurons (N_{MAX}) and Spread of the radial basis functions (S_{rbf}). Based on experiments, N_{MAX} = 50 and S_{rbf} = 1 have been found to be appropriate for the collected data set.

In addition, the RF model has been constructed in MATLAB's programming environment (Matwork, 2017) as an ensemble of 100 decision trees. As mentioned earlier, to implement MARS, this study relies on the toolbox developed by Jekabsons (2016). As suggested by Friedman (1991) and Jekabsons (2016), the penalty coefficient (*c*) and the maximum number of basis functions (*BFs*) of MARS are determined by a grid search procedure (Hoang and Bui, 2018) within the ranges of [2, 4] for *c* and [10,50] for *BFs*.

Moreover, to confirm the statistical difference between the

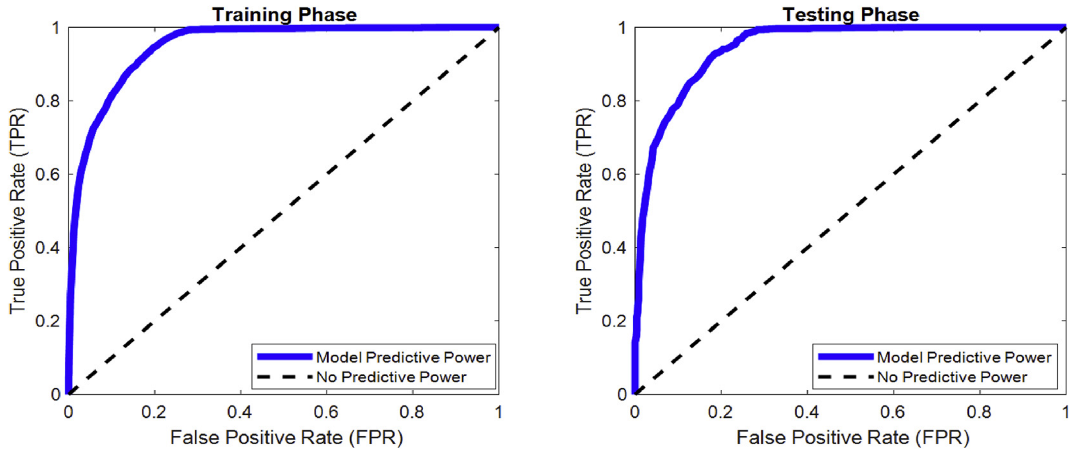

Fig. 10. Average ROCs of MARS-DFP obtained from the repeated subsampling.

Table 2
Model performance obtained from the repeated subsampling (Std: standard deviation).

Phase	Performance	Prediction Models											
		MARS-DFP		BPANN		ANFIS		RBFANN		RF		MARS	
		Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Training	AUC	0.95	0	0.91	0.1	0.93	0.01	0.94	0	0.87	0.03	0.97	0.01
	CAR (%)	87.44	0.55	83.97	7.97	87.43	0.98	87.73	0.83	84.94	1.11	90.08	1.85
	TPR	0.95	0.01	0.89	0.14	0.93	0.01	0.91	0.01	1.00	0.00	0.95	0.02
	FPR	0.20	0.01	0.21	0.07	0.18	0.01	0.16	0.01	0.30	0.02	0.15	0.03
	FNR	0.05	0.01	0.11	0.14	0.07	0.01	0.09	0.01	0.00	0.00	0.05	0.02
	TNR	0.8	0.01	0.79	0.07	0.82	0.01	0.84	0.01	0.70	0.02	0.85	0.03
Testing	AUC	0.95	0.02	0.90	0.1	0.87	0.04	0.85	0.03	0.86	0.04	0.89	0.04
	CAR (%)	86.57	2.88	83.97	8.64	80.98	3.42	79.22	3.21	84.55	2.59	82.05	3.39
	TPR	0.95	0.03	0.9	0.15	0.86	0.06	0.81	0.06	1.00	0.01	0.86	0.04
	FPR	0.21	0.05	0.22	0.08	0.24	0.06	0.22	0.06	0.31	0.05	0.22	0.06
	FNR	0.05	0.01	0.1	0.14	0.14	0.01	0.19	0.01	0.00	0.00	0.14	0.02
	TNR	0.79	0.05	0.78	0.08	0.76	0.06	0.78	0.06	0.69	0.05	0.78	0.06

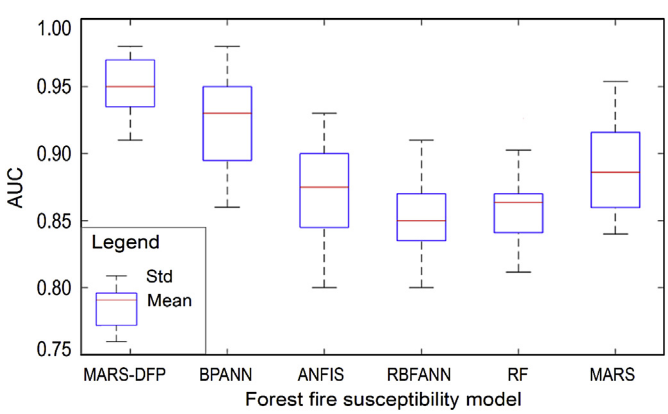

Fig. 11. Prediction performance comparison of the forest fire models (Std: Standard deviation).

Table 3
Matrix of *p*-value.

	MARS-DFP	BPANN	ANFIS	RBFANN	RF	MARS
MARS-DFP	–	0.0142	0.0001	0.0001	0.0001	0.0001
BPANN	0.0142	–	0.0146	0.0022	0.0057	0.0304
ANFIS	0.0001	0.0146	–	0.1197	0.3702	0.2471
RBFANN	0.0001	0.0022	0.1197	–	0.6273	0.0045
RF	0.0001	0.0057	0.3702	0.6273	–	0.0064
MARS	0.0001	0.0304	0.2471	0.0045	0.0064	–

prediction models, the Wilcoxon signed-rank test (WSRT) is employed in this section of the study. As a non-parametric statistical hypothesis test, WSRT is usually employed for comparing classification performance of different models (Tien Bui and Hoang, 2017). The significance level of WSRT is fixed to be 0.05. The *p*-values of the test obtained from the model prediction results are computed. If the *p*-value is less than the significance level = 0.05, it is able to confirm that the classification performances of the two models are statistically different.

5. Result and discussion

5.1. Preliminary evaluation of influencing factors

The MI-based analysis outcome is reported in Fig. 7. It can be seen observed from the figure that the influencing factor (IF_6), which is the NDVI, has received the highest MI values. This result is explainable since NDVI exhibits the condition of vegetation in a location. Hence, this factor highly influences the forest fire occurrence. This analysis outcome is in accordance with the previous findings of Biswajeet et al. (2007) and Tien Bui et al. (2017c). Moreover, since MI values of all influencing factors are not null, it can be concluded that there is no redundant factor and all the factors should be used for the model training and prediction processes.

5.2. Model performance and result comparison

The optimization process of the MARS-DFP prediction model is demonstrated in Fig. 8 which reports the best and the mean values of

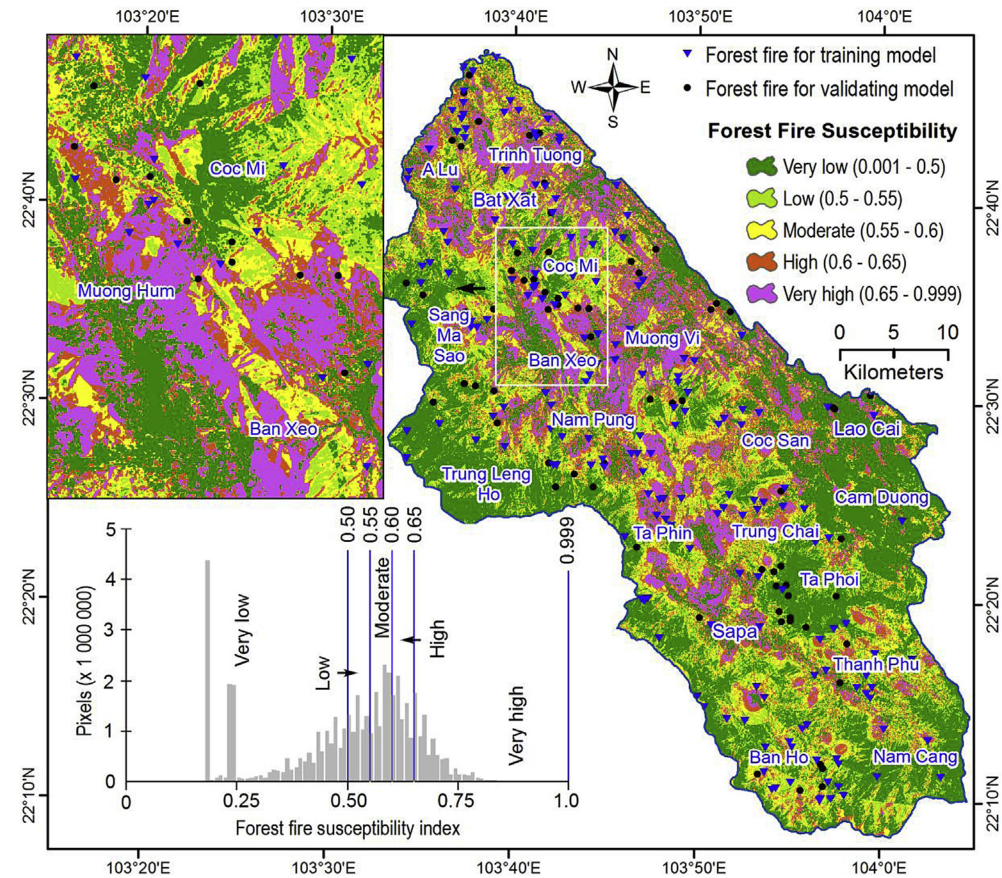

Fig. 12. Forest fire susceptibility map for the study area using the MARS-DFP model.

the cost function. The DFP metaheuristic is used to minimize the cost function.

As can be seen from the figure, after about 40 generations, the DFP has identified a good set of tuning parameters (*BFs* = 26 and *c* = 1.53) which helps to attain high values of AUC in both model training and prediction phases. Additionally, ROCs of the MARS-DFP model in both training and validation phases have been provided in Fig. 9. The AUCs of the proposed method in the training and validation phases are 0.95 and 0.91, respectively. These facts demonstrate that MARS-DFP has attained a very good forest fire susceptibility modeling outcome.

In addition, the performance of the proposed MARS-DFP prediction model is compared to those of BPANN, ANFIS, RBFANN, and RF. The experimental results of the four models in the training and validation phase are provided in Table 1. The predictive outcome of MARS-DFP is the highest (AUC = 0.91 and CAR = 84.29%), followed by MARS (AUC = 0.88 and CAR = 82.47%), RBFANN (AUC = 0.87 and CAR = 80.00%), BPANN (AUC = 0.84 and CAR = 83.57%), ANFIS (AUC = 0.84 and CAR = 80.00%), and RF (AUC = 0.83 and CAR = 83.12%).

Furthermore, to reliably evaluate the predictive capability of all the models, a random subsampling with 20 runs has been carried out. In each run, 30% of the data set is randomly extracted to create the validation data set; the rest of the data is used as the training data set. It is noted that the data in both classes of “non-fire pixels” and “non-fire pixels” have been drawn randomly to constitute the training and testing sets. The average ROCs of the proposed hybrid model collected from 20 runs is presented in Fig. 10. The outcome of the random subsampling process is reported in Table 2.

The experimental results assure that MARS-DFP (AUC = 0.95 and CAR = 86.57%) is superior to other machine learning models. The BPANN is the second-best approach with AUC = 0.90 and CAR = 83.97%, followed by MARS (AUC = 0.89 and CAR = 82.05%). ANFIS (AUC = 0.87 and CAR = 80.98%), RBFANN (AUC = 0.85 and CAR = 79.22%), and RF (AUC = 0.86 and CAR = 84.55%). It can be observed that the outcome of the random subsampling process is different from that of the training-validation process. Nevertheless, MARS-DFP has proved its superiority in both experiments.

The boxplots in Fig. 11 graphically present the predictive capability of the machine learning algorithm obtained from 20 runs.

The outcome of WSRT is reported in Table 3. Based on the computed p-value (which is less than 0.05) of MARS-DFP and BPANN (p = 0.0142), MARS-DFP and ANFIS (p = 0.0001), MARS-DFP and RBFANN (p = 0.0001), MARS-DFP and RF (p = 0.0001), and MARS-DFP and MARS (p = 0.0001). It could be confirmed that the classification performances of the MARS-DFP model and the other models are statistically different. Overall, it can be concluded that the newly constructed MARS-DFP is the most capable method compared to BPANN, ANFIS, RBFANN, RF, and MARS for forest fire spatial prediction in the study area.

5.3. Generation of the forest fire danger map for the study area

Since the proposed MARS-DFP model was successfully trained and validated, this model is used to establish a forest fire susceptibility map for the study area of Lao Cai province. MARS-DFP is employed to compute forest fire danger indices for all the pixels of the study area. The forest fire danger in Lao Cai province area was cartographically shown in terms of five categories (see Fig. 12): very high danger, high danger, moderate danger, low danger, and very low danger.

This map can be useful for the local authority to formulate effective fire prevention strategies and land-use plans. Visual interpretation of the susceptibility map shows that tropical forests of the Hoang Lien national park at Sapa, Trung Chai, Ta Phin, and Nam Phung have high susceptible to fire (Fig. 12). This is due to characteristics of the terrain which exposures to strong sun light in the dry season. In addition, slash-and-burn practices of the local people (Schliesinger, 2015) make the tropical forest fires in these areas are more flammable; therefore, these areas should be received high attentions in developing measures for fire preventions. In contrast, forests in Lao Cai and Cam Duong areas are less susceptible to forest fire. This is because in these areas, new plantation forests are dominant which have been protected and managed better.

6. Conclusion

In this research, a new hybrid computational approach based on MARS and DFP, named as MARS-DFP, has proposed for predicting tropical forest fire danger. Both MARS and DFP are state-of-the art computational method with the first one is used to generalize a classification model to predict fire susceptibility; whereas second one helps to optimize the MARS model by determining the best set of model parameters. To the best of our knowledge, they have never been explored for forest fire danger modeling; therefore, the current study is an attempt to fill this gap in the literature.

The high-performance result of the MARS-DFP model with the case study in the tropical forest at the northwest region of Vietnam indicates that the proposed model is capable to tackle the complexity of forest fire danger modeling. Compared to those deriving from benchmarks (BPANN, ANFIS, RBFANN, and RF), the proposed model has achieved better performance significantly. This fact demonstrates that the proposed MARS-DFP is a very promising alternative to help local authorities in hazard mitigation and land use planning tasks.

Future extensions of the current work include applying hybridization of MARS and DFP in spatial modeling of forest fire for other study areas and investigating the possibility of advanced feature reduction or extraction methods to improve the model predictive ability.

Acknowledgement

This research was supported by the Geographic Information Science Research group, Ton Duc Thang University, Ho Chi Minh city, Vietnam.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jenvman.2019.01.108.

References

Arpaci, A., Malowerschnig, B., Sass, O., Vacik, H., 2014. Using multi variate data mining techniques for estimating fire susceptibility of Tyrolean forests. Appl. Geogr. 53, 258–270.
Bar Massada, A., Syphard, A.D., Stewart, S.I., Radeloff, V.C., 2013. Wildfire ignition-distribution modelling: a comparative study in the Huron–Manistee national forest, Michigan, USA. Int. J. Wildland Fire 22, 174–183.
Bax, V., Francesconi, W., 2018. Environmental predictors of forest change: an analysis of natural predisposition to deforestation in the tropical Andes region, Peru. Appl. Geogr. 91, 99–110.
Bishop, C.M., 2011. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer0387310738 April 6, 2011.
Bisquert, M., Caselles, E., Sánchez-Tomás, J., Caselles, V., 2012a. Application of artificial neural networks and logistic regression to the prediction of forest fire danger in Galicia using MODIS data. Int. J. Wildland Fire 21, 1025–1029.
Bisquert, M., Caselles, E., Sánchez, J.M., Caselles, V., 2012b. Application of artificial neural networks and logistic regression to the prediction of forest fire danger in Galicia using MODIS data. Int. J. Wildland Fire 21, 1025–1029.
Biswajeet, P., Mohd, D.H.B.S., Mohamad, A.B.A., 2007. Forest fire susceptibility and risk mapping using remote sensing and geographical information systems (GIS). Disaster Prev. Manag.: Int. J. 16, 344–352.
Boulanger, Y., Gauthier, S., Burton, P., 2014. A refinement of models projecting future Canadian fire regimes using homogeneous fire regime zones. Can. J. For. Res. 44, 365–376.
Bradley, A.P., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recogn. 30, 1145–1159.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Catry, F.X., Rego, F.C., Bação, F.L., Moreira, F., 2010. Modeling and mapping wildfire ignition risk in Portugal. Int. J. Wildland Fire 18, 921–931.
Chatterjee, T., Chowdhury, R., 2017. Adaptive bilevel approximation technique for multiobjective evolutionary optimization. J. Comput. Civ. Eng. 31, 04016071.
Chen, S., Cowan, C.F.N., Grant, P.M., 1991. Orthogonal least squares learning algorithm for radial basis function networks. IEEE Trans. Neural Netw. 2, 302–309.
Cheng, M.-Y., Hoang, N.-D., 2015. Typhoon-induced slope collapse assessment using a novel bee colony optimized support vector classifier. Nat. Hazards 78, 1961–1978.
Chiu, S., 1994. Fuzzy model identification based on cluster estimation. J. Intell. Fuzzy Syst. 2.
Chuvieco, E., Aguado, I., Yebra, M., Nieto, H., Salas, J., Martín, M.P., Vilar, L., Martínez, J., Martín, S., Ibarra, P., de la Riva, J., Baeza, J., Rodríguez, F., Molina, J.R., Herrera, M.A., Zamora, R., 2010. Development of a framework for fire risk assessment using remote sensing and geographic information system technologies. Ecol. Model. 221, 46–58.
Correia, R., Teodoro, A., Duarte, L., 2017. PI2GIS: Processing Image to Geographical Information Systems, a Learning Tool for QGIS, Earth Resources and Environmental Remote Sensing/GIS Applications VIII. International Society for Optics and Photonics, pp. 104281H.
Crawford, B., Soto, R., Astorga, G., García, J., Castro, C., Paredes, F., 2017. Putting continuous metaheuristics to work in binary search spaces. Complexity 2017, 19.
Duarte, L., Teodoro, A.C., Gonçalves, J.A., Ribeiro, J., Flores, D., Lopez-Gil, A., Dominguez-Lopez, A., Angulo-Vinuesa, X., Martin-Lopez, S., Gonzalez-Herraez, M., 2017. Distributed temperature measurement in a self-burning coal waste pile through a GIS open source desktop application. ISPRS Int. J. Geo-Inf. 6, 87.
Emary, E., Zawbaa, H.M., Hassanien, A.E., 2016. Binary grey wolf optimization approaches for feature selection. Neurocomputing 172, 371–381.
Ferreira-Santiago, A., López-Martín, C., Yáñez-Márquez, C., 2016. Metaheuristic optimization of multivariate adaptive regression splines for predicting the schedule of software projects. Neural Comput. Appl. 27, 2229–2240.
Friedman, J.H., 1991. Multivariate adaptive regression splines. Ann. Stat. 19, 1–67.
García-Nieto, P.J., García-Gonzalo, E., Alonso Fernández, J.R., Díaz Muñiz, C., 2016. Using evolutionary multivariate adaptive regression splines approach to evaluate the eutrophication in the Pozón de la Dolores lake (Northern Spain). Ecol. Eng. 94, 136–151.
Goh, A.T.C., Zhang, W., Zhang, Y., Xiao, Y., Xiang, Y., 2016. Determination of Earth Pressure Balance Tunnel-Related Maximum Surface Settlement: a Multivariate Adaptive Regression Splines Approach. Bulletin of Engineering Geology and the Environment.
Hagan, M.T., Demuth, H.B., Beale, M.H., 1996. Neural Network Design. PWS Publishing Co., Boston, MA, USA.
Haghiabi, A.H., 2017. Prediction of river pipeline scour depth using multivariate adaptive regression splines. J. Pipeline Syst. Eng. Pract. 8, 04016015.
Herawati, H., Santoso, H., 2011. Tropical forest susceptibility to and risk of fire under changing climate: a review of fire nature, policy and institutions in Indonesia. For. Policy Econ. 13, 227–233.
Ho, T.K., 1995. Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition, vol. 271. pp. 278–282.
Hoang, N.-D., Bui, D.T., 2018. Predicting earthquake-induced soil liquefaction based on a hybridization of kernel Fisher discriminant analysis and a least squares support vector machine: a multi-dataset study. Bull. Eng. Geol. Environ. 77, 191–204.
Hoang, N.-D., Chen, C.-T., Liao, K.-W., 2017. Prediction of chloride diffusion in cement mortar using multi-gene genetic programming and multivariate adaptive regression splines. Measurement 112, 141–149.
Hoang, N.-D., Tien Bui, D., 2016. Predicting earthquake-induced soil liquefaction based on a hybridization of kernel Fisher discriminant analysis and a least squares support vector machine: a multi-dataset study. Bull. Eng. Geol. Environ. 1–14.
Hoang, N.-D., Tien Bui, D., 2017. Chapter 18 - slope stability evaluation using radial basis function neural network. In: Least Squares Support Vector Machines, and Extreme Learning Machine, Handbook of Neural Computation. Academic Press, pp. 333–344.
Hoang, N.-D., Tien Bui, D., Liao, K.-W., 2016. Groutability estimation of grouting processes with cement grouts using differential flower pollination optimized support vector machine. Appl. Soft Comput. 45, 173–186.
Hong, H., Naghibi, S.A., Moradi Dashtpagerdi, M., Pourghasemi, H.R., Chen, W., 2017. A comparative assessment between linear and quadratic discriminant analyses (LDA-QDA) with frequency ratio and weights-of-evidence models for forest fire susceptibility mapping in China. Arab. J. Geosci. 10, 167.
Hong, H., Tsangaratos, P., Ilia, I., Liu, J., Zhu, A.X., Xu, C., 2018. Applying genetic algorithms to set the optimal combination of forest fire related variables and model forest fire susceptibility based on data mining models. The case of Dayu County, China. Sci. Total Environ. 630, 1044–1056.
Jaafari, A., Gholami, D.M., Zenner, E.K., 2017. A Bayesian modeling of wildfire probability in the Zagros Mountains, Iran. Ecol. Inf. 39, 32–44.
Jaafari, A., Zenner, E.K., Pham, B.T., 2018. Wildfire spatial pattern analysis in the Zagros Mountains, Iran: a comparative study of decision tree based classifiers. Ecol. Inf. 43, 200–211.
Jang, J.-S.R., Sun, C.-T., Mizutani, E., 1997. Neuro-fuzzy and Soft Computing: a Computational Approach to Learning and Machine Intelligence. Prentice Hall.
Jekabsons, G., 2016. ARESLab: Adaptive Regression Splines Toolbox for Matlab/Octave. Technical Report. Riga Technical University available at: http://www.cs.rtu.lv/jekabsons/.
Lee, M.-W., Kwak, K.-C., 2016. An incremental radial basis function network based on information granules and its application. Comput. Intell. Neurosci. 2016, 6.
Leuenberger, M., Parente, J., Tonini, M., Pereira, M.G., Kanevski, M., 2018. Wildfire

susceptibility mapping: deterministic vs. stochastic approaches. Environ. Model. Softw 101, 194–203.
Li, W., Kong, D., Wu, J., 2017. A new hybrid model FPA-SVM considering cointegration for particular matter concentration forecasting: a case study of Kunming and Yuxi, China. Comput. Intell. Neurosci. 2017, 11.
Liao, K.-W., Hoang, N.-D., Gitomarsono, J., 2018. A probabilistic safety evaluation framework for multi-hazard assessment in a bridge using SO-MARS learning model. KSCE J. Civ. Eng. 22, 903–915.
Lokuge, W., Wilson, A., Gunasekara, C., Law, D.W., Setunge, S., 2018. Design of fly ash geopolymer concrete mix proportions using Multivariate Adaptive Regression Spline model. Constr. Build. Mater. 166, 472–481.
López, V., Fernández, A., García, S., Palade, V., Herrera, F., 2013. An insight into classification with imbalanced data: empirical results and current trends on using data intrinsic characteristics. Inf. Sci. 250, 113–141.
Lozano, F.J., Suárez-Seoane, S., Kelly, M., Luis, E., 2008. A multi-scale approach for modeling fire occurrence probability using satellite data and classification trees: a case study in a mountainous Mediterranean region. Rem. Sens. Environ. 112, 708–719.
Marchal, J., Cumming, S.G., McIntire, E.J.B., 2017. Land cover, more than monthly fire weather, drives fire-size distribution in Southern Québec forests: implications for fire risk management. PLoS One 12, e0179294.
MathWorks, 2012. Fuzzy Logic Toolbox. The MathWorks, Inc.
Matwork, 2017. Statistics and Machine Learning Toolbox User's Guide. Matwork Inc.. https://www.mathworks.com/help/pdf_doc/stats/stats.pdf, Accessed date: 28 April 2018.
Ministry of Agriculture and Rural Development of Vietnam, 2016. The Vietnam's Fire Watch System for Online Monitoring and Management of Forest Fires. Ministry of Agriculture and Rural Development of Vietnam Hanoi. http://www.kiemlam.org.vn/firewatchvn.
Moayed, R.Z., Kordnaeij, A., Mola-Abasi, H., 2017. Compressibility indices of saturated clays by group method of data handling and genetic algorithms. Neural Comput. Appl. 28, 551–564.
Molina, J.R., Moreno, R., Castillo, M., Rodríguez y Silva, F., 2018. Economic susceptibility of fire-prone landscapes in natural protected areas of the southern Andean range. Sci. Total Environ. 619–620, 1557–1565.
Nami, M.H., Jaafari, A., Fallah, M., Nabiuni, S., 2018. Spatial prediction of wildfire probability in the Hyrcanian ecoregion using evidential belief function model and GIS. Int. J. Environ. Sci. Technol. 15, 373–384.
Oliveira, S., Zêzere, J.L., Queirós, M., Pereira, J.M., 2017. Assessing the social context of wildfire-affected areas. The case of mainland Portugal. Appl. Geogr. 88, 104–117.
Parisien, M.-A., Carol, M., Sean, A.P., Evan, R.D., François-Nicolas, R., Mike, D.F., 2016. The spatially varying influence of humans on fire probability in North America. Environ. Res. Lett. 11, 075005.
Peng, H., Fulmi, L., Ding, C., 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. Pattern Anal. Mach. Intell. 27, 1226–1238.
Peterson, A.T., Papeş, M., Soberón, J., 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. Ecol. Model. 213, 63–72.
Pham, B.T., Jaafari, A., Prakash, I., Bui, D.T., 2018. A novel hybrid intelligent model of support vector machines and the multi boost ensemble for landslide susceptibility modeling. Bull. Eng. Geol. Environ. https://doi.org/10.1007/s10064-018-1281-y.
Pourghasemi, H.r., Beheshtirad, M., Pradhan, B., 2016. A comparative assessment of prediction capabilities of modified analytical hierarchy process (M-AHP) and Mamdani fuzzy logic models using Netcad-GIS for forest fire susceptibility mapping. Geomatics, Nat. Hazards. Risk 7, 861–885.
Pourtaghi, Z.S., Pourghasemi, H.R., Aretano, R., Semeraro, T., 2016. Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques. Ecol. Indicat. 64, 72–84.
Prayogo, D., Susanto, Y.T.T., 2018. Optimizing the prediction accuracy of friction capacity of driven piles in cohesive soil using a novel self-tuning least squares support vector machine. Adv. Civ. Eng. 2018, 9.
Price, K.V., Storn, R.M., Lampinen, J.A., 2005. Differential Evolution A Practical Approach to Global Optimization. Springer-Verlag, Berlin, Germany.
Rajan, N.M., Shanmugam, J., 2018. Estimation of Forest Fire Spread by Modeling in GIS Platform.
Razavi Termeh, S.V., Kornejady, A., Pourghasemi, H.R., Keesstra, S., 2018. Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms. Sci. Total Environ. 615, 438–451.
Robinne, F.-N., Parisien, M.-A., Flannigan, M., 2016. Anthropogenic influence on wildfire activity in Alberta, Canada. Int. J. Wildland Fire 25, 1131–1143.
Rodrigues, M., de la Riva, J., 2014. An insight into machine-learning algorithms to model human-caused wildfire occurrence. Environ. Model. Softw 57, 192–201.
Sahana, M., Ganaie, T.A., 2017. GIS-based landscape vulnerability assessment to forest fire susceptibility of Rudraprayag district, Uttarakhand, India. Environ. Earth Sci. 76, 676.
Satir, O., Berberoglu, S., Donmez, C., 2016. Mapping regional forest fire probability using artificial neural network model in a Mediterranean forest ecosystem. Geomatics, Nat. Hazards. Risk 7, 1645–1658.
Schliesinger, J., 2015. Hill Tribes of Vietnam Vol 2: Profile of the Existing Hill Tribe Groups. Booksmango.
Schweizer, D.W., Cisneros, R., 2017. Forest fire policy: change conventional thinking of smoke management to prioritize long-term air quality and public health. Air Quality, Atmos. Health 10, 33–36.
Silva, G.L., Soares, P., Marques, S., Dias, M.I., Oliveira, M.M., Borges, J.G., 2015. A Bayesian Modelling of Wildfires in Portugal. Springer International Publishing, Cham, pp. 723–733.
Storn, R., Price, K., 1997. Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. J. Global Optim. 11, 341–359.
Suman, S., Das, S.K., Mohanty, R., 2016. Prediction of friction capacity of driven piles in clay using artificial intelligence techniques. Int. J. Geotech. Eng. 10, 469–475.
Tehrany, M.S., Jones, S., Shabani, F., Martínez-Álvarez, F., Bui, D.T., 2018. A novel ensemble modeling approach for the spatial prediction of tropical forest fire susceptibility using logitboost machine learning classifier and multi-source geospatial data. Theor. Appl. Climatol. 1–17.
Teodoro, A., Duarte, L., Sillero, N., Gonçalves, J., Fonte, J., Gonçalves-Seco, L., da Luz, L.P., dos Santos Beja, N., 2015. An Integrated and Open source GIS Environmental Management System for a Protected Area in the South of Portugal, Earth Resources and Environmental Remote Sensing/GIS Applications VI. International Society for Optics and Photonics, pp. 96440U.
Tien Bui, D., Anh Tuan, T., Hoang, N.-D., Quoc Thanh, N., Nguyen, B.D., Van Liem, N., Pradhan, B., 2017a. Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a novel hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. Landslides 14, 447–458.
Tien Bui, D., Bui, Q.-T., Nguyen, Q.-P., Pradhan, B., Nampak, H., Trinh, P.T., 2017b. A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. Agric. For. Meteorol. 233, 32–44.
Tien Bui, D., Bui, Q.-T., Nguyen, Q.-P., Pradhan, B., Nampak, H., Trinh, P.T., 2017c. A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. Agric. For. Meteorol. 233, 32–44.
Tien Bui, D., Hoang, N.D., 2017. A Bayesian framework based on a Gaussian mixture model and radial-basis-function Fisher discriminant analysis (BayGmmKda V1.1) for spatial prediction of floods. Geosci. Model Dev. 10, 3391–3409.
Tien Bui, D., Le, K.-T., Nguyen, V., Le, H., Revhaug, I., 2016a. Tropical forest fire susceptibility mapping at the cat Ba national park area, Hai phong city, Vietnam, using GIS-based kernel logistic regression. Rem. Sens. 8, 347.
Tien Bui, D., Nguyen, Q.P., Hoang, N.-D., Klempe, H., 2016b. A novel fuzzy K-nearest neighbor inference model with differential evolution for spatial prediction of rainfall-induced shallow landslides in a tropical hilly area using GIS. Landslides 1–17.
Tien Bui, D., Tran, A.T., Klempe, H., Pradhan, B., Revhaug, I., 2015. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides 1–18.
Tien Bui, D., Tuan, T.A., Hoang, N.-D., Thanh, N.Q., Nguyen, D.B., Van Liem, N., Pradhan, B., 2016c. Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. Landslides 1–12.
Valdez, M.C., Chang, K.-T., Chen, C.-F., Chiang, S.-H., Santos, J.L., 2017. Modelling the spatial variability of wildfire susceptibility in Honduras using remote sensing and geographical information systems. Geomatics, Nat. Hazards. Risk 8, 876–892.
Verde, J., Zêzere, J., 2010. Assessment and validation of wildfire susceptibility and hazard in Portugal. Nat. Hazards Earth Syst. Sci. 10, 485–497.
Vietnamnews, 2016. Fire Destroys 8ha of Forest in Lào Cai. http://vietnamnews.vn/society/293417/fire-destroys-8ha-of-forest-in-lao-cai.html#bs5rVArZRMlof5vK.99, Accessed date: 3 January 2018.
Vietnamplus, 2015. Lao Cai Alerted to Forest Fire Risk. https://en.vietnamplus.vn/lao-cai-alerted-to-forest-fire-risk/73036.vnp, Accessed date: 3 January 2018.
Wang, X., Wotton, B.M., Cantin, A.S., Parisien, M.-A., Anderson, K., Moore, B., Flannigan, M.D., 2017. cffdrs: an R package for the Canadian forest fire danger rating system. Ecol. Process. 6, 5.
West, A.M., Kumar, S., Jarnevich, C.S., 2016. Regional modeling of large wildfires under current and potential future climates in Colorado and Wyoming, USA. Climatic Change 134, 565–577.
Wu, Z., He, H.S., Yang, J., Liang, Y., 2015. Defining fire environment zones in the boreal forests of northeastern China. Sci. Total Environ. 518–519, 106–116.
Xue, X., 2017. Prediction of slope stability based on hybrid PSO and LSSVM. J. Comput. Civ. Eng. 31, 04016041.
Yang, X.-S., 2012. Flower pollination algorithm for global optimization. In: Durand-Lose, J., Jonoska, N. (Eds.), Unconventional Computation and Natural Computation. Springer Berlin Heidelberg, pp. 240–249.
Yang, X.-S., 2014. Nature-Inspired Optimization Algorithms.
Yao, J., Raffuse, S.M., Brauer, M., Williamson, G.J., Bowman, D.M.J.S., Johnston, F.H., Henderson, S.B., 2018. Predicting the minimum height of forest fire smoke within the atmosphere using machine learning and data from the CALIPSO satellite. Rem. Sens. Environ. 206, 98–106.
Yilmaz, B., Aras, E., Nacar, S., Kankal, M., 2018. Estimating suspended sediment load with multivariate adaptive regression spline, teaching-learning based optimization, and artificial bee colony models. Sci. Total Environ. 639, 826–840.
York, T.P., Eaves, L.J., Oord, E.J.C.G.v.d., 2006. Multivariate adaptive regression splines: a powerful method for detecting disease–risk relationship differences among subgroups. Stat. Med. 25, 1355–1367.
Zhang, J., Sanderson, A.C., 2009. JADE: adaptive differential evolution with optional external archive. IEEE Trans. Evol. Comput. 13, 945–958.