Journal of Environmental Informatics 36(1) 43-57 (2020)

Journal of Environmental Informatics

www.iseis.org/jei

# A Hybrid Intelligence System Based on Relevance Vector Machines and Imperialist Competitive Optimization for Modelling Forest Fire Danger Using GIS

H. V. Le[1], Q. T. Bui[2], D. Tien Bui[3, 4 *], H. H. Tran[5], and N. D. Hoang[6]

[1] *Faculty of Information Technology, Hanoi University of Mining and Geology, Duc Thang, Bac Tu Liem, Hanoi, Vietnam*
[2] *Faculty of Geography, VNU University of Science, 334 Nguyen Trai, Thanh Xuan, Hanoi, Vietnam*
[3] *Geographic Information Science Research Group, Ton Duc Thang University, Ho Chi Minh City, Vietnam*
[4] *Faculty of Environment and Labour Safety, Ton Duc Thang University, Ho Chi Minh City, Vietnam*
[5] *Faculty of Geomatics and Land Administration, Hanoi University of Mining and Geology, Duc Thang, Bac Tu Liem, Hanoi, Vietnam*
[6] *Faculty of Civil Engineering, Institute of Research and Development, Duy Tan University, P809 - K7/25 Quang Trung, Danang, Vietnam*

Received 18 July 2017; revised 6 May 2018; accepted 23 July 2018; published online 22 December 2018

**ABSTRACT.** This article proposes and verifies a novel intelligence approach for modelling forest fire danger developed based on a hybrid model of Imperialist Competitive Algorithm (ICA) and Relevance Vector Machine (RVM). The hybrid model is named as ICA-RVM. They are state-of-the art machine learning techniques that have not been investigated for forest fire danger modeling. RVM is used to establish a prediction model that computes probability of fire danger, whereas ICA is adopted to optimize the prediction model. The tropical forest at Gia Lai province, Central Highland (Vietnam), was used as a case study. A geographic information system (GIS) database featuring 12 fire ignition factors has been established to train and verify the hybrid intelligence model. Area under the curve (AUC) and statistical measures were used to assess the model performance. The result showed that the proposed model achieves high performances; AUC is 0.842 and 0.793 on the training dataset and the validation dataset, respectively. Compared to two benchmarks, Random Forests and Support Vector Machine, the proposed model performs better. Therefore, the propose ICA-RVM is a valid alternative system for forest fire danger modeling.

*Keywords:* forest fire; Imperialist Competitive Algorithm; Relevance Vector Machines; GIS; Gia Lai; Vietnam

## 1. Introduction

Forest fire, which is currently a global problem, is a result from complex interactions among vegetation fuels, weather and climate and human landuse activities (Dupire et al., 2017). Over the past decades, due to climate change, forest fires have increaseingly become a serious natural hazard that threaten local communities, destroy vast amounts of natural resources, causing soil degradation and air pollution (Conard et al., 2017; You et al., 2017). In addition, change of climate i.e., prolonged dry weather with high temperature is expected to increase in both number of fires and areas burned in many regions in the world (Doerr and Santín, 2016). Therefore, a better knowledge of fire risk valuably helps to minimize losses to residents, economic activeities, and buildings within territories vulnerable to forest fires.

A forest fire danger map is able to reveal areas highly vulnerable and affected by the hazard and therefore assists the development of land use planning. In addition, information on the spatial distribution of fires is a requisite to meliorate fire prevention strategies and tactics (Pourghasemi et al., 2016), quantify economic losses from wildfires (Alcasena et al., 2016), study the human influence on fire ignition (Fusco et al., 2016), or investigate the effects of climatic and local factors on fire occurrences (Wu et al., 2014). Furthermore, the spread of fires in forests is a complicated phenomenon conditioned by the fuel area, wind speed, wind direction, slope, and other factors (You et al., 2017). These facts make spatial modeling forest fire is indeed still a challenging task.

Recent advancements in geographic information system (GIS) and remote sensing (RS) technologies have facilitated many research works on spatial modeling of forest fires (Chuvieco et al., 2010; Duarte, 2013; Teodoro et al., 2015; Teodoro and You et al., 2017) due to the ability to handle large-scale databases with multi-layered information of spatial characteristics (Tien Bui et al., 2017a). Various studies have employed probabilistic and physical models including Fuel Moisture Content (FMC) (García et al., 2008), Fire Area Simulator (FARSITE) (Krasnow et al., 2009), Maximum Entropy (Renard et al., 2012), mathematical models (Grishin and Filkov, 2011), analytic hierarchy process (Güngöroğlu, 2017), and numerical simulation (Sanjuan et al., 2016). However, due to the multivariate nature of the problem at hand, it is still difficult to predict future forest fires.

* Corresponding author. Tel.: +47 35952724; fax: +47 35952601.
*E-mail address:* buitiendieu@tdtu.edu.vn (D. Tien Bui).

ISSN: 1726-2135 print/1684-8799 online
 doi:10.3808/jei.201800404

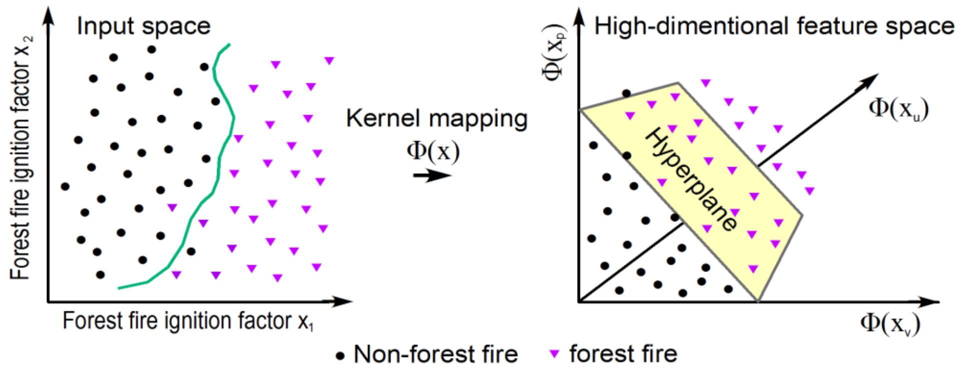


**Figure 1.** An illustration of RVM classification concept.

Therefore, advanced statistical and machine learning methods have drawn the attention of researchers in spatial modeling of various natural hazards (Bui et al., 2016; Hoang and Tien Bui, 2016a; Pham et al., 2016; Tien Bui et al., 2016c; Shirzadi et al., 2017). Since the problem of fire danger mapping can be formulated as a pattern classification task, machine learning algorithms can effectively help to establish accurate prediction models based on GIS databases. Carvalho et al. (2006) attempted to construct a forest fire modelling method based on fuzzy logic and cellular automata. Koetz et al. (2008) relied on Support vector machine (SVM) for performing multi-source land cover classification for forest fire management. Bisquert et al. (2012) applied artificial neural networks (ANN) and logistic regression (LR) to predict forest fire danger. Pourtaghi et al. (2015) employed Shannon's entropy to construct forest fire susceptibility maps in Iran. Pierce et al. (2012) and Pourtaghi et al. (2016) used random forests (RF) to construct forest fire susceptibility maps. Adaptive neuro-fuzzy inference systems have been employed Tien Bui et al. (2017b). A common finding of these studies is that machine learning provides an effective solution for analyzing large-scale data sets and deriving highly accurate prediction models for forest fire mapping. Therefore, exploration of new machine learning algorithms for forest fire modeling is highly necessary.

In this study, we aim at extending the body of knowledge by proposing a new alternative for the problem of forest fire danger modeling. The proposed approach is a novel integration of Relevance Vector Machine (RVM), the Imperialist Competitive Algorithm (ICA), and a GIS database. RVM, which was proposed by Tipping (2000), is a Bayesian inference approach for probabilistic classification. Since forest fire evaluation is undoubtedly a complicated and uncertain problem (Brunette et al., 2017; Tien Bui et al., 2017b) and it is very desirable for the decision-makers to obtain a prediction model that can exhibit the uncertainties associated with the estimations, probabilistic models for forest fire susceptibility mapping is a practical need. Furthermore, the model construction of RVM requires an appropriate setting of its tuning parameter, so ICA is integrated with RVM to assist the model establishment. ICA has been demonstrated in the previous studies to be an effective metaheuristic approach for solving optimization method in continuous space (Hosseini and Al Khaled, 2014). The tropical forest of the Gia Lai province in the Central Highland of Vietnam was selected as a case study to train and verify the proposed integrated model. This province has particularly faced with the forest fire problem during the last ten years. Especially, this is one of the most sensitive provinces to El Niño Southern Oscillation that caused serious droughts and forest fire in 2016 (CGIAR, 2016).

## 2. Theoretical Background of the Methods Used

### 2.1. Relevance Vector Machine

Developed by Tipping (2000), Relevance vector machine (RVM) is a Bayesian inference method which can be applied for probabilistic pattern recognition. The functional form of RVM is similar to that of the SVM proposed by Vapnik (1998). Moreover, RVM is capable of yielding probabilistic classification. Different from SVM, expectation maximization-based method is employed to establish the RMV prediction model. This algorithm also relies on a kernel function to deal with data nonlinearity. RVM first maps the input data from the original space to a high dimensional feature space where the data can be linearly separated (see Figure 1).

The following section briefly describes the RVM-based classification model. For more detail of this algorithm as well as its implementation, readers are guided to previous works of Tipping (2000) and Tipping (2009). Given a training data consisting of a set of fire ignition factors $\mathbf{X}=\{\mathbf{x}_i\}_{i=1}^{N}$, with corresponding target classes $C=\{C_i\}_{i=1}^{2}$, the task at hand is to construct a classification boundary that separates the input vector $\mathbf{X}$ into the two classification regions. In the current research context, forest modeling is considered to be a binary classification problem, the target outputs $c_i$ has two possible class labels, 0 for the non-forest fire class and 1 for the forest fire class. Given the input vector, the conditional distribution of the class label is provided as follows (Bishop and Tipping, 2000):

$$P(c_i|\mathbf{x},w)\sigma(y) \tag{1}$$

where $\sigma(y)=1/(1+e^{-y})$ denotes a logistic sigmoid function;

$y$ is a linearly-weighted sum of $M$ fixed basis function $\varphi_m(x)$:

$$y(\mathbf{x},w)=\sum_{m=1}^{M} w_m\varphi_m(\mathbf{x})+\mathbf{w}_o=\mathbf{w}\boldsymbol{\varphi};$$
$$\boldsymbol{\varphi}=(1,\varphi_1(\mathbf{x}_t),\varphi_2(\mathbf{x}_t),\ldots,\varphi_m(\mathbf{x}_t)) \tag{2}$$

where $w$ denotes a vector of the model weights; $\varphi$ represents a vector of basis function. The basis function that is often employed is the Gaussian radial basis function (Samui, 2012):

$$\varphi_m(\mathbf{x})=\exp(\frac{-\|\mathbf{x}-\mathbf{x}_m\|^2}{2r^2}) \tag{3}$$

where $r$ represents the width of the radial basis function. It is proper to note that the parameter $r$ affects the smoothness of the classification boundary, and therefore, this parameter should be appropriately selected for the training process of the RVM model.

The likelihood function for all data points is given in the form of a Bernoulli distribution (Tipping, 2000) as follows:

$$P(\mathrm{C}\mid \mathrm{w})=\prod_{i=1}^{N}\sigma(y_i)^{c_i}.(1-\sigma(y_i))^{1-c_i} \tag{4}$$

In addition, a sparse weight prior distribution can be obtained by assigning a different variance parameter for each weight (Tipping, 2001). Hence, the prior distribution over $w$ is given in the following manner (Tipping and Faul, 2003):

$$\begin{aligned} p(\mathrm{w}\mid\alpha)&=\prod_{m=1}^{M}N(\mathrm{w}_m\mid 0,\alpha_m^{-1})\\ &=(2\pi)^{-M/2}\prod_{m=1}^{M}\alpha_m^{1/2}\exp(-\frac{\alpha_m \mathrm{w}_m^2}{2}) \end{aligned} \tag{5}$$

where $N(z|\mu, S)$ is a multivariate Gaussian distribution over $z$ with mean $\mu$ and covariance $S$; $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_M)$ is a vector of independent hyper-parameters; each element $\alpha_j$ of the vector controls the strength of the prior over its associated weight $w_j$.

Given initial values of the hyper-parameter $\alpha$, it is noted that $p(\mathrm{w}\mid \mathrm{C},\alpha)\propto p(\mathrm{C}\mid \mathrm{w})\,p(\mathrm{w}\mid\alpha)$ and the most probable weight $\mu$ can be found by maximizing the penalized negative log-likelihood function over $w$ (Tipping and Faul, 2003):

$$\begin{aligned} &\log(P(\mathrm{C}\mid \mathrm{w})\,p(\mathrm{w}\mid\alpha))\\ &=\sum_{i=1}^{N}(c_i\log y_i+(1-c_i)\log(1-y_i))-0.5\mathbf{w}^T\mathbf{A}\mathbf{w} \end{aligned} \tag{6}$$

where $\mathbf{A} = diag\ (\alpha_1, \alpha_2, \ldots, \alpha_M)$.

The iteratively reweighted least squares algorithm (Nabney, 1999) and the Laplace approximation procedure are then applied to solve the above optimization problem; the most probable weight $\mu$ and its covariance $\Sigma$ are found as follows (Tipping, 2000; Tipping and Faul, 2003):

$$\mu=\Sigma\boldsymbol{\varphi}^{\mathrm{T}}\mathbf{BC} \quad \text{and} \quad \Sigma=(\boldsymbol{\varphi}^{\mathrm{T}}\mathbf{B}\boldsymbol{\varphi}+\mathbf{A})^{-1} \tag{7}$$

where $B = diag\ (\beta_1, \beta_2, \ldots, \beta_N)$ with $\beta_i = \sigma\{y(x_i)\}.\ [1 - \sigma\{y(x_i)\}]$.

Notably, when the optimization process ends, an interesting outcome is that many elements of the hyper-parameter vector $\alpha$ approach infinity; thus, the weight vector w only has a few non-zero elements which are considered as relevant vectors (Tipping, 2000). After the training process, the vector of model weights $w$ is then used to predict the posterior of the class label $c_i$ given an input vector $x$ using Equation1.

### 2.2. Imperialist Competitive Algorithm

Imperialist Competitive Algorithm (ICA), proposed by Atashpaz-Gargari and Lucas (2007), is inspired from the field of human social evolution. This algorithm belongs to the group of swarm intelligence which can effectively deal with continuous optimization problems (Hosseini and Al Khaled, 2014). However, exploration of ICA for forest fire modeling has not been carried out.

Basically, as other metaheuristics, ICA is specifically designed for solving optimization problem in which no exact algorithm can be applied in polynomial time, also called Non-deterministic Polynomial time (NP)-hard problem (Marandi et al., 2014). The task of finding an optimal value for a machine learning model can be classified as a NP-hard problem since the interaction between the model structure and the training data is highly complex and there is an infinite number of possible values of the tuning parameter. Since successful applications of ICA have been widely observed (Hosseini and Al Khaled, 2014; Sadowski and Nikoo, 2014), this algorithm can be helpful to assist the training phase of RVM by selecting an optimal width of the radial basis function.

In essence, ICA is a population-based stochastic search inspired by imperialistic competition. This algorithm attempts to mimic the social policy of imperialism in the real world. When an empire raises, it dominates more colonies and take advantage of their sources; if one empire falls, other empires will compete to take its possession. In ICA, individuals within the population represent countries and they interact with each other to form empires that possess colonies.

ICA begins with an initial population and a pre-specified objective function (see Section 4.2). Based on the objective function value, the most powerful countries are chosen as imperialists and the others are colonies of them. The algorithm then simulates the competition among imperialists in order to acquire more colonies. The best imperialist typically has more chance to occupy more colonies. A population of ICA is illustrated in Figure 2.

After colonies have been assigned to each imperialist, these colonies move towards their corresponding imperialists. These movements of colonies are demonstrated in Figure 3. In this figure, it is noted that $\alpha$ represents a uniformly distributed random number, generated as follows:

$$a\sim U(0,\theta\times S) \tag{8}$$

where $\theta$ denotes a constant variable; typically, $\theta$ is greater than 1 (e.g., 1.5) (Atashpaz-Gargari and Lucas, 2007). $S$ is the distance between the colony and the imperialist.

During the competition process, the weakest empire gradually loses their colonies and other powerful empires attempt to obtain them. Moreover, if the colonies demonstrate more power than its relevant imperialist, they will exchange their positions (Kaveh and Talatahari, 2010). The empire that has no colonies will collapse and eventually the most powerful empire will dominate all other empires and represent the final optimal solution for the optimization problem at hand. The overall algorithm of ICA is demonstrated in **Algorithm 1**.

**Begin algorithm**

*Select population size, maximum iteration, and define objecttive function*

*Initialization of the algorithm*

// Generate some random solution in the search space and create initial empires.

**For** each iteration **do**

*1-Move the colonies toward their possessed imperialist.*

*2-Exchange the position of a colony and the imperialist based on the objective function value.*

*3-Calculate the objective function of empires.*

*4-Assign the weakest colony to one of the most powerful empires.*

*Delete the powerless empires.*

*If there is just one empire, stop; otherwise, go to 1*

*If the current iteration exceeds its maximum value, stop, if not go to 1.*

**End for**

**End algorithm**

**Algorithm 1**. The ICA optimization process

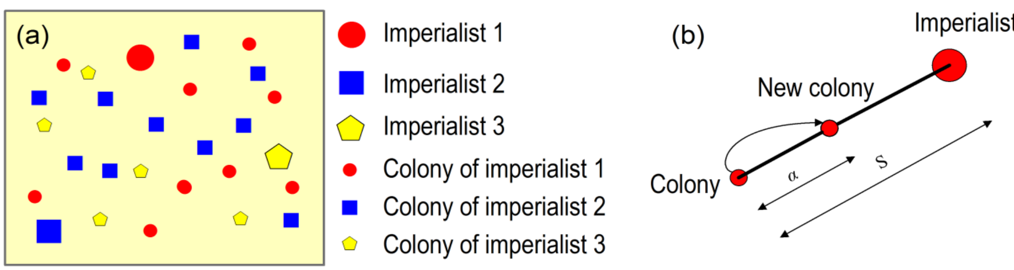


**Figure 2.** (a) An initial population of the Imperialist Competitive Algorithm and (b) Movements of colonies.

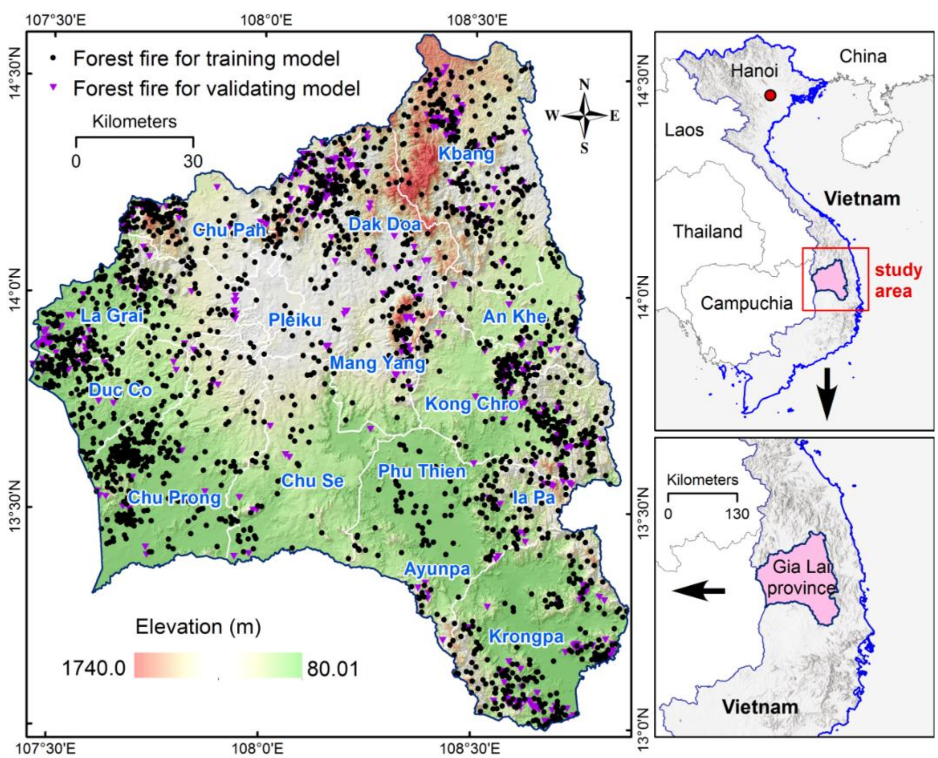


**Figure 3.** The study area and forest fire locations.

### 2.3. Performance Evaluation Metrics

To evaluate the goodness-of-fit and prediction power of the forest fire susceptibility models are evaluated based on statistical measures (Tien Bui et al., 2016a) such as classification rate (CR), specificity, sensitivity were computed as follows:

$$CR = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{10}$$

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

*CR* is the number of forest fires and non-fire points that are correctly classified divided to the total points. Sensitivity is the percentage of correct forest fire points whereas specificity is the percentage of correct non-fire points in the training or validation datasets. In addition, specificity provides a measure of the proportion of non-fire samples that are accurately categorized as such.

Moreover, the global performance of the forest fire models is evaluated using the receiver operating characteristic (ROC) curve that is constructed based on sensitivity (true positivity rate) and specificity (false negative rate) (Tien Bui et al., 2017a). To quantify the global performance, area under the curve (AUC) that varies between 0.5 and 1 is also used. AUC values of 0.5 ~ 0.6 indicate insufficient whereas values of 0.6 ~ 0.7 indicate poor performance. Models with AUC values of 0.7 ~ 0.8 denote moderate performance while models with AUC values 0.7 ~ 0.8 denote good performance. Models with AUC values of 0.8 ~ 0.9 will indicate very good performance (Peterson et al., 2008).

## 3. The Study Area and GIS Database

### 3.1. Description of the Study Area

The study area is the Gia Lai province that is located at the Central Highlands region in Vietnam (Figure 3.), between longitudes 107º26’00”E and 108º51’00”E, and between latitudes 12º58’00”N and 14º35’00”N. It covers an area of 15512.8 km$^2$. The elevation is from 80 m to 1740 m above the sea level, with the mean is 522.3 m and the standard deviation is 278.2 m.

The province is inhabited by nearly 1.4 million people with the average population density of 90 persons/km$^2$ in 2015 and around 70% of the settlements are non-urban areas (GSO, 2015). Due to its special location, Gia Lai is crucial for socio-economic connection with the Southern Central coast of Vietnam and is the center of the Vietnam-Laos-Cambodia triangle development zone. The overall poverty rate in the province is 11.36%, and the annual GDP per capita is about 1100 USD (GSO, 2015). The economy of the province is mainly based on agriculture with 500 thousand hectares of annual crops and perennial trees). The industry is agro-forestry processing and hydroelectric power with 82 hydroelectric projects in the whole province.

Climate of the study area is characterized by a tropical monsoon and plateau climate (Van et al., 2014) with two seasons: a rainy season from June to October and a dry season from November to May. The average annual temperature is 18 ~ 20 ºC for areas with elevation is above 600 m, whereas the average temperature is around 25 ºC for the other areas. The average annual rainfall is 1634 mm, the average annual humidity is 80%, and the number of hours of sunshine is about 2757 hours (GSO, 2015).

According to the land use statistics (CGIAR, 2016), forest and perennial crop lands cover approximately 64.5% of the total study area. The production forest land and the protection forest land account for 33.4% and 9.7% of the total study area, respectively. Due to climate changes such as prolonged dry weather in recent years (Van et al., 2014), the forest fire problem at the province seems to be severe i.e. a huge forest fire occurred in December 2013 at the protected pine forest at Bac Bien Ho areas destroyed around 250 ha; a huge forest fire also occurred on March 2015 at the tourist spot of Ham Rong mountain destroyed more than 3 ha pine forests; therefore, study of forest fire for this province is an urgent task.

### 3.2. Forest Fire

Historical fires, their locations, and their influencing factors are a main key for modelling forest fires (Massada et al., 2011). Therefore, an inventory map with a total of 2530 historical fire location for the study area was prepared first. These fires that occurred during the last 10 years (2007 – 2016) were provided by the Department of Forest Protection of Vietnam (Ministry of Agriculture and Rural Development of Vietnam, 2016). This is the official forest fire database in Vietnam, and is now available at http://www.kiemlam.org.vn/firewatchvn/.

Temporal analysis of the forest fires (Figure 4) shows that 88.8% of the total fires occurred in the dry season, from January to May, in which the forest fires are at the peak on March (39.4%). In contrast, few forest fires occurred in the rainy season. Many forest fires occurred in years 2015 (33.9%), 2016 (16.3%), 2010 (19%), and 2013 (13.2%), whereas, few forest fires occurred for the other years. Surprisingly, no fire occurred in 2011. It is emphasis that a series of floods occurred in 2011 (Tien Bui et al., 2016a) whereas prolonged dry weather with worst droughts occurred years of 2010, 2013, 2015, and 2016 due to El Niño Southern Oscillation (CGIAR, 2016).

### 3.3. Ignition Factor

Occurrence of forest fires is influenced by interactions of various factors such as topography, fuels, and climate patterns (Cary et al., 2009); therefore, determination of fire ignition factors is a crucial task in forest fire modeling. In this study, a total of 12 ignition factors were considered (Tien Bui et al., 2017b), slope (º), aspect, elevation (m), curvature, NDVI (Normalized Difference Vegetation Index), NDWI (Normalized Difference Water Index), NDMI (Normalized Difference

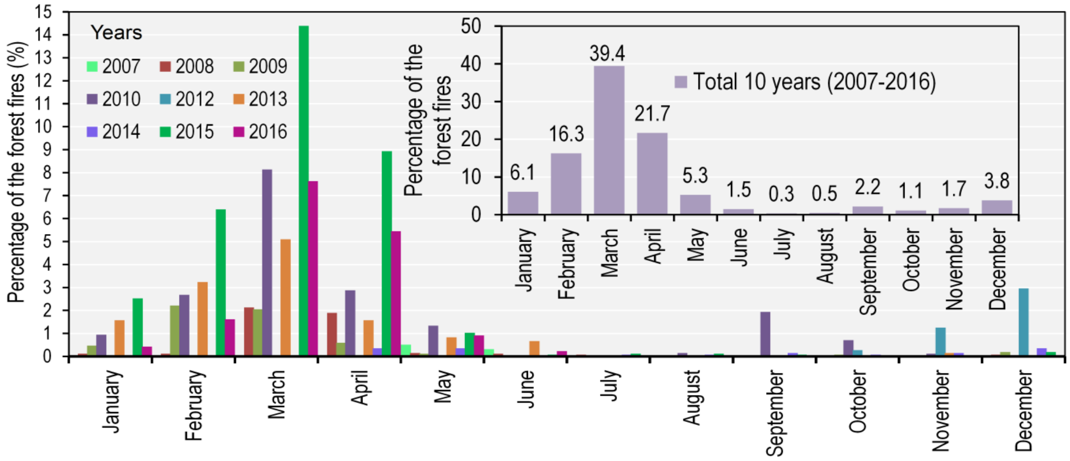


**Figure 4.** Temporal analysis of the forest fires occurrence for the study area.

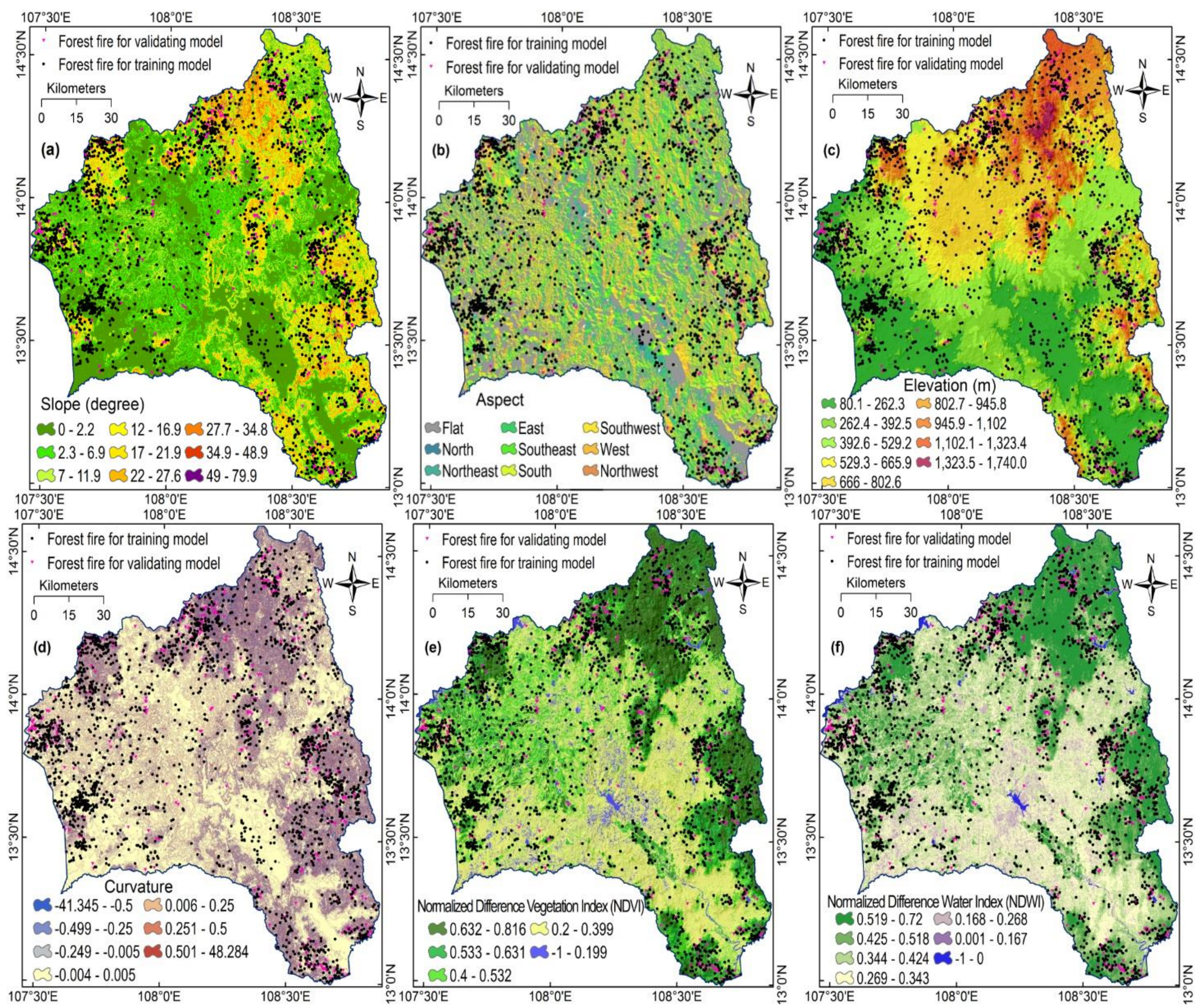


**Figure 5.** Forest fire ignition factors used in this study: (a) Slope map; (b) Aspect map; (c) Elevation map; (d) Curvature map; (e) Landuse map; (f) NDVI map.

Moisture Index), land use, temperature (º), wind speed (m/s), relative humidity (%), and rainfall (mm).

Slope and aspect should be selected because slope influences fire spread rate, whereas aspect is related to wind speeds, a major factor affecting fires propagation (Pimont et al., 2012). Elevation influences vegetation distribution, fuel moisture, and air humidity that indirectly relate to forest fires (Verde and Zêzere, 2010). Curvature is used for the forest fire modeling because lower local curvature has proven to spread fires faster than those with higher local curvature (Hilton et al., 2017), therefore it may influence behaviors of forest fires. In this study, the slope map (Figure 5a), the aspect map (Figure 5b), and the elevation map (Figure 5c) were generated with 9 classes, whereas the curvature map was compiled with 7 categories (Figure 5d). These classes and categories were determined based on histogram distribution analysis of the data values of these maps using the natural breaks algorithm with the Jenks optimization algorithm (Jenks, 1977) in ArcGIS 10.4. Accordingly, the data in the four maps were analyzed to exploit natural breakpoints in which the inter-variance of the homogenous classes is maximized and the intra-variance of the homogenous classes is minimized.

Topography influences fire behavior because it affects the local climate, wind direction, and vegetation cover (Oliveira et al., 2012); therefore, it was considered for analyzing forest fire occurrence in this study. For this purpose, a Digital Elevation Model (DEM) of the study area was generated using national digital topographic maps at a scale of 1:50,000 using ArcGIS 10.4 software. Based on the DEM, slope, aspect, elevation, and curvature were generated.

NDVI is considered as a proxy for vegetation health status that related to the flammability and quantity of surface fuel (Zylstra et al., 2016), whereas, behaviors of forest fires are also influenced by vegetation water content (Maki et al., 2004) and live fuel moisture (Dennison et al., 2005). Therefore, these factors should be considered in forest fire modeling. For this research, reflectance (Vlassova and Pérez-Cabello, 2016) in Landsat-8 Operational Land Imagery spectral bands (30 m resolution) and available at http://earthexplorer.usgs.gov) was used to compute the three indices, i.e., NDVI using Equation 11 (DeFries and Townshend, 1994). The vegetation water content was derived using the Normalized Difference Water Index (NDWI) in Equation 12 (Gao, 1996), whereas the live fuel moisture was estimated through Normalized Difference Moisture Index (NDMI) in Equation 13 (Xu, 2006) as follows:

$$NDVI = (\text{Band } 5 - \text{Band } 4)/(\text{Band } 5 + \text{Band } 4) \tag{11}$$

$$NDVI = (\text{Band } 3 - \text{Band } 5)/(\text{Band } 3 + \text{Band } 5) \tag{12}$$

$$NDVI = (\text{Band } 5 - \text{Band } 6)/(\text{Band } 5 + \text{Band } 6) \tag{13}$$

where Band 3 is the Green band (0.53 – 0.59 µm); Band 4 is the Red band (0.64 ~ 0.67 µm); Band 5 is the Near-infrared band (0.85 ~ 0.88 µm); and Band 6 is the Short-wave Infrared (SWIR) band (1.57 ~ 1.65 µm).

For the forest file modeling, the NDVI map (Figure 5e) with five classes was generated, whereas for the NDWI map (Figure 5f) and the NDMI map (Figure 5g), seven classes were considered. These classes were determined using the natural breaks algorithm in ArcGIS 10.4 as mentioned above. For land use, this factor should be used for the forest fire modeling because it relates to anthropogenic activities, an important ignition sources for fire occurrence (Huesca et al., 2009). In this study, the land use map (Figure 5h) with eleven groups at a scale of 1:50,000, which was provided by the authority of the Gia Lai Province, was used. This map was produced in the national project on land use inventory carried out in 2013 by General Department of Land Administration of Vietnam.

Regarding weather, this is considered as one of the largest driven factor to burned areas (Abatzoglou and Kolden, 2013), and in most cases, forest fire happened based on intersections of ignition sources, fuel, and dry weather (Jolly et al., 2015). Therefore, climatic factors should be used. Specifically, four factors (temperature, wind speed, relative humidity, and were selected for this study because these factors have proven influencing both spread rates and intensities of fires other researches (Jolly et al., 2015). In this analysis, climatic data for the period 2007 ~ 2014 provided by CFSR (Climate Forecast System Reanalysis, available at https://www.ncdc. noaa.gov/) that consist of average maximum monthly temper-ature, average monthly wind speed, average monthly relative humidity, and average monthly total sum of precipitation were used. The temperature map (Figure 5i) was constructed with nine classes, whereas six classes were considered for the wind speed map (Figure 5j). For the relative humidity map, nine classes were adopted, and for the case of the rainfall map (Figure 5k), eight classes were employed. We determined these classes using the natural breaks algorithm (Tien Bui et al., 2017b) available in ArcGIS 10.4 that was mentioned above.

## 4. Proposed ICA-RVM for Modelling and Predicting Tropical Forest Fire Danger Using GIS

The overall structure of the proposed ICA-RVM model which is a combination of RVM and ICA algorithms is shown in Figure 6. ICA-RVM used a pattern recognition approach to distinguish the pixels in the study area into the forest fire class and the non-forest fire class. As result, probability of a pixel belongs to the forest fire class was used as forest fire danger index.

It is worth to notice that these data were acquired, processed, and integrated using ArcGIS 10.4 and IDRISI Selva 16 packages. The relevance vector machine (RVM) is available at http://www.miketipping.com/sparsebayes.htm, while the proposed ICA-RMV model was developed by the authors in Matlab. In addition, a geospatial tool developed in C++ also by the authors was used to transform forest fire danger indices into a rater format for implementation in ArcGIS package.

### 4.1. GIS Database, Training Dataset and Validation Dataset

In order to construct a machine learning based model for predicting tropical forest fire danger, it is necessary to establish

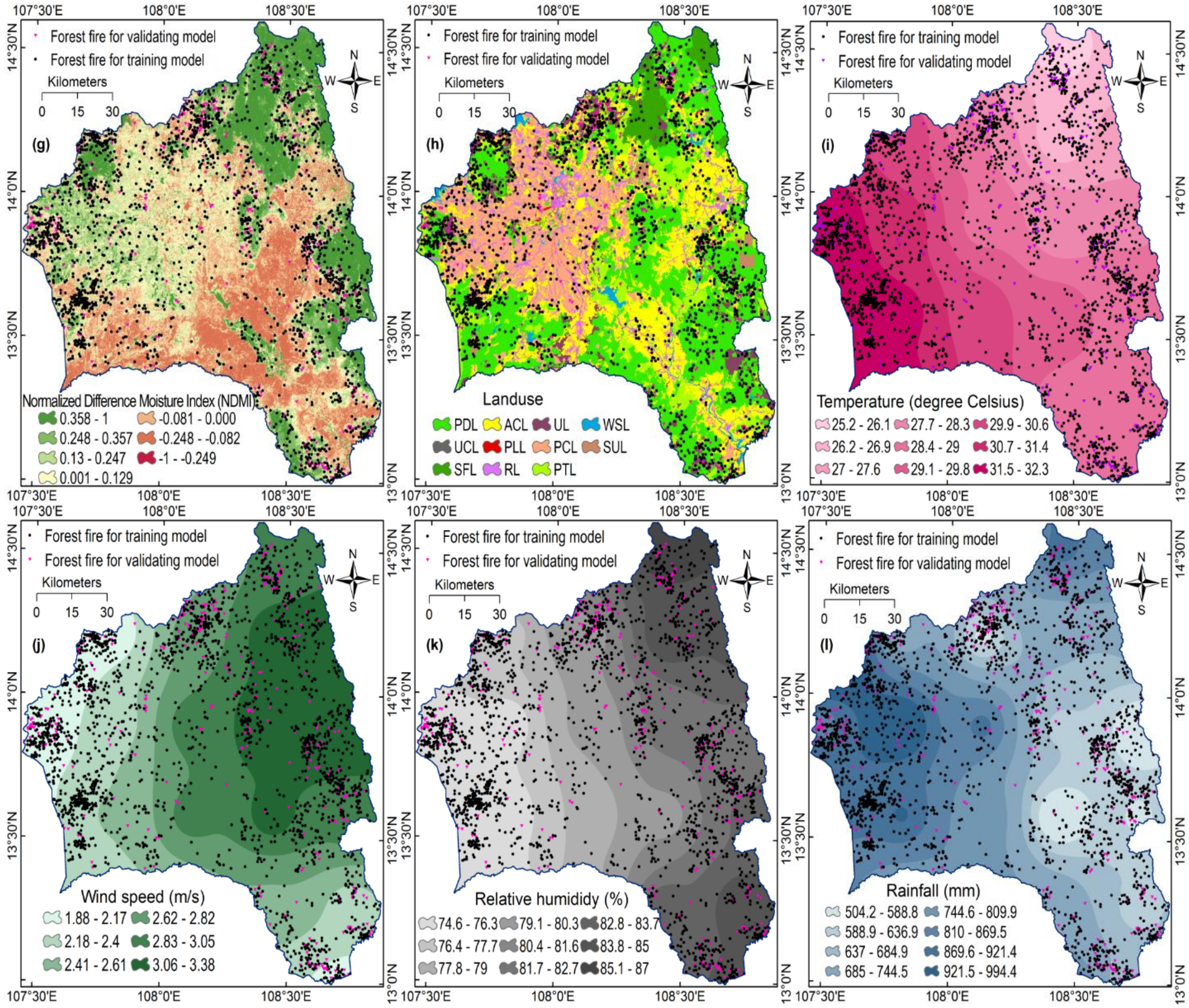


**Figure 5.** (continue); (g) NDWI map; (h) NDMI map; (g) NDWI map; (h) NDMI map; i) Temperature map; (j) Wind speed map; (k) Relative humidity map; and (l) Rainfall.

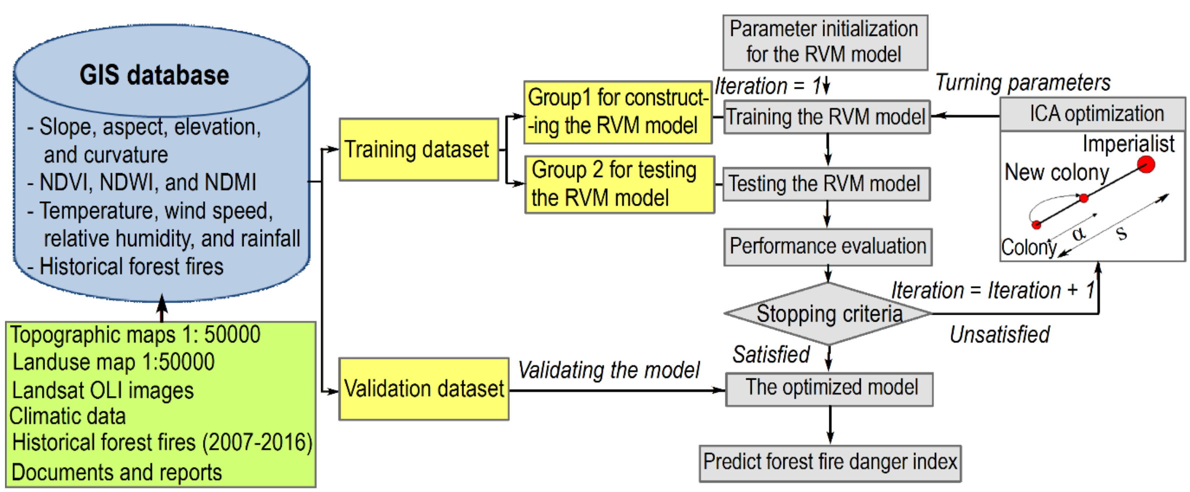


**Figure 6.** The proposed ICA-RVM model for the forest fire modeling in this study.

a GIS database. Accordingly, digital topographic maps and land use map at a scale of 1:50,000, Landsat-8 OLI images with a resolution of 30 m, climatic data (temperature, wind speed, re-lative humidity, and rainfall), and 2530 historical forest fires have been attained, processed, and integrated into the mentioned GIS database. All factors were converted into a raster format with 30 m resolution.

Because we formulate the forest fire danger modeling as a binary supervised learning task, therefore it is necessity to generate a training dataset and a validation dataset. The first dataset was used to train the proposed model whereas the second dataset was used to validate the model and confirm its prediction accuracy. Accordingly, the data of 2530 forest fire locations were split into two subsets, the first subset consists of 2118 fire locations, occurred from 2007 to 2015, was used for training model, whereas the second subset contains fire locations, occurred in 2016 only (412 forest fire locations), was used for the model validation. The same amounts of non-forest points were randomly generated in non-forest areas with NDVI is less than 0. The forest fire data were assigned to the fire class label of $C_1 = 1$ whereas the label of non-forest fire data were denoted as $C_2 = 0$.

Finally, a sampling process was conducted to extract the values of 12 ignition factors to build the training dataset (4236 samples) and the validation dataset (824 samples). To facilitate the modeling process, normalization of input data should be carried out (Hoang and Tien Bui, 2016b); therefore, the 12 factors were converted from categorical classes into continuous values within the range of 0.01 and 0.99 using a method described in Tien Bui et al. (2016e).

### 4.2. Determination of Objective Function for the ICA Optimization

The behavior of the RVM model is influenced by the RBF basis width that is explained in section 2.1; therefore, the RBF basis width should be carefully determined. In this research, the ICA optimization is utilized to optimize the RBF basis width. Accordingly, the fitness of the RVM model is measured using our proposed objective function (Equation 15). Thus, the training dataset (4236 samples) was further divided into two groups, Group 1 (70% or 2962 samples) was used for the construction of the RVM model, whereas Group 2 (30% or 1271 samples) that served as unknown patterns was used to test the RVM model.

The purpose of the aforementioned separation of the training dataset is to alleviate the potential overfitting problem. It is noted that one may simply identify the most suitable RBF basis width parameter by considering the model prediction performance on the whole the training dataset. However, the prediction result on the training set alone is not a good indicator of the model generalization due to overfitting issue (Hoang et al., 2016). Overfitting generally occurs when a model learns the training data very well but it predict poorly with the data outside the training set. Therefore, the data in Group 2 is utilized to penalize over-fitted model (Hoang et al., 2016):

$$f_{RVM} = \frac{1}{CR_{Group1} + CR_{Group2}} \tag{14}$$

where $CR_{Group1}$ and $CR_{Group2}$ denote classification rates (CR) of the two groups of interest. CR is computed using Equation 9.

### 4.3. Training and Validating the ICA-RVM Model

Since the objective function has been defined, the ICA optimization was carried out in the training process. We select a stopping criterion was 1000 iteration (Hosseini-Moghari et al., 2015). At the first iteration, the RBF basis width of RVM is generated randomly within the range of 0 and 1. During the searching process, ICA gradually recognized suitable values of the tuning parameter and discard inappropriate ones, and when the stopping criterion was reached, the ICA optimization was stopped and the most desirable RBF basis width has been identified. In the next step, the optimized ICA-RVM model is retrained with the whole training dataset to get the final ICA-RVM model and the prediction of outcome on the validation dataset was obtained.

Once the ICA-RVM model is successfully trained and validated, the final model is used to calculate the forest fire danger indices for all the pixels in the study area. These indices were converted to the rater format using the geospatial tool developed by the authors mentioned above, and then, open the result in ArcGIS software.

### 4.4. Software and Data Availability

It is noted that the integrated ICA-RVM model operates in Matlab environment. The machine learning model RVM is provided in the Sparse Bayesian Models toolbox of Tipping (2009). Meanwhile, the source code of ICA metaheuristic has been adopted and modified from Yarpiz (2016). The overall model has been programmed by the authors.

## 5. Results and Discussion

### 5.1. Feature Selection

In forest fire modeling, the performance of the prediction model could be degraded if noisy features exist in the training dataset; therefore, feature selection should be carried out (Tien Bui et al., 2017b). In this study, two feature selection techniques, Information Gain Ratio (IGR) and Pearson correlation (Tien Bui et al., 2016e) were used to detect potential noisy features to ensure the objective of selected features. Accordingly, predictive ability of the twelve ignition factors were quantified and evaluated, and factors with null-predictive value were considered irrelevant and being eliminated.

Since both IGR and Pearson correlation measure the correlation of each factor with the forest fire, they provide no information on if some factors may have their co-effects on the forest fire modeling. To assess the merit of the factors including their co-effects on the forest fire modeling, the wrapper evaluation method (Vafaei et al., 2018) combined with Random Forests (RF) classifier (Breiman, 2001) was further used. Accordingly, the RF with 500 trees as suggested by Stevens et al. (2015) and the classification rate (CR) as a statistical measure were used to assess the merit of each ignition factors.

The feature selection result is shown in Table 2. The result showed that NDVI has the highest predictive ability values (0.217 for IGR and 0.357 for Pearson correlation), followed by NDWI (0.100 for IGR and 0.32 for Pearson correlation), NDMI (0.069 for IGR and 0.278 for Pearson correlation), humidity (0.023 for IGR and 0.082 for Pearson correlation), wind speed (0.021 for IGR and 0.15 for Pearson correlation). In contrast, aspect and elevation has the lowest predictive ability (IGR is 0.0005 and Pearson correlation is 0.006). The predictive merit of the factors including their co-effects shows that NDVI has the highest predictive merit (0.238) whereas aspect and elevation have the lowest one (Table 1).

The finding is in agreement with Tien Bui et al. (2017b); (2016a) and is reasonable since NDVI is related to tree covers that affect the variability of fuel load, a main factor controlling the fire mechanism (Holsinger et al., 2016). NDWI estimates the leaf water content at canopy level and NDMI captures the variations of moisture in vegetated areas, therefore, it is understandable that they are among the most predictive factors for forest fires. Since no factor reveals non-predictive value, all the factors are considered as relevant factors for the modelling process.

**Table 1.** Predictive Ability of the Fire Ignition Factors Using Information Gain Ratio and Pearson Correlation with a 10-fold Cross Validation

| No. | Forest fire ignition factor | Predictive ability | | Predictive merit |
|---|---|---|---|---|
| | | Information Gain Ratio | Pearson correlation | |
| 1 | NDVI | 0.217 | 0.357 | 0.238 |
| 2 | NDWI | 0.100 | 0.32 | 0.155 |
| 3 | NDMI | 0.069 | 0.278 | 0.156 |
| 4 | Humidity (%) | 0.023 | 0.082 | 0.063 |
| 5 | Wind speed (m/s) | 0.021 | 0.150 | 0.080 |
| 6 | Temperature (°) | 0.016 | 0.083 | 0.074 |
| 7 | Slope (°) | 0.012 | 0.117 | 0.062 |
| 8 | Rainfall (mm) | 0.011 | 0.049 | 0.068 |
| 9 | Land use | 0.009 | 0.022 | 0.060 |
| 10 | Curvature | 0.005 | 0.015 | 0.048 |
| 11 | Aspect | 0.001 | 0.016 | 0.019 |
| 12 | Elevation (m) | 0.0005 | 0.006 | 0.019 |

### 5.2. Model Performance and Evaluation

Using the training dataset, the ICA-RVM model was successfully trained and the result is shown in Table 2 and Figure 7. The result shows that the ICA-RVM model has high goodness-of-fit with the training dataset. The classification rate is 76.82% indicating a good classification result, whereas AUC is 0.842 indicating that the global fit is 84.2%. Sensitivity is 73.83% indicating that 73.83% of the forest fires are correctly classified whereas specificity is 80.67% indicating that 80.67% of the non-forest fires are correctly classified. Kappa index is 0.536 indicating a moderate agreement between the training dataset and the estimated forest fires of the model.

The ICA-RVM model was validated using the validation. The result (Table 3 and Figure 7) shows that AUC is 0.793 indicating that the prediction capability of the model is 79.3%. The classification rate is 74.51% indicating an acceptable result, whereas Kappa index is 0.490 indicating a moderate agreement between the validation dataset and the predicted outcome of the model. Sensitivity and specificity are 70.78 and 79.88% indicating that the model correctly classifies 70.78% the forest fires and 79.88% the non-forest fire, respectively.

**Table 2.** Performance of the Proposed ICA-RVM Model, the SVM Model, and the RF Model Using the Training Dataset

| No | Statistical measure | ICA-RVM | SVM | RF |
|---|---|---|---|---|
| 1 | True positive | 1760 | 1822 | 1748 |
| 2 | True negative | 1494 | 1372 | 1432 |
| 3 | False positive | 358 | 296 | 370 |
| 4 | False negative | 624 | 746 | 686 |
| 5 | Sensitivity (%) | 73.83 | 70.95 | 71.82 |
| 6 | Specificity (%) | 80.67 | 82.25 | 79.47 |
| 7 | CR (%) | 76.82 | 75.40 | 75.07 |
| 8 | AUC | 0.842 | 0.813 | 0.830 |
| 9 | Kappa statistic | 0.536 | 0.508 | 0.501 |

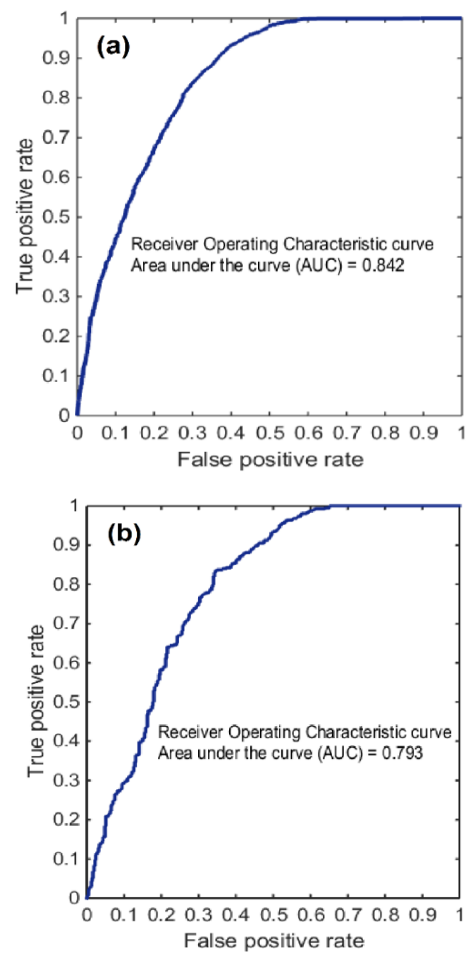


**Figure 7.** ROC curve and AUC of the ICA-RVM model using (a) the training dataset and (b) the validation dataset.

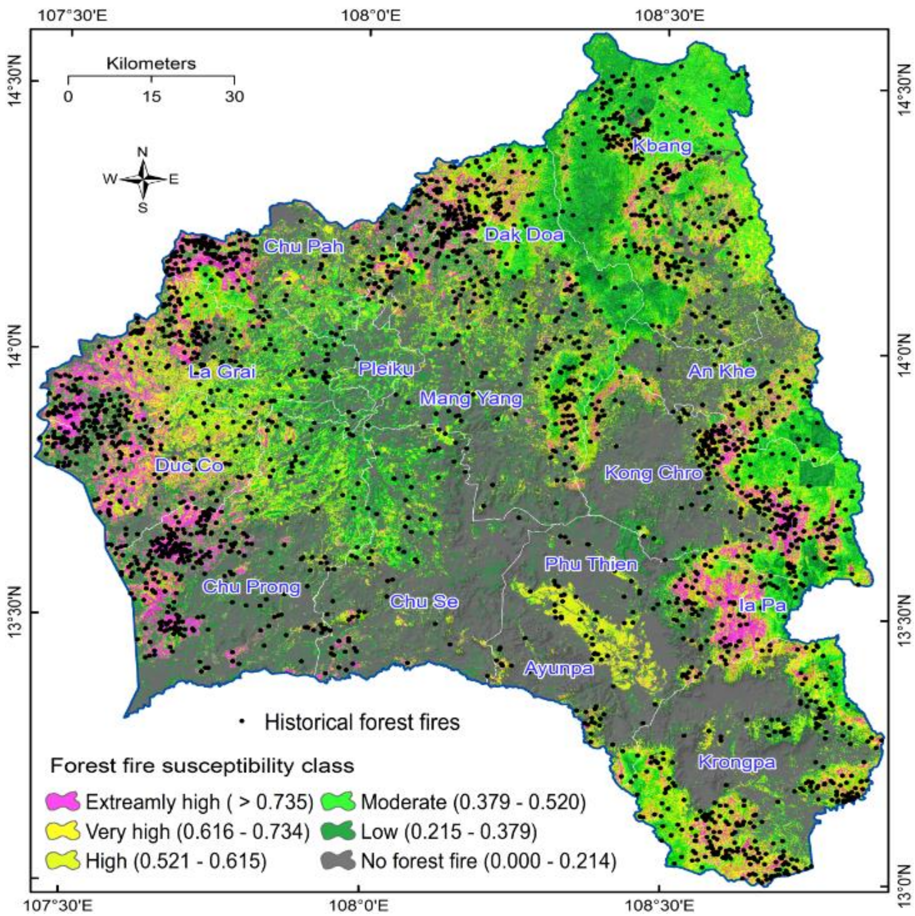


**Figure 8**. Forest fire danger map for the study area using the ICA-RVM model.

**Table 3.** Performance of the Proposed ICA-RVM Model, the SVM Model, the RF Model Using the Validation Dataset

| No | Statistical measure | ICA-RVM | SVM | RF |
|---|---|---|---|---|
| 1 | True positive | 344 | 359 | 340 |
| 2 | True negative | 270 | 252 | 259 |
| 3 | False positive | 68 | 53 | 72 |
| 4 | False negative | 142 | 160 | 153 |
| 5 | Sensitivity (%) | 70.78 | 69.17 | 68.97 |
| 6 | Specificity (%) | 79.88 | 82.62 | 78.25 |
| 7 | CR (%) | 74.51 | 74.15 | 72.69 |
| 8 | AUC | 0.793 | 0.786 | 0.790 |
| 9 | Kappa statistic | 0.490 | 0.483 | 0.454 |

### 5.3. Model Comparison

Because this is the first time the ICA-RVM model is proposed for the forest fire danger modelling, therefore, the validity of the model should be assessed and compared with benchmarks. For this purpose, we selected SVM and RF as two benchmarks because the first one has been recognized as an efficient method for modelling of complex real-world problems (Tien Bui et al., 2016b), whereas the second one is considered as a state-of-the art method for classification (Oliveira et al., 2012; Belgiu and Drăguţ, 2016).

For forest fire danger with SVM in this research, Radial Basic Function (RBF) kernel was used as suggested in Hoang and Tien Bui (2016a). Accordingly, the two turning parameters of SVM, the regularization and the RBF kernel width were derived using the grid-search method as in Tien Bui et al. (2016e). Consequently, the regularization equals 10 and the RBF kernel width equals 0.065 were found the best for the data at hand.

The training and validation results of SVM and RF in this study are shown in Tables 2 and 3. It could be seen that classification rate of the SVM model is 75.40% in the training dataset and 74.15% in the validation dataset, whereas classification rate of the RF model is 75.07% with the training dataset and is 72.69% in the validation dataset. These classification rates are lower than those of the ICA-RVM model. In addition, AUC of the ICA-RVM model is slightly better than the two benchmark models in the training dataset. Furthermore, Sensitivity of the SVM model and the RF model is slightly lower than that of the ICA-RVM model in both the training dataset and the validation dataset. The other parameters are in the Tables 2 and 3. Overall, based on the above analysis, it could be concluded that the ICA-RVM model performed best with the forest fire data in this study. It is noted that compared to the SVM and RF models, the proposed ICA-RVM has achieved improvements of 0.36 and 1.82% in terms of classification rate. Forest fire is complex phenomenon; thus, even a small improvement in the prediction accuracy can result in a significantly better forest fire hazard management and prevention in a regional scale.

### 5.4. Generation of A Forest Fire Danger Map

Since the proposed ICA-RVM model was successfully trained and validated, the model was then used to compute forest fire danger indices for all the pixels of the study area. It is noted that the ICA-RVM model estimated two probability of fire danger for each pixel that are probabilities belongs to the forest fire class and the non-forest fire class. We used the probability of pixel belongs to the forest fire class for the forest fire danger index. These fire danger indices were then converted to a raster format using the geospatial tool mentioned in Section 4 and developed by the authors to open in ArcGIS 10.4.

The forest fire danger map (Figure 8) was cartographically visualized by means of six classes (Tien Bui et al., 2017b): extremely high (10%, 1551.5 $km^2$), very high (10%, 1551.5 $km^2$), high (10%, 1551.5 $km^2$), moderate (15%, 2326.9 $km^2$), low (15%, 2327.1 $km^2$), and no-forest fire (40%, 6204.2 $km^2$). These thresholds for these classes were determined based on a graphic curve (Tien Bui et al., 2017b) that was constructed using percentage of the forest fires versus percentage of the forest fire danger map (Figure 9).

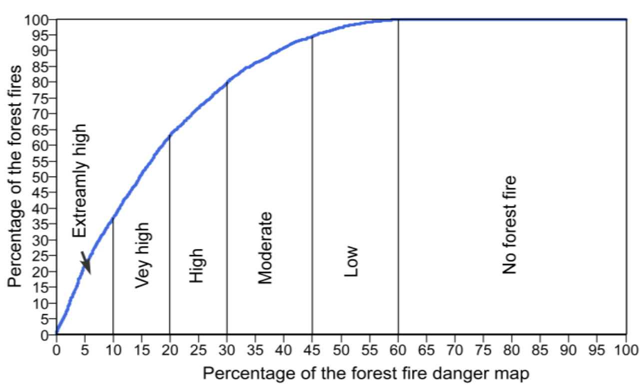


**Figure 9.** Forest fire danger map for the study area using the ICA-RVM model.

## 6. Concluding Remarks

In this research, a new hybrid intelligence system approach based on RVM and ICA for modeling and predicting tropical forest fire danger is proposed and verified with a case study at the tropical forest of the Gia Lai province in the Central Highland (Vietnam). According to current literature, RVM and ICA are state-of the art soft computing techniques that have not been explored for modeling forest fire danger, a typical non-linear and complex real world problem. The advantage of the RVM model is that this technique uses Bayesian framework to infer probability of forest fire danger, therefore we used the inferred probabilities as fire danger indices and compiled the forest fire danger map. Since the RVM model is influenced by the RBF band width, the ICA is utilized to search the most desirable RBF basis width in the optimization process.

Using the forest fire database of the study area, the ICA-RVM model was successfully trained and validated. Experimental results demonstrated that the behavior of the RVM model is strongly dependent on the RBF band width. The performance of the ICA-RVM model on both the training dataset and the validation dataset indicating that ICA is a good algorithm should be considered for optimizing the RVM model' parameters. The performance of the ICA-RVM model is further compared to those produced from two benchmarks, the SVM model and the RF model, using the same data. Since the ICA-RVM model performed better, it could be concluded that the proposed ICA-RVM model is a valid tool that should be considered for modeling of forest fire danger.

Because performance of forest fire danger models is influenced not only by the method used but also by the selection of ignition factors. Therefore, selection of ignition factors is an interested issue, particularly. In this study, twelve ignition factors were selected based on analyses of the historical forest fires in the study areas, and in addition, these are popular ignition factors in literature. All of these factors revealed the predictive ability objectively and good performance of the ICA-RVM model demonstrating that these factors were selected, processed, and coded successfully. In addition, the historical forest fires were temporally separated with the training dataset consists of these forest fires 2007 ~ 2015, whereas the validation dataset contains these forest fires occurred in 2016 only. Consequently, the good prediction outcome of the ICA-RVM model on the fires occurred in 2016 indicate that the ICA-RVM model is efficient and possess a good predictive capability. Aerial interpretation of the resulting map (Figure 8) shows that areas in Chu Pah, La Grai, Duc Co, and Ia Pa have high probability of forest fire danger. Therefore, these areas should be critically considered for developing prevention measures. Whereas, areas at Chu Se and Kong Chro have lower probabilities of forest fires due to low forest canopy in these areas.

The limitation of this work is that only the metaheuristic of ICA was explored and investigated for optimization of the RVM model, therefore, the performance of the model may be enhanced by employing other alternative metaheuristic algorithms. In addition, only 12 ignition factors were used. Therefore, further factors should be considered to improve the prediction performance of the model. Despite such limitation, the result obtained from this works has been demonstrated to be useful for forest management and planning in forest fire danger areas. Future extensions of the current study may include applying the proposed ICA-RVM for constructing forest fire susceptibility map in other regions, investigating other advanced machine learning solution for forest fire pattern recognition, as well as integrating novel feature selection methods for potentially ameliorating the model predictive accuracy.

**Acknowledgements.** This research was supported by the Geographic Information Science research group and Ton Duc Thang University, Vietnam.

## References


Abatzoglou, J.T. and Kolden, C.A. (2013). Relationships between climate and macroscale area burned in the western United States. *Int. J. Wildland Fire*, 22(7), 1003-1020. https://doi.org/10.1071/WF13019

Alcasena, F.J., Salis, M., Nauslar, N.J., Aguinaga, A.E., and Vega-García, C. (2016). Quantifying economic losses from wildfires in black pine afforestations of northern Spain. *For. Policy Econ.*, 73, 153-167. https://doi.org/10.1016/j.forpol.2016.09.005

Atashpaz-Gargari, E. and Lucas, C. (2007). Imperialist competitive algorithm: An algorithm for optimization inspired by imperialistic competition, In Proc. of 2007 IEEE Congress on Evolutionary Computation, 4661-4667. https://doi.org/10.1109/CEC.2007.4425083

Belgiu, M. and Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.*, 114, 24-31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Bishop, C.M. and Tipping, M.E. (2000). Variational Relevance Vector Machines. *Proc. of the 16th Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 46-53.

Bisquert, M., Caselles, E., Sánchez, J.M., and Caselles, V. (2012). 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(8), 1025-1029. https://doi.org/10.1071/WF11105

Breiman, L. (2001). Random forests. *Mach. learning*, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

Brunette, M., Couture, S., and Pannequin, F. (2017). Is forest insurance a relevant vector to induce adaptation efforts to climate change? *Ann. For. Sci.*, 74(2), 41. https://doi.org/10.10 07/s13595-017-0639-9

Bui, K.T.T., Tien Bui, D., Zou, J., Van Doan, C., and Revhaug, I. (2016). A novel hybrid artificial intelligent approach based on neural fuzzy inference model and particle swarm optimization for horizontal displacement modeling of hydropower dam. *Neural Comput. Appl.*, 1-12.

Carvalho, J.P., Carola, M., and Tome, J.A.B. (2006). Forest fire modelling using rule-based fuzzy cognitive maps and voronoi based cellular automata, *Proc. of NAFIPS 2006 - 2006 Annual Meeting of the North American Fuzzy Information Processing Society*, Quebec, Canada, 217-222. https://doi.org/10.1109/NAFIPS.2006.365411

Cary, G.J., Flannigan, M.D., Keane, R.E., Bradstock, R.A., Davies, I.D., and Lenihan, J.M. (2009). Relative importance of fuel management, ignition management and weather for area burned: evidence from five landscape-fire-succession models. International *J. Wildland Fire*, 18(2), 147-156. https://doi.org/10.1071/WF07085

CGIAR, (2016). *The Drought Crisis in the Central Highlands of Vietnam: Assessment Report*, CGIAR Research Centers in Southeast Asia

Chuvieco, E., Aguado, I., Yebra, M., Nieto, H., Salas, J., and Martín, M.P., et al. (2010). Development of a framework for fire risk assessment using remote sensing and geographic information system technologies. *Ecol Model.*, 221(1), 46-58. https://doi.org/10.1016/j.ecolmodel.2008.11.017

Conard, S.G., Doerr, S., and Foster, J. (2017). Twenty-five years of international journal of wildland fire. *Int. J. Wildland Fire*, 25(12). https://doi.org/10.1071/WFv25n12_FO

DeFries, R. and Townshend, J. (1994). NDVI-derived land cover classifications at a global scale. *Int. J. Remote Sens.*, 15(17), 3567-3586. https://doi.org/10.1080/01431169408954345

Dennison, P.E., Roberts, D., Peterson, S., and Rechel, J. (2005). Use of normalized difference water index for monitoring live fuel moisture. *Int. J. Remote Sens.*, 26(5), 1035-1042. https://doi.org/10.1080/0143116042000273998

Doerr, S.H. and Santín, C. (2016). Global trends in wildfire and its impacts: perceptions versus realities in a changing world. *Philos. Trans. R. Soc. Lond.*, 371(1696). https://doi.org/10.1098/rstb.2015.0345

Dupire, S., Curt, T., and Bigot, S. (2017). Spatio-temporal trends in fire weather in the French Alps. *Sci. Total Environ.*, 595, 801-817. https://doi.org/10.1016/j.scitotenv.2017.04.027

Fusco, E.J., Abatzoglou, J.T., Balch, J.K., Finn, J.T., and Bradley, B.A. (2016). Quantifying the human influence on fire ignition across the western USA. *Ecol. Appl.*, 26(8), 2390-2401. https://doi.org/10.1002/eap.1395

Gao, B.-C. (1996). NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.*, 58(3), 257-266. https://doi.org/10.1016/S0034-4257(96)00067-3

García, M., Chuvieco, E., Nieto, H., and Aguado, I. (2008). Combining AVHRR and meteorological data for estimating live fuel moisture content. *Remote Sens. Environ.*, 112(9), 3618-3627. https://doi.org/10.1016/j.rse.2008.05.002

Grishin, A.M. and Filkov, A.I. (2011). A deterministic-probabilistic system for predicting forest fire hazard. *Fire Saf. J.*, 46(1-2), 56-62. https://doi.org/10.1016/j.firesaf.2010.09.002

GSO, (2015). General Statistics Office of Vietnam. http://www.gso.gov.vn.

Güngöroğlu, C. (2017). Determination of forest fire risk with fuzzy analytic hierarchy process and its mapping with the application of GIS: The case of Turkey/Çakırlar. *Hum. Ecol. Risk Assess.*, 23(2), 388-406. https://doi.org/10.1080/10807039.2016.1255136

Hilton, J., Miller, C., Sharples, J., and Sullivan, A. (2017). Curvature effects in the dynamic propagation of wildfires. *Int. J. Wildland Fire*, 25(12), 1238-1251. https://doi.org/10.1071/ WF16070

Hoang, N.D. and Tien Bui, D. (2016a). A novel relevance vector machine classifier with cuckoo search optimization for spatial prediction of landslides. *J. Comput. Civ. Eng.*, 30(5). https://doi.org/10.1061/(ASCE)CP.1943-5487.0000557

Hoang, N.D. and Tien Bui, D. (2016b). 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., and 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. https://doi.org/10.1016/j.asoc.2016.04.031

Holsinger, L., Parks, S.A., and Miller, C. (2016). Weather, fuels, and topography impede wildland fire spread in western US landscapes. *For. Ecol. Manage.*, 380, 59-69. https://doi.org/10.1016/j.foreco.2016.08.035

Hosseini-Moghari, S.M., Morovati, R., Moghadas, M., and Araghinejad, S. (2015). Optimum operation of reservoir using two evolutionary algorithms: imperialist competitive algorithm (ICA) and cuckoo optimization algorithm (COA). *Hydrol. Water Resour. Manage.*, 29(10), 3749-3769. https://do i.org/10.1007/s11269-015-1027-6

Hosseini, S. and Al Khaled, A. (2014). A survey on the Imperialist Competitive Algorithm metaheuristic: Implementation in engineering domain and directions for future research. *Appl. Soft Comput.*, 24, 1078-1094. https://doi.org/10.1016/j.asoc.2014.08.024

Huesca, M., Litago, J., Palaciosorueta, A., Montes, F., Seba-stiánlópez, A., and Escribano, P. (2009). Assessment of forest fire seasonality using MODIS fire potential: A time series approach. *Agric. For. Meteorol.*, 149(11), 1946-1955. https://doi.org/10.1016/j.agrformet.2009.06.022

Jenks, G. (1977). Optimal data classification for choropleth maps, *Dept. Geogr.*, University of Kansas, USA.

Jolly, W.M. et al. (2015). Climate-induced variations in global wildfire danger from 1979 to 2013. Nat. Commun., 6. https://doi.org/10.1038/ncomms8537

Kaveh, A. and Talatahari, S. (2010). Optimum design of skeletal structures using imperialist competitive algorithm. *Comput. Struct.*, 88(21-22), 1220-1229. https://doi.org/10.1016/j.compstruc.2010.06.011

Koetz, B., Morsdorf, F., van der Linden, S., Curt, T., and All-göwer, B.

(2008). Multi-source land cover classification for forest fire management based on imaging spectrometry and LiDAR data. *For. Ecol. Manage.*, 256(3), 263-271. https://doi.org/10.1016/j.foreco.2008.04.025

Krasnow, K., Schoennagel, T., and Veblen, T.T. (2009). Forest fuel mapping and evaluation of LANDFIRE fuel maps in Boulder County, Colorado, USA. *For. Ecol. Manage.*, 257(7), 1603-1612. https://doi.org/10.1016/j.foreco.2009.01.020

Maki, M., Ishiahra, M., and Tamura, M. (2004). Estimation of leaf water status to monitor the risk of forest fires by using remotely sensed data. *Remote Sens. Environ.*, 90(4), 441-450. https://doi.org/10.1016/j.rse.2004.02.002

Marandi, A., Wang, Z., Takata, K., Byer, R.L., and Yamamoto, Y. (2014). Network of time-multiplexed optical parametric oscillators as a coherent Ising machine. *Nat. Photonics*, 8(12), 937. https://doi.org/10.1038/nphoton.2014.249

Massada, A.B., Syphard, A.D., Hawbaker, T.J., Stewart, S.I., and Radeloff, V.C. (2011). Effects of ignition location models on the burn patterns of simulated wildfires. *Environ. Model. Software*, 26(5), 583-592. https://doi.org/10.1016/j.envsoft.2010.11.016

Ministry of Agriculture and Rural Development of Vietnam, (2016). The Vietnam's FireWatch system for online monitoring and management of forest fires, http://www.kiemlam.org.vn/firewatchvn.

Nabney, I.T. (1999). Efficient training of RBF networks for classification. *Proc. of the Ninth International Conference on Artificial Neural Networks ICANN 99*, Edinburgh, Scotland, 1, 210-215. https://doi.org/10.1049/cp:19991110

Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A., and Pereira, J.M.C. (2012). Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random *Forest. For. Ecol. Manage.*, 275, 117-129. https://doi.org/10.1016/j.foreco.2012.03.003

Peterson, A.T., Papeş, M., and Soberón, J. (2008). Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecol. Model.*, 213(1), 63-72. https://doi.org/10.1016/j.ecolmodel.2007.11.008

Pham, B.T., Pradhan, B., Tien Bui, D., Prakash, I., and Dholakia, M.B. (2016). A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India). *Environ. Model. Software*, 84, 240-250. https://doi.org/10.1016/j.envsoft.2016.07.005

Pierce, A.D., Farris, C.A., and Taylor, A.H. (2012). Use of random forests for modeling and mapping forest canopy fuels for fire behavior analysis in Lassen Volcanic National Park, California, USA. *For. Ecol. Manage.*, 279, 77-89. https://doi.org/10.1016/j.foreco.2012.05.010

Pimont, F., Dupuy, J.L., and Linn, R. (2012). Coupled slope and wind effects on fire spread with influences of fire size: a numerical study using FIRETEC. *Int. J. Wildland Fire*, 21(7), 828-842. https://doi.org/10.1071/WF11122

Pourghasemi, H.r., Beheshtirad, M., and 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(2), 861-885. https://doi.org/10.1080/19475705.2014.984247

Pourtaghi, Z.S., Pourghasemi, H.R., Aretano, R., and Semeraro, T. (2016). Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques. *Ecol. Indicators*, 64, 72-84. https://doi.org/10.1016/j.ecolind.2015.12.030

Pourtaghi, Z.S., Pourghasemi, H.R., and Rossi, M. (2015). Forest fire susceptibility mapping in the Minudasht forests, Golestan province, Iran. *Environ. Earth Sci.*, 73(4), 1515-1533. https://doi.org/10.1007/s12665-014-3502-4

Renard, Q., Pélissier, R., Ramesh, B.R., and Kodandapani, N. (2012). Environmental susceptibility model for predicting forest fire occurrence in the Western Ghats of India. *Int. J. Wildland Fire*, 21(4), 368-379. https://doi.org/10.1071/WF10109

Sadowski, L. and Nikoo, M. (2014). Corrosion current density prediction in reinforced concrete by imperialist competitive algorithm. *Neural Comput. Appl.*, 25(7), 1627-1638. https://doi.org/10.1007/s00521-014-1645-6

Samui, P. (2012). Application of relevance vector machine for prediction of ultimate capacity of driven piles in cohesionless soils. *Geotech. Geol. Eng.*, 30(5), 1261-1270. https://doi.org/10.1007/s10706-012-9539-9

Sanjuan, G., Brun, C., Margalef, T., and Cortés, A. (2016). Determining map partitioning to minimize wind field uncertainty in forest fire propagation prediction. *J. Comput. Sci.*, 14, 28-37. https://doi.org/10.1016/j.jocs.2016.01.006

Shirzadi, A., Bui, D.T., Pham, B.T., Solaimani, K., Chapi, K., Kavian, A., Shahabi, H., and Revhaug, I. (2017). Shallow landslide susceptibility assessment using a novel hybrid intelligence approach. *Environ. Earth Sci.*, 76(2), 60. https://doi.org/10.1007/s12665-016-6374-y

Stevens, F.R., Gaughan, A.E., Linard, C., and Tatem, A.J. (2015). Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLoS One*, 10(2), e0107042. https://doi.org/10.1371/journal.pone.0107042

Teodoro, A., Duarte, L., Sillero, N., Gonçalves, J.A., Fonte, J., Gonçalves-Seco, L., Pinheiro da Luz, L.M., and dos Santos Beja, N.M.R. (2015). An integrated and open source GIS environmental management system for a protected area in the south of Portugal, *Proc. of the SPIE*, pp. 96440U-96440U-12.

Teodoro, A.C. and Duarte, L. (2013). Forest fire risk maps: a GIS open source application - a case study in Norwest of Portugal. *Int. J. Geogr. Inf. Sci.*, 27(4), 699-720. https://doi.org/10.1080/13658816.2012.721554

Tien Bui, D., Tuan, T.A., Hoang, N.D., Thanh, N.Q., Nguyen, D.B., Van Liem, N., and 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(2), 447-458. https://doi.org/10.1007/s10346-016-0711-9

Tien Bui, D., Bui, Q.T., Nguyen, Q.P., Pradhan, B., Nampak, H., and 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. https://doi.org/10.1016/j.agrformet.2016.11.002

Tien Bui, D., Le, K.-T., Nguyen, V., Le, H., and 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. *Remote Sens.*, 8(4), 347. https://doi.org/10.3390/rs8040347

Tien Bui, D., Pham, T.B., Nguyen, Q.-P., and Hoang, N.-D. (2016b). Spatial prediction of rainfall-induced shallow landslides using hybrid integration approach of Least-Squares Support Vector Machines and differential evolution optimization: a case study in Central Vietnam. *Int. J. Digit. Earth*, 9(11), 1077-1097. https://doi.org/10.1080/17538947.2016.1169561

Tien Bui, D., Pradhan, B., Nampak, H., Bui, Q.T., Tran, Q.A., and Nguyen, Q.P. (2016c). Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibility modelling in a high-frequency tropical cyclone area using GIS. *J. Hydrol.*, 540, 317-330. https://doi.org/10.1016/j.jhydrol.2016.06.027

Tien Bui, D., Tuan, T.A., Klempe, H., Pradhan, B., and Revhaug, I. (2016e). Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides*, 13, 361-378. https://doi.org/10.1007/s1034

6-015-0557-6
Tipping, M.E. (2000). The relevance vector machine. *Adv. in Neural Inf. Process. Syst.*, 12(3), 652-658.
Tipping, M.E. (2001). Sparse bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.*, 1, 211-244
Tipping, M.E. (2009). Sparse Bayesian Models (and the RVM). http://www.miketipping.com/sparsebayes.htm.
Tipping, M.E. and Faul, A.C. (2003). Fast marginal likelihood maximisation for sparse Bayesian models. *Proc. of the Ninth International Workshop on Artificial Intelligence and Statistics*, Key West, FL, Jan 3-6.
Vafaei, S. et al. (2018). Improving accuracy estimation of forest aboveground biomass based on incorporation of ALOS-2 palsar-2 and sentinel-2a imagery and machine learning: A case study of the hyrcanian forest area (Iran). *Remote Sens.*, 10(2), 172. https://doi.org/10.3390/rs10020172
Van, N.K., Ly, P.T., and Hong, N.T. (2014). Bioclimatic map of Tay Nguyen at scale 1: 250,000 for setting up sustainable ecological economic models. *Vietnam J. Earth Sci.*, 36(4), 504-514.
Vapnik, V.N. (1998). *Statistical Learning Theory*. John Wiley & Sons, Inc., Printed in the United States of America.
Verde, J. and Zêzere, J. (2010). Assessment and validation of wildfire susceptibility and hazard in Portugal. *Nat. Hazards Earth Syst. Sci.*, 10(3), 485-497. https://doi.org/10.5194/nhess-10-485-2010
Vlassova, L. and Pérez-Cabello, F. (2016). Effects of post-fire wood management strategies on vegetation recovery and land surface temperature (LST) estimated from Landsat images. Int. *J. Appl. Earth Obs. Geoinf.*, 44, 171-183. https://doi.org/10.1016/j.jag.2015.08.011
Wu, Z., He, H.S., Yang, J., Liu, Z., and Liang, Y. (2014). Relative effects of climatic and local factors on fire occurrence in boreal forest landscapes of northeastern China. *Sci. Total Environ.*, 493, 472-480. https://doi.org/10.1016/j.scitotenv.2014.06.011
Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int. J. Remote Sens.*, 27(14), 3025-3033. https://doi.org/10.1080/01431160600589179
Yarpiz. (2016). Imperialist Competitive Algorithm (ICA) in MATLAB, http://yarpiz.com/247/ypea118-imperialist-competi tive-algorithm.
You, W., Lin, L., Wu, L., Ji, Z., Yu, J., Zhu, J., Fan, Y., and He, D. (2017). Geographical information system-based forest fire risk assessment integrating national forest inventory data and analysis of its spatiotemporal variability. *Ecol. Indicators*, 77, 176-184. https://doi.org/10.1016/j.ecolind.2017.01.042
Zylstra, P., Bradstock, R.A., Bedward, M., Penman, T.D., Doherty, M. D., and Weber, R. O., et al. (2016). Biophysical mechanistic modelling quantifies the effects of plant traits on fire severity: species, not surface fuel loads, determine flame dimensions in eucalypt forests. *PloS one*, 11(8), e0160715. https://doi.org/10.1371/journal.pone.0160715