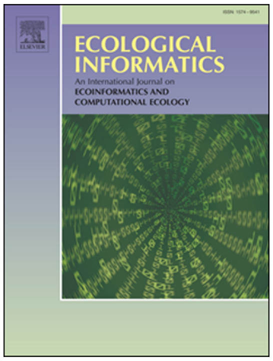


Spatial pattern assessment of tropical forest fire danger at Thuan Chau area (Vietnam) using GIS-based advanced machine learning algorithms: A comparative study

Nguyen Ngoc-Thach, Dang Bao-Toan Ngo, Pham Xuan-Canh, Nguyen Hong-Thi, Bui Hang Thi, Hoang NhatDuc, Tien Bui Dieu

PII:	S1574-9541(18)30085-2
DOI:	doi:10.1016/j.ecoinf.2018.05.009
Reference:	ECOINF 861
To appear in:	*Ecological Informatics*
Received date:	26 March 2018
Revised date:	27 May 2018
Accepted date:	28 May 2018

Please cite this article as: Nguyen Ngoc-Thach, Dang Bao-Toan Ngo, Pham Xuan-Canh, Nguyen Hong-Thi, Bui Hang Thi, Hoang NhatDuc, Tien Bui Dieu , Spatial pattern assessment of tropical forest fire danger at Thuan Chau area (Vietnam) using GIS-based advanced machine learning algorithms: A comparative study. Ecoinf (2017), doi:10.1016/j.ecoinf.2018.05.009

Spatial Pattern Assessment of Tropical Forest Fire Danger at Thuan Chau area (Vietnam) using GIS-Based Advanced Machine Learning Algorithms: A comparative study

Author: **Ngoc-Thach, Nguyen**, Ph.D
Position: Associate Professor
Affiliation: Faculty of Geography, VNU University of Science
Address: 334 Nguyen Trai, Thanh Xuan, Hanoi, Vietnam
Email: nguyenngocthachhus@gmail.com.

Author: **Bao-Toan Ngo, Dang**
Affiliation: Faculty of Geography, VNU University of Science
Address: 334 Nguyen Trai, Thanh Xuan, Hanoi, Vietnam
Email: toanrs@yahoo.com
Author: **Xuan-Canh, Pham**
Affiliation: Faculty of Geography, VNU University of Science
Address: 334 Nguyen Trai, Thanh Xuan, Hanoi, Vietnam
Email: phamxuancanhhus@gmail.com

Author: **Hong-Thi, Nguyen**
Affiliation: Faculty of Geoloy, VNU University of Science
Address: 334 Nguyen Trai, Thanh Xuan, Hanoi, Vietnam
Email: nguyenthihong.hus@gmail.com

Author: **Hang Thi, Bui**
Affiliation: Faculty of Geography, VNU University of Science
Address: 334 Nguyen Trai, Thanh Xuan, Hanoi, Vietnam
Email: buithihang.hus@gmail.com

Author: **NhatDuc, Hoang**, Ph.D
Position: Lecture
Affiliation: Faculty of Civil Engineering, Institute of Research and Development, Duy Tan University,
Address: P809, 03 Quang Trung, Da Nang 550000, Vietnam
Email: hoangnhatduc@dtu.edu.vn

Author: **Dieu, Tien Bui***, Ph.D.
Position: Associate Professor
Affiliation: GIS Group, Department of Business and IT, University College of Southeast Norway
Address: Gullbringvegen 36, N-3800 Bø i Telemark, Norway
Email: Dieu.T.Bui@usn.no; BuiTienDieu@gmail.com,
* Corresponding author

Abstract

Thuan Chau is a serious district affected by forest fire in Vietnam, especially in 2016; however, no forest fire prediction research has been conducted for this region. Thus, knowledge of spatial patterns of fire danger of the district plays a key role in forest succession and ecological implications. This study's aim was to analyze the spatial pattern of fire danger for the tropical forest of Thuan Chau district using advanced machine learning algorithms, Support Vector Machine classifier (SVMC), Random Forests (RF), and Multilayer Perceptron Neural Network (MLP-Net). For this purpose, a GIS database for the study area was established with 564 forest fire locations and ten forest fire variables. Then, Pearson correlation method was used to assess the correlation of the variables with the forest fire. In the next step, three forest fire danger models, SVMC, RF, and MLP-Net, were trained and validated. Finally, global performance of these models was assessed using the classification accuracy (ACC), Kappa statistics (KS), Area under the curve (AUC). In addition, Wilcoxon signed-rank test was employed to check the prediction performance of these models. The result shows the three models performed well; however, the MLP-Net model has the highest prediction performance (ACC=81.7, KS = 0.633, and AUC = 0.894), followed by the RF model (ACC=81.1, KS = 0.621, and AUC = 0.883), and the SVMC model (ACC=80.2, KS = 0.604, and AUC = 0.867). The result in this study is useful for the local authority and forest manager in forest management and fire suppression.

Key words: Forest fire; Support vector machines; Random forests; Neural networks; GIS; Vietnam

1. Introduction

Forest fire seems to be inevitable and plays an important role in vegetation succession and landscape transformation (Chuvieco et al., 2010; Hong et al., 2018; Wang et al., 2017). This phenomenon has been widely recognized as a serious hazard that brings about negative impacts on the environment and the society in many countries around the globe (Fox et al., 2016; Wenhua, 2004). In recent years, high-intensity forest fires have been observed in many countries i.e. USA (Odion et al., 2014), Sweden (Brown and Giesecke, 2014), China (Wu et al., 2014), Portugal (Teodoro and Amaral, 2017), and Indonesia (Sumarga, 2017), due to effects of extreme weather events i.e. long dry period combined with dry and hot weather conditions (Mason et al., 2017). It is anticipated that the occurrence of forest fire will be increased in the future due to change of climate is continuing (Brown et al., 2017; Whitburn et al., 2016). Therefore, development of reliable prediction models of forest fire danger is important for public safety, forest management, and suppression planning.

Development of high accuracy prediction models of forest fire is still a difficult task because forest fire is typical a nonlinear and complex process that is governed by many influencing factors. Therefore, it is challenging to model and predict the occurrence of forest fires. Thus, fire occurrence is influenced by not only ignition factors and biomass fuels but also weather conditions for combustion (Pettinari and Chuvieco, 2017). Changes of climate influence these factors both multiple timescales and in complex ways (Moritz et al., 2012). Therefore, various models have been proposed for forest fire danger prediction, varying from simple statistical techniques i.e. Poisson regression (Wotton et al., 2003), geographically weighted regression (Koutsias et al., 2010), logistic regression (Arndt et al., 2013), and linear regression (Oliveira et al., 2012) to more complex models i.e. Pareto distribution (Bermudez et al., 2009), favorability functions (Verde and Zêzere, 2010), and an approach based on numerical simulation (Conedera et al., 2011), and complex sophisticated mathematical models i.e. ELMFIRE (Lautenberger, 2013). Brief reviews of these models could be seen in (Pimont et al., 2016; Sileshi, 2014; Teodoro and Duarte, 2013; Tien Bui et al., 2017b).

In recent years, machine learning has been considered for predicting the spatial pattern of fire dangers such as neural networks (Cheng and Wang, 2008; Satir et al., 2015), support vector machines (Sakr et al., 2011), random forests (Arpaci et al., 2014; Oliveira et al., 2012), logistic regression classifier with kernel function (Tien Bui et al., 2016a), and neural fuzzy (Tien Bui et al., 2017b). Common conclusion from the above researches is that machine learning models have proven abilities to deliver better results (Massada et al., 2013; Tien Bui et al., 2017b). Nevertheless, it is a debate on which method or technique is the best for modeling forest fire. Therefore, comparison of methods and techniques is highly necessary to gather reasonable conclusions for forest fire prediction.

The main objective of this research was to analyze the spatial pattern of fire danger for the tropical forest of Thuan Chau district using advanced machine learning algorithms, Support Vector Machine classifier (SVMC), Random Forests (RF), and Multilayer Perceptron Neural Network (MLP-Net). SVMC and RF are selected because they are state-of-the art machine learning techniques that have proven efficient for environmental modeling (Abdel-Rahman et al., 2014; Pham et al., 2016a; Pourtaghi et al., 2016; Were et al., 2015), whereas MLP-Net is selected because it is considered to be the most widely used for forest fire danger (Cheng and Wang, 2008; Satir et al., 2015; Vasconcelos et al., 2001).

Thuan Chau district was selected because this was the most serious district affected by forest fire, especially in 2016, in Vietnam. However, no research on forest fire modeling has been carried out. Should be noted that during the last two decades, forest fire has been a big problem in Vietnam mainly due to both the change of climate and anthropogenic issues i.e. burning in rice fields, hunting, and clear cut logging activities (Le et al., 2014). Thus, together with farming, clear-cut logging, and development activities, forest fire is one of the key drivers for the deforestation. Forest fires are a cause for severe pollutions, especially in the dry season (November to April), with more than 6 million hectares forest are in a high probability of forest fire risk across the country (Pham et al., 2017b), especially in the Son La province, including the Thuan Chau district. Therefore, study on forest fire, including developments of high accuracy models for fire prediction in order to determine appropriate prevention measures, is an urgent task.

2. Study area and geospatial data

2.1 Description of the Thuan Chau district

Thuan Chau district (Son La province) is located in northwest region of Vietnam, around 300 km to the west Hanoi capital (Fig. 1). It covers an area of 1533.8 km^2, between latitudes 21^{o}11’51” N and 21^{o}37’52” N, and between longitudes 103^{o}19’20” E and 103^{o}59’50” E. In the study area, slope varies from 0 to 75.9^o with the mean of 24.2^o. The altitude of the district ranges from 140 m to 1839 m above sea level, with the mean of 874.8 m and the standard deviation of 296.4 m.

According to statistical analyses in 2015 (Tue, 2015), the agriculture land of the district accounts for 73.5% of the total land, in which the protective forest land is 54.2 % (61231.2 ha), the special forest land is 4.2% (4726.1 ha), the productive forest land is 2.9% (3312.8 ha), and the perennial crop land is 4.3% (4811.6 ha).

The district is located in a tropical monsoon climate with two separate seasons, dry and rainy. The dry season extends from October to March, whereas the rainy season lasts from April to September. The lowest average temperature is 14^{o}C whereas the highest average temperature is 26^{o}C, and the annual average temperature is around 21.4^{o}C (Tue, 2015). The average total annual rainfall in the district is 1372 mm, in which the rainfall during the rainy season accounts for up to 80% of the total rainfall.

The study area belongs to the most serious forest fire province in Vietnam, especially in the dry season due to effects of Foehn wind from Laos (Nguyen and Reiter, 2014) with low humidity and hot weather. According to Ministry of Agriculture and Rural Development of Vietnam, in 2016 only, a total of 2400 ha of forest destroyed by 181 large forest fires occurred in across 19 provinces in northern Vietnam, Son La province occupied the largest destroyed forest with around 969 ha (VNA, 2016).

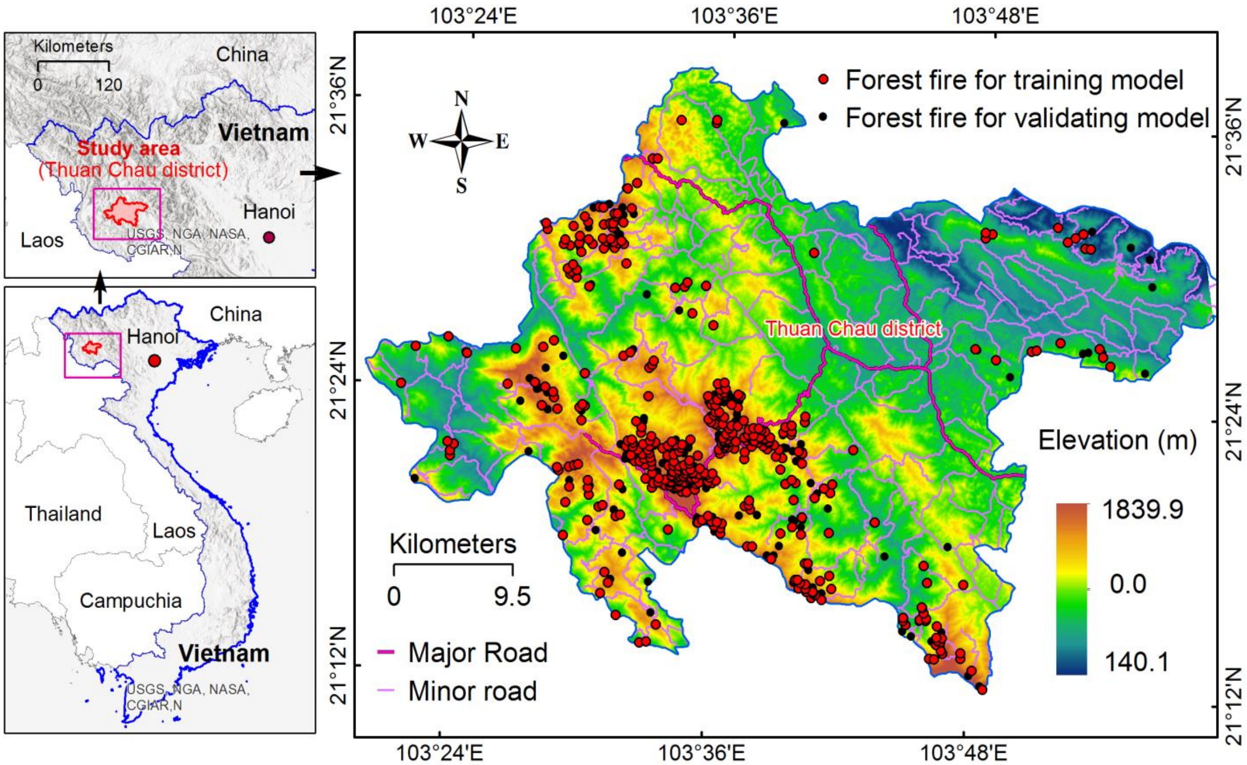


Fig. 1. Location of the study area and forest fire inventories.

2.2 Data used

2.2 1. Historic forest fire

In this study, a total of 564 fire locations that occurred in 2016 at the Thuan Chau district were collected to construct a forest fire inventory map for modeling. These fire locations that occurred in 2016 are from the national forest fire database of Vietnam (available at http://www.kiemlam.org.vn/firewatchvn/) and were produced by Ministry of Agriculture and Rural Development of Vietnam. It is emphasis that prolonged dry seasons in 2016 has significantly increased forest fire activities in the study area.

Temporal analysis of the historic fire data occurred in 2016 in this study area (Fig. 2) shows that fires occurred mainly in April and March that account for 85.6% of the total forest fires. They were followed by February (4.4%), November (3.7%), December (2.7%), and May (2.5%). In contrast, no fire has been recognized in July and September and very few fires occurred on January, October, June, and August. It is noted that in 2016, Vietnam was suffered the worst drought occurred in 90 years due to the El Niño weather event and the peak of the El Niño drought was from February to May, whereas from June to October had moderate rainfall intensity due to effect of La Niña activities (UNCT, 2016).

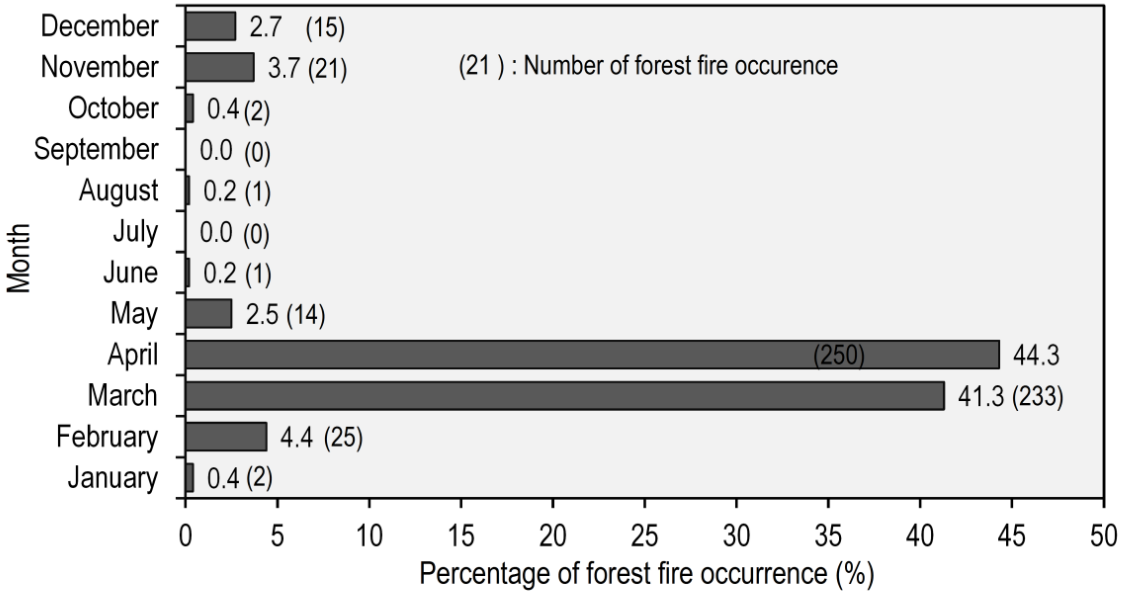


Fig. 2. Temporal statistics of the forest fires occurred in 2016 in this study area.

2.2 2. Forest fire influencing variables

Identification of fire variables for forest fire danger modeling is important task, and for finer spatial scales (less than 5000 ha), topography, biomass fuels, and weather condition have been determined to be the most important factors influencing intensity and severity of forest fire (Birch et al., 2015). In addition, human related factors, which have been responsible for igniting (Mann et al., 2016), should be considered. Therefore, in this research, a total of ten forest fire variables were selected and the detailed descriptions of these classes/categories and original sources are shown in Table 1.

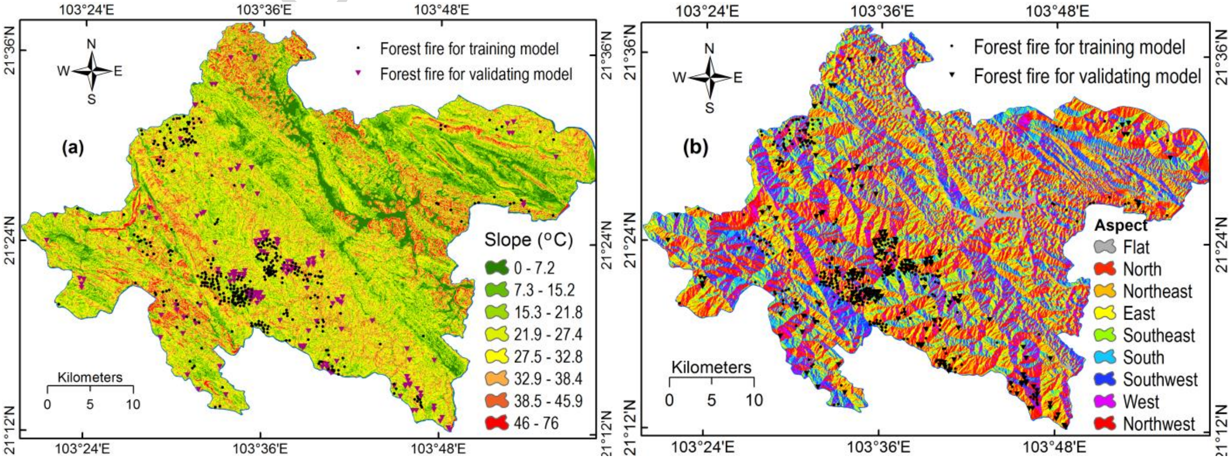
Table 1. Forest fire variable.

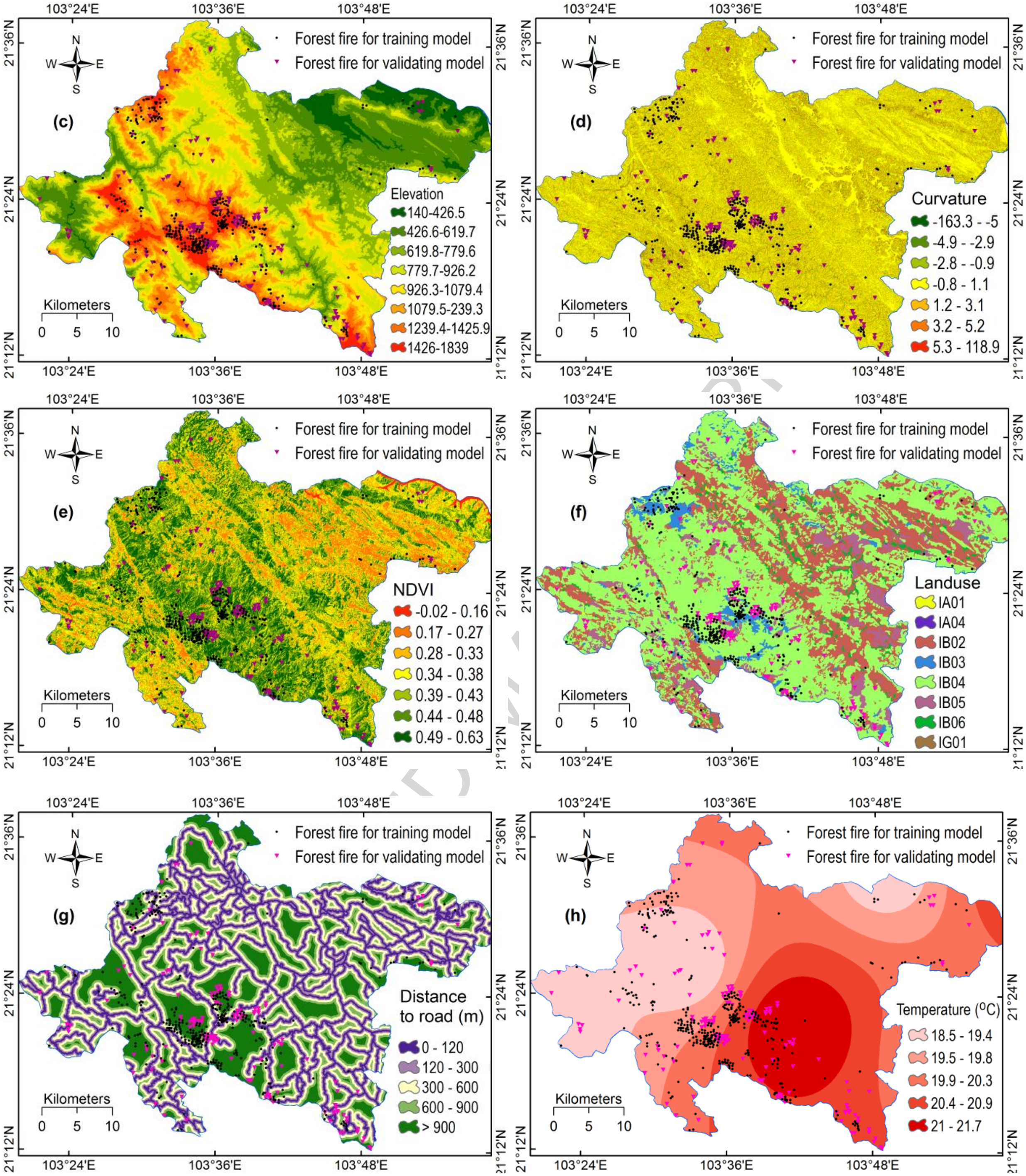
Forest fire variable	**Coding**	**Class/category**	**Original sources**
Slope (^{o})	SLO	(1) 0–7.2; (2) 7.3–15.2; (3) 15.3–21.8; (4) 21.9–27.4; (5) 27.5–32.8; (6) 32.9–38.4; (7) 38.5–45.9; (8) 46–76	Topographic maps 1:50,000 scale (from Ministry of Natural Resources and Environment of Viet Nam-MONRE)
Aspect	ASP	Flat; North; Northeast; East; Southeast; South; Southwest; West; Northwest	
Elevation (m)	ELEV	(1) 140 – 426.5; (2) 426.6 – 619.7; (3) 619.8 – 779.6; (4) 779.9 – 926.2; (5) 926.3 – 1079.4; (6) 1079.5 – 1239.3; (7) 1239.4 – 1425.9; (8) 1426 – 1839	
Curvature	CUR	(1) -163.3 – -5; (2) -4.9 – -2.9; (3) -2.8 – -0.9; (4) -0.8–1.1; (5) 1.2–3.1; (6) 3.2–5.2; (7) 5.3–118.9	

Distance to road (m)	DISR	(1) 0–120; (2) 120–300; (3) 300–600; (4) 600–900; (5) > 900	
NDVI	NDVI	(1) -0.02– 0.16; (2) 0.17–0.27; (3) 0.28–0.33; (4) 0.34–0.38; (5) 0.39–0.43; (6) (7) 0.44-0.48; (8) 0.49-0.63	Landsat-8 OLI 30 m, http://earthexplorer.usgs.gov
Land use	LAND	IA01; IA04; IB02; IB03; IB04;IB05;IB06; IG01	From the authority of the Thuan Chau district
Temperature (^{o})	TEMP	(1) 18.5–19.4; (2) 19.5–19.8; (3)19.9–20.3; (4) 20.4–20.9; (5) 21–21.7	Weather data (From Ministry of Natural Resources and Environment)
Humidity (%)	HUMI	(1) 62.2–64.9; (2) 65–67.3; (3) 67.4–69.5; (4) 69.6–71.2; (5) 71.3–72.9; (6) 73–74.9; (6) 75–78.2	
Rainfall (mm)	RAIN	(1) 162.7–183.3; (2) 183.4–200; (3) 201–216.7; (4) 216.8–233.3; (5) 233.4–256.7; (6) 256.8–285.3; (7) 285.4 – 332.7	

A Digital Elevation Model (DEM) with 30 m pixel size for the study area was generated using topographic maps 1:50,000 scale. These are official national maps that were constructed and provided by MONRE (Ministry of Natural Resources and Environment of Vietnam). Should be noted that the interval terrain contour of these maps are 5 m for areas with slopes less than 2°, 10 m for slope areas from 2° to 15°, and 20 m for the other areas (Tien Bui et al., 2017a). Using the DEM, we extract four variables: slope (SLO), aspect (ASP), elevation (ELEV), and curvature (CUR). SLO and ASP were considered for fire modeling because SLO controls the rate of fire spread (Cruz and Alexander, 2017). The highest rate often occurs on steeper slopes, and thus, fire progression is faster on areas with upper slopes than that of lower-slope (Viegas and Pita, 2004; Viegas and Simeoni, 2011). Whereas ASP influences fuel moisture that relate to fire behaviors (Nyman et al., 2015).

ELEV was considered because it influences solar radiation, temperature, and evapotranspiration of the terrain that are indirectly related to forest fires (Camp et al., 1997). Regarding CUR, this variable has proven important effecting the forest fire behavior, in which, the rate of fires spread at lower curvature area is higher than those at higher curvatures (Hilton et al., 2017).




(c)
Forest fire for training model
Forest fire for validating model
Elevation
140-426.5
426.6-619.7
619.8-779.6
779.7-926.2
926.3-1079.4
1079.5-239.3
1239.4-1425.9
1426-1839
(d)
Curvature
-163.3 - -5
-4.9 - -2.9
-2.8 - -0.9
-0.8 - 1.1
1.2 - 3.1
3.2 - 5.2
5.3 - 118.9
(e)
NDVI
-0.02 - 0.16
0.17 - 0.27
0.28 - 0.33
0.34 - 0.38
0.39 - 0.43
0.44 - 0.48
0.49 - 0.63
(f)
Landuse
IA01
IA04
IB02
IB03
IB04
IB05
IB06
IG01
(g)
Distance to road (m)
0 - 120
120 - 300
300 - 600
600 - 900
> 900
(h)
Temperature (°C)
18.5 - 19.4
19.5 - 19.8
19.9 - 20.3
20.4 - 20.9
21 - 21.7
N
S
W
E
Kilometers
0
5
10
103°24'E
103°36'E
103°48'E
21°36'N
21°24'N
21°12'N

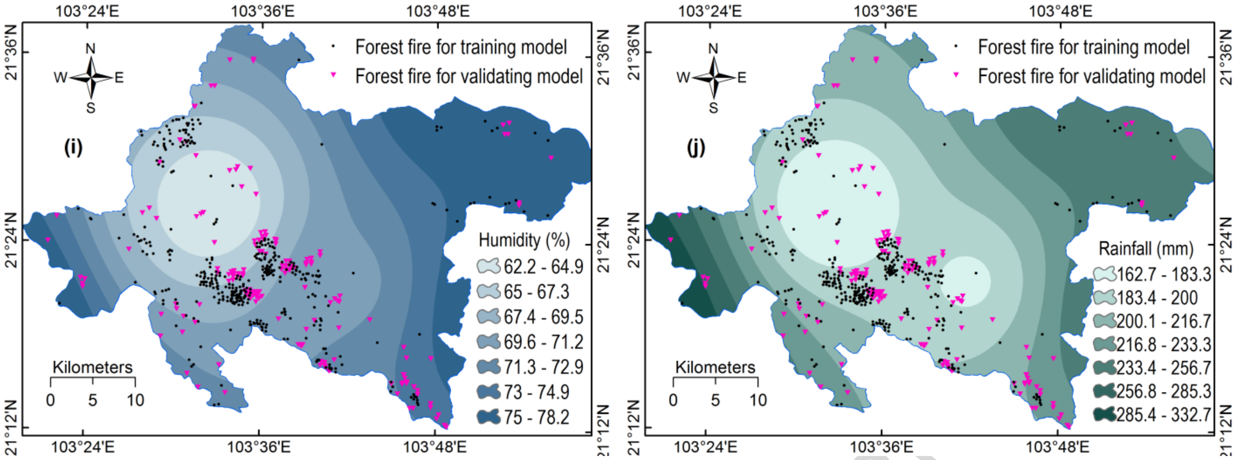


Fig. 3. **(a)** SLO; **(b)** ASP; **(c)** ELEV; and **(d)** CUR **(e)** NDVI map; **(f)** land use map; **(g)** Distance to road map; **(h)** Temperature**;** **(i)** Humidity; **(j)** Rainfall map.

For this research, the SLOP map (Fig. 3a) and the ELEV map (Fig. 3c) were built consisting 8 categories, meanwhile, the curvature map with 7 categories (Fig. 3d) was employed. We determined these categories through analyzing the pixel based-histogram of these maps with the assist of the natural-breaks technique (North, 2009) in ArcGIS 10.4 software. More specifically, we obtained the breakpoint for these categories using the Jenks method (Jenks, 1977). In which, these categories must satisfied two conditions, their inter-variance is maximized and the intra-variance is minimized. For the ASP map, nine categories were used (Fig. 3b).

Behavior of forest fire is influenced by fuels, therefore, NDVI that is an indicator of vegetation status was selected (Bajocco et al., 2015). In this research, NDVI was computed using Landsat-8 OLI (Operational Land Imagery) 30 m resolution that was derived from http://earthexplorer.usgs.gov using Eq.1, and then, the NDVI map was compiled with seven classes (Fig. 3e). These classes were determined using the natural breaks technique that indicated above. The NDVI is calculated according to the following equation (Reed et al., 1994) :

$$NDVI = (NIR - RED)/(NIR + RED) \qquad (1)$$

where NIR and RED indicate the surface reflectance values from the near-infrared band and the red band, respectively.

In addition, anthropogenic activities have been reported to be important sources for forest fire (Huesca et al., 2009; Nepstad et al., 2008). Hence, in this study, land use (LAND) and distance to road (DISTR) were considered as sources of ignition (Rolstad et al., 2017; Woo et al., 2017). Accordingly, the land use map at scale of 1:50,000 (Fig. 3f) were prepared with nine classes. This map has been established based on the outcome of the national land use project map in 2015 and was provided by the authority of the Thuan Chau district. For the DISR map, this map was created by buffering the road network extracting from the national topographic map 1:50,000 scale. Accordingly, there are five categories within these maps as shown in Fig. 3g.

Weather patterns such as temperature, relative humidity, and wind are considered as principal factors that strongly affects forest fire behavior, in which, forest fire is more likely to occur under weather conditions of hot, windy, and dry (Chang et al., 2013). For this study, the weather data in 2016 that available at Ministry of Natural Resources and Environment (Vietnam) were used including average maximum monthly climatic related data: temperature (TEMP), humidity

(HUMI), and total sum of rainfall (RAIN). Accordingly, the TEMP map (Fig. 3i) was created with five classes, whereas the HUMI map (Fig. 3k) was compiled with eight classes, and the RAIN map was generated with seven classes. These classes were determined based on the natural breaks technique that was described above.

3. Background of the method used

3.1 Support vector machine classifier

Support vector machine classifier (SVMC) is one of the most successfully machine learning techniques for classification and regression (Smola and Vapnik, 1997; Vapnik, 1998). SVMC has proven as an efficient for environmental modeling such as in landslide (Pham et al., 2016b; Tien Bui et al., 2016d) and forest biomass (Wu et al., 2017). However, it has been rarely explored for forest fire (Tien Bui et al., 2016a). The main advantage of SVMC is that it could produce high accuracy model using small sample set (Mountrakis et al., 2011).

Consider a training dataset $\mathrm{TD}(x_i, y_i)_{i=1}^{N}$ with $x_i \in R^n$ is the input vector, N is the number samples of TD, *n* is the dimension of TD, $y_i \in \{-1, +1\}$ denotes a class label. In this research context, the input vector consists of the ten forest fire influencing variables, whereas the label is the forest fire and the non-forest fire classes. The aim of the SVM mode is to find the best decision function (Eq.2) that is capable to separate the samples in the two classes. In this study, the probability of data samples belonging to the forest fire is employed as forest fire danger indices.

$$f(x) = \mathrm{sign}\left[\sum y_i \alpha_i K(x_i, x_j) + b\right] \tag{2}$$

where b is the offset value; α_i is the Lagrange multiplier; and $K(x_i, x_j)$ is kernel function.

It is noted that the performance of the SVMC model is dependent on the kernel function used. In this research, the Radial Basis Function (RBF) kernel was used because the SVMC model using RBF has outperformed other functions in term of prediction capability in various works (Hoang et al., 2016; Hoang and Tien Bui, 2016). Since the performance of the RBF-SVMC model is controlled by the two parameters of the regularization (C) and the kernel width (γ), these parameters must be selected properly (Hoang and Tien Bui, 2016). In this work, the commonly used grid search method (Hsu et al., 2003) was utilized because of its fast computation and acceptable performance (Kavzoglu and Colkesen, 2009).

3.2 Random forests

Random Forests (RF) classifier is a popular machine learning technique proposed by Breiman (2001) that have been widely used for classification, regression, and evaluation of the relative importance of input factors (Yu et al., 2017). Although RF has received huge attentions in other fields i.e. remote sensing due to the very high classification accuracy and computation speed, and robustness outliers (Belgiu and Drăguţ, 2016), exploration of RF for forest fire modeling has been rarely carried out. Basically, RF is an ensemble learning approach in which a set of decision tree classifiers is established to make a prediction. Accordingly, various sub-datasets are randomly generated through replacement from the training dataset. Subsequently, each sub-dataset is used to construct a decision tree using the Classification And Regression Tree *(*CART) algorithm (Breiman et al., 1984).

It is noted that the performance of RF is influenced by both the number of decision tree (N-tree) in the forest and the number of input factor (N-fact) that is random selected. Therefore, these two parameters must be set appropriately according to different data sets. Since N-tree is less sensitive to the classification accuracy, 500 trees is selected study to ensure the diversity of the RF model as suggested in Lawrence et al. (2006) and Stevens et al. (2015). For the case of N-fact, this parameter is selected equal the total number of input factors (ten) as suggested in Ghosh et al. (2014).

3.3 Multilayer perceptron neural network

Artificial neural network (ANN) is a mathematical model that is designed as a pattern of interconnection nodes. ANN has been successfully used for modeling of complex real world problems such as total suspended matter (Teodoro et al., 2007), biomass estimation (Pham et al., 2017a), landslide modeling (Tien Bui et al., 2016c), earthquake assessment (Asencio-Cortés et al., 2017), and energy prediction (Ascione et al., 2017). Therefore, ANN was selected for this research. Since the problem of forest fire susceptibility prediction can be modeled as a two-label classification task, a multilayer perceptron neural network (MLP-Net) structure is selected for the current study. It is because MLP-Net has been used in previous works of spatial modeling of forest fire risk (Cheng and Wang, 2008; Satir et al., 2015; Vasconcelos et al., 2001).

Structurally, MLP-Net comprises three layers: input, hidden and output, in which, an activation function is used to connect the input and hidden layers, whereas a linear function is used for the hidden and the output layers (Haykin, 1998). Performance of the MLP-Net model is influenced by connection weights of the three layers and these weights are updated and adjusted to fit the training dataset. In this research, the back-propagation algorithm (Hirose et al., 1991) was used since the algorithm has been proven to be efficient in various complex real world problems. Accordingly, the activation function of the logistic sigmoid was selected, whereas other parameters i.e. momentum of 0.2, learning rate of 0.3, and training iteration of 500 (Tien Bui et al., 2016b) were used.

4. Methodological flow used for this study

This section provides descriptions of the methodological flow (Fig. 4) used in this study for forest fire danger modeling. In this work, ArcGIS 10.4 has been employed for the data processing, coding, and visualizing purposes. The modeling process was carried out by Weka 3.9. In addition, a Python application was developed to connect the model results to an open GIS environment.

4.1 Establishment of GIS database, the training dataset, and the validation dataset

As mentioned earlier, the historical forest fires and the ten variables have been collected from various sources. Therefore, a geographic processing has been employed to construct a GIS database for this research. Accordingly, an ESRI file geodatabase format was established with the help of the ArcGIS 10.4. Subsequently, a coding process was performed by ArcGIS Model Builder tool to convert all maps to a raster format with resolution of 30 m. In addition, all categories of these maps were transferred to numeric format by a method suggested in Tien Bui et al. (2012b). Finally, the data normalization process was conducted for the ten forest fire influencing factors to hedge the potential bias caused by unbalanced magnitudes of the factors.

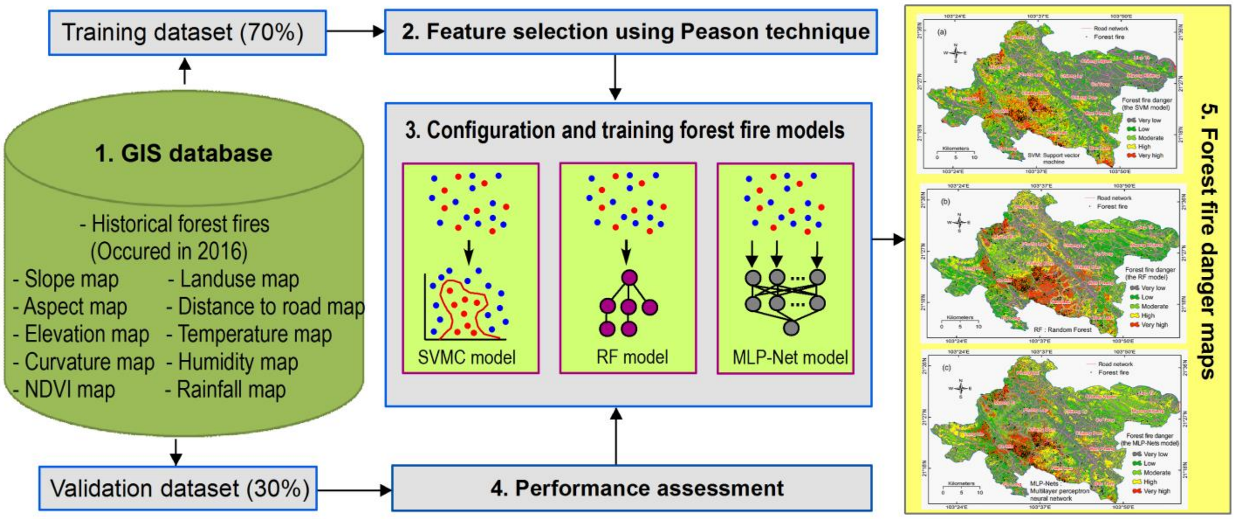


Fig. 4. Workflow of the methodology employed in this study.

For constructing forest file danger models, 564 forest fire locations were randomly split into two parts. The first part that consists of 395 fire locations (70%) was used for training the three machine learning models whereas the second part (169 fire locations, 30%) was reserved for assessing the prediction capability of the models. Since the forest fire danger modeling is considered as an on-off classification where samples were classified to either the forest fire class or the non-forest fire class. Therefore, the same amount of non-forest fire data points were randomly sampled from intact areas. The identification of intact areas were identified by the use of the estimated NDVI of the study area (i.e. bare lands and water areas with NDVI < 0.1) (Tien Bui et al., 2017b). At last, all the values of both the forest fires and the non-forest fire samples are extracted to establish the training and the validation datasets for the study area.

4.2 Feature selection

In machine learning applications, feature selection should be carried out because it may not only improve the predictive performances of classification models, but also reduce their computation costs and provide a better understanding of the input data (Chandrashekar and Sahin, 2014; Martínez-Álvarez et al., 2013). For forest fire danger modeling, although machine learning models do not require the input factors to be normally distributed, the prediction capability of these models may be reduced if noise features are included (Tien Bui et al., 2016a). Therefore, feature selection should be employed with the aim at identifying the most useful input variables from the original feature set. Thereby, the predictive ability of each feature is initially quantified in the connection with the output. Features with non-predictive ability should be eliminated to reduce bias and enhance the model performance.

In this research, the Pearson technique that has been proven to be efficient in forest fire modeling (Tien Bui et al., 2017b) was employed for the task of feature selection and quantifying the predictive ability of the ten forest fire variables. Accordingly, a merit value of a factor was determined by measuring the Pearson correlation (Guyon and Elisseeff, 2003) between that factor and the output. This merit factor is computed in Eq.3 as follows:

$$COR_i = cov(x_i, y) / \sqrt{var(x_i) * var(y)} \tag{3}$$

where COR_i is the Pearson correlation value of the forest fire influencing variable *i*; x_i is the forest fire influencing variable *i; Y* denotes the class label; $cov(.)$ and $var(.)$ represent the covariance and the variance, respectively.

4.3 Configuration and training forest fire models

For the SVMC model, the grid search was employed to search for the most desired values of C and γ with the employment of the training dataset. The grid space for C was from 2^{-5} to 2^{10} and for γ was from 2^{10} to 2^{-4} (Tien Bui et al., 2012a). Experimental result points out that C = 44 and γ = 0.075 are the most appropriate values. For the case of the RF model, N-tree = 500 was used to ensure the diversity of the RF model (Gislason et al., 2006) and all forest fire valuables (N-fact = 10) was used to establish the model. Regarding MLP-Net, performance of the MLP-Net model is strongly influenced by the number of hidden neurons; therefore, in this research, an empirical test as suggested in (Tien Bui et al., 2016c) was employed by testing numbers of hidden neurons versus MAE (Mean Absolute Error) on both the training and the validation datasets. Finally, the MLP-Net model with 4 neurons in the hidden layer found to be the best neural network structure for the data at hand. Using the training dataset and the above configurations, the SVMC, the RF, and the MLP-Net models were finally trained and obtained.

4.4 Performance assessment

As mentioned earlier, forest fire danger modeling is considered as an on-off classification that deals with the categorization of pixels of the study areas into one of the two classes: the forest fire and the non-forest fire. Accordingly, the four outcomes of the forest fire models with respect to the two classes are defined as true positive (TP), true negative (TN), false positive (FP), and false negative (FN) (Zweig and Campbell, 1993). These metrics are combined to form a confusion matrix that has been widely used to derive other evaluation metrics i.e. classification accuracy (ACC), sensitivity (SENS), specificity (SPEC), positive predictive ability (PPA), and negative predictive ability (NPA) (Witten et al., 2011). These metrics were used to assess the performance of the forest fire models in this research. The closer to 100 the evaluation metrics (ACC, SENS, SPEC, PPA, and NPA) are, the better the forest fire model is.

In addition, the Area under the ROC (Receiver Operating Characteristic) Curve (AUC) that summarizes the aforementioned metrics is employed to measure the global performance of the three forest fire susceptibility prediction models (Fernandes et al., 2004; Šturm and Podobnikar, 2017). Accordingly, a model is interpreted as follows: (i) excellent with AUC of (0.9–1.0); good with AUC of (0.8–0.9), fair with AUC of (0.7–0.8), and poor with AUC <0.7 (Kantardzic, 2011; Zhang et al., 2017). In machine learning, comparison of models should take into account the cost of error; therefore, Kappa Statistics are recommended to use (Ozcift and Gulten, 2011). Accordingly, performances of the forest fire susceptibility prediction models are considered to be more realistic if their Kappa statistics values are closer to 1.

5. Results and discussion

5.1 Predictive ability assessment of forest fire variables

The result of the feature selection using the Pearson method for this study is reported in Table 2. It could be seen that the ten forest fire influencing variables have shown clearly their predictive abilities in forest fire susceptibility prediction. The highest predictive ability value is NDVI (0.510), followed by and the ELEV (0.431), HUMI (0.140), DISR (0.120), RAIN (0.113), LAND (0.050), TEMP (0.048), and ASP (0.035). In contrast to these variables, CUR (0.015) and SLO (0.010) have the lowest predictive abilities. This is because the distribution of the forest fire locations in the classes of both CUR and SLO was more even. Nevertheless, because all forest fire variables are associated with positive values (Table 2), all of them are included in the forest fire modeling process.

Table 2. Predictive values of the ten fire forest fire influencing variables for the Thuan Chau district.

No	Forest fire variable	Predictive values	
		Average value	Standard deviation
1	NDVI	0.510	± 0.006
2	ELEV	0.431	± 0.011
3	HUMI	0.140	± 0.010
4	DISR	0.120	± 0.016
5	RAIN	0.113	± 0.014
6	LAND	0.050	± 0.010
7	TEMP	0.048	± 0.011
8	ASP	0.035	± 0.009
9	CUR	0.015	± 0.007
10	SLO	0.010	± 0.007

5.2 Model training and analysis

Once the SVMC model, the RF model, and the MLP-Net model had been configured, the training process was carried out to establish the final models. The training result is shown in Table 3. It could be seen that the MLP-Net model has attained the highest PPA (84.0%, Table 3). This fact indicates that the true positive rate of the MLP-Net model is 84.0%, followed by the SVMC model (75.6%), and finally the RF model (70.2%). The highest NPA value is obtained from RF (85%) and MLP-Net (83.7%). This results point out that that the probability these two models correctly classify the samples to the non-forest fire class is relatively high. In terms of NPA, the SVMC model (79.1%) is inferior to the two aforementioned prediction methods.

The highest SENS (83.8%) is achieved by the MLP-Net model denoting that 83.8% of the forest fire locations were classified to the forest fire class correctly. It is followed by the RF model (82.4%) and the SVMC model (78.4%). For SPEC, the MLP-Net model has the highest value of 83.9% denoting that 83.9% of the non-forest fire locations are classified to the non-forest fire class correctly.

Overall, all the three models have high degree-of-fit with the training dataset. However, the MLP-Net model has the highest one (ACC = 83.8% and AUC = 0.904). The SVMC model and the RF model have almost equal performances in terms of ACC and AUC. However, the SVMC model is considered to be better than the RF model in term of PPA. Kappa statistics for the three models are from 0.547 (the SVMC model) to 0.679 (the MLP-Net model) indicating a satisfied result.

Table 3. Performance of the three machine learning models (SVMC, RF, and MLP-Net) for forest fire danger modeling in the training stage. TP: True positive; TN: True negative; FP: False positive; FN: False negative; SENS: Sensitivity; SPEC: Specificity; PPA: Positive predictive ability; NPA: Negative predictive ability; ACC: Classification accuracy: AUC: Area under the receiver operating characteristic curve.

No	Statistical index	SVMC	RF	MLP-Net
1	TP	297	276	330
2	TN	311	334	329
3	FP	96	117	63
4	FN	82	59	64
5	PPA (%)	75.6	70.2	84.0
6	NPA (%)	79.1	85.0	83.7
7	SENS (%)	78.4	82.4	83.8
8	SPEC (%)	76.4	74.1	83.9
9	ACC (%)	77.4	77.6	83.8
10	Kappa statistics	0.547	0.552	0.679
11	AUC	0.844	0.844	0.904

5.3 Validation of the trained forest fire danger models

In this research, the three fire danger models were assessed using the validation dataset. The result is shown in Table 4. It could be observed that the MLP-Net model has the highest PPA of 81.1%. Accordingly, the probability that the MLP-Net model correctly classifies new samples to the forest fire class is 81.1%. In terms of PPA, the SVMC model (78.7%) and the RF model (74.6%) are the second and the third best approaches. However, for the case of NPA, the RF model has the highest value of 87.6% indicating that the probability the RF model correctly classifies new samples to the non-forest fire class is 87.6%. The values of NPA obtained from the MLP-Net model (82.2%) and the SVMC model (81.7%) are significantly lower than that of produced by the MLP-Net model.

Regarding the SENS metric, the RF model has the highest value of 85.7%. This result implies that 85.7% of the the forest fire locations are correctly classified to the forest fire class. The performance of the RF method in terms of SENS is followed by the MLP-Net model (82.0%) and the SVMC model (81.1%). Considering the specificity, the MLP-Net model has the highest value of 81.3% denoting that 81.3% of the non-forest fire locations are classified to the non-forest fire class correctly. The values of specificity of the SVMC model (79.3%) and the RF model (77.5%) are inferior to that of the neural network approach. For ACC and Kappa statistics, the MLP-Net model (ACC=81.7%, Kappa statistics = 0.633) has the highest values, followed by the RF model (ACC= 81.1%, Kappa statistics = 0.621) and the SVMC model (ACC= 80.2%, Kappa statistics = 0.604). ROC curves and AUC values of the three models in Fig. 5 show that the MLP-Net model has the highest one (AUC = 0.894), followed by the RF model (AUC = 0.883) and the SVMC model (AUC = 0.867).

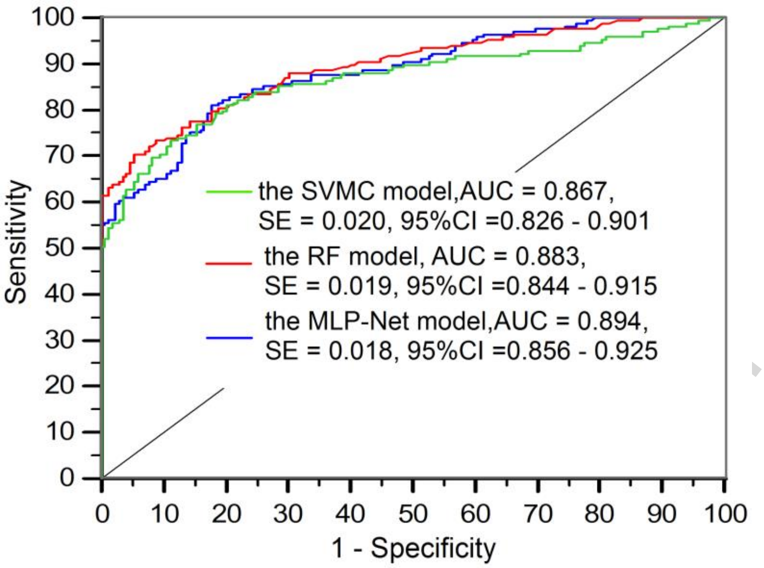


Fig. 5. ROC curve and AUC of the SVMC model, the RF model, and the MLP-Net model using the validation dataset (SE: Standard Error and CI: Confidence Interval).

To confirm the statistical significance of the three models' prediction performances, the non-parametric Wilcoxon signed-rank test (Wilcoxon and Wilcox, 1964), which is considered to be a powerful test when comparing the performance of machine learning models (Kate, 2016), was used for paired comparisons. This test uses the null-hypothesis (H_o) that there is no significant difference between the performances of the two forest fire danger prediction models at 95% CI (confidence intervals). Accordingly, the two-tailed probability (p-value) and the test statistics *(*z-value) are then computed on the prediction values produced from the forest fire prediction models. If p -value is less than 0.05 and z-value is outside of the critical values of [-1.96 and +1.96], H_o will be rejected and the performance of the models features a statistically significant difference (Goldfarb and King, 2016).

Table 4. Prediction capability of the three machine learning models (SVMC, RF, and MLP-Net) in the validation process.

No	Statistical index	SVMC	RF	MLP-Net
1	TP	133	126	137
2	TN	138	148	139
3	FP	36	43	32
4	FN	31	21	30
5	PPA (%)	78.7	74.6	81.1
6	NPA (%)	81.7	87.6	82.2
7	SENS (%)	81.1	85.7	82.0
8	SPEC (%)	79.3	77.5	81.3
9	ACC (%)	80.2	81.1	81.7
10	Kappa statistics	0.604	0.621	0.633

Result of the Wilcoxon signed-rank test is shown in Table 5. We see that the prediction performance of the MLP-Net model is significantly different from those of the SVMC model (P-value < 0.0001 and Z-value =4.625) and the RF model (P-value < 0.0001 and Z-value = 4.666). In contrast, the prediction performances of the SVMC model and the RF model have no significant difference (P-value = 0.936 and Z-value = 0.080). Overall, from above analysis, it could be concluded that the MLP-Net model is best suited for forest fire danger modeling in the study area.

Table 5. Result of pairwise comparison of the three forest fire danger models using Wilcoxon signed-rank test.

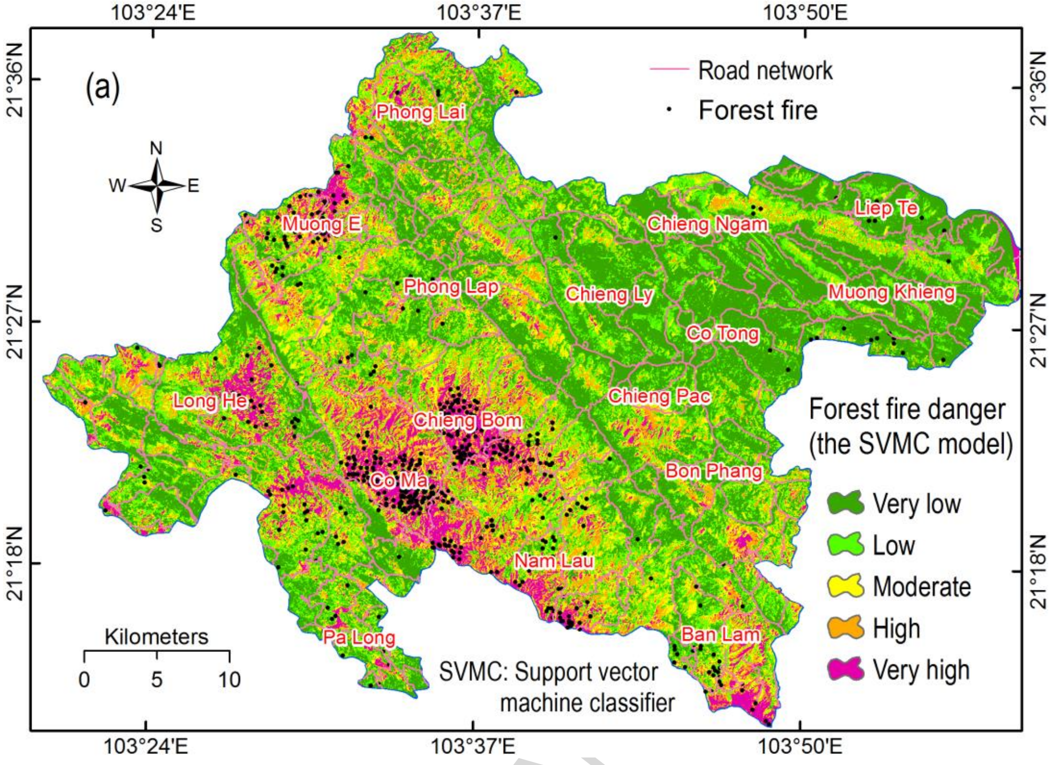
No	**Pairwise comparison**	Z-value	P-value	Significance
1	SVMC vs. RF	0.080	0.936	No
2	SVMC vs. MLP-Net	4.625	< 0.0001	Yes
3	RF vs. MLP-Net	4.666	< 0.0001	Yes

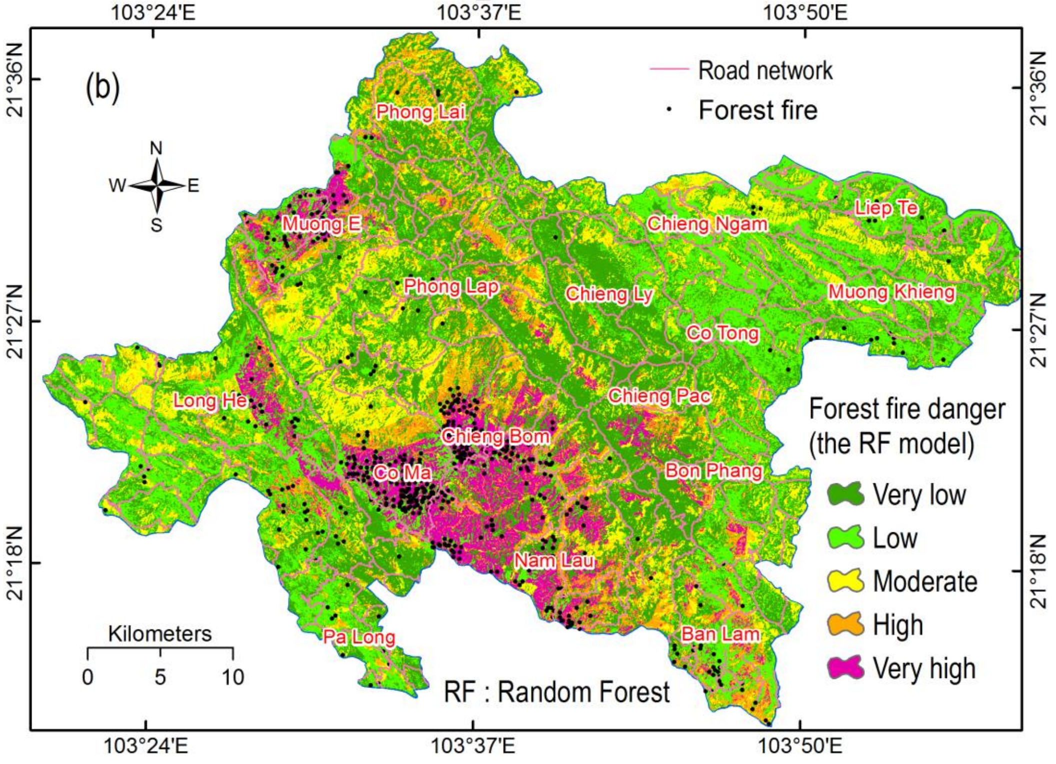
5.4 Compiling Forest fire danger map

Since the MLP-Net model delivered the best prediction performance for forest fire danger prediction in this study area, this machine learning model was further used to compute fire-danger indices for the Thuan Chau district. These indices were then converted to a grid format to be manageable in GIS environment and generated the forest fire danger maps.

Fig. 6 shows the three forest fire danger maps for the study area deriving from (i) the SVMC model (Fig. 6a), (ii) the RF model (Fig. 6b), and (iii) the MLP-Net model (Fig. 6c). These maps were reclassified into five danger classes with different colors: very low (40%), low (25%), moderate (15%), high (10%), and very high (10%). The thresholds of these classes were determined through overlaying the historical forest fires to each of the forest fire danger maps. For the detailed description of the overlaying technique, readers are guided to the previous work Tien Bui et al. (2017b).

Forest fire density analysis of the five forest fire danger classes was further carried out by calculating frequency ratios of the forest fire data for each class. Ideally, the density should be increased from the very low class to the very high class (Pradhan and Lee, 2010). The result (Table 6) showed that the very high class of the three models has the highest density value, whereas the very low class shows low density value. Overall, the MLP-Net model performed better than the other models because the fire density of the very high class has the highest value.


103°24'E
103°37'E
103°50'E
21°36'N
21°27'N
21°18'N
(a)
Road network
Forest fire
N
W
E
S
Phong Lai
Muong E
Chieng Ngam
Liep Te
Phong Lap
Chieng Ly
Muong Khieng
Co Tong
Long He
Chieng Pac
Chieng Bom
Co Ma
Bon Phang
Nam Lau
Pa Long
Ban Lam
Forest fire danger
(the SVMC model)
Very low
Low
Moderate
High
Very high
Kilometers
0
5
10
SVMC: Support vector
machine classifier


103°24'E
103°37'E
103°50'E
21°36'N
21°27'N
21°18'N
(b)
Road network
Forest fire
N
W
E
S
Phong Lai
Muong E
Chieng Ngam
Liep Te
Phong Lap
Chieng Ly
Muong Khieng
Co Tong
Long He
Chieng Pac
Chieng Bom
Co Ma
Bon Phang
Nam Lau
Pa Long
Ban Lam
Forest fire danger
(the RF model)
Very low
Low
Moderate
High
Very high
Kilometers
0
5
10
RF : Random Forest

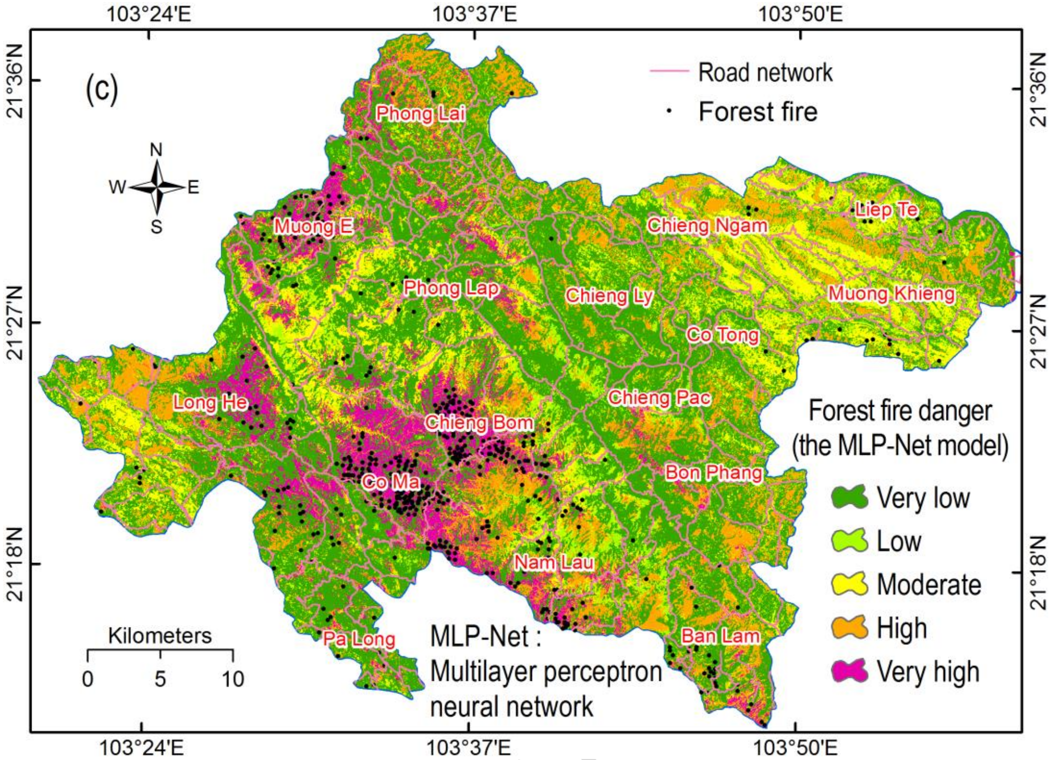


Fig. 6. Forest fire danger map derived from: (a) the SVMC model; (b) the RF model; (c) the MLP-Net model.

Table 6. Characteristics of the forest fire susceptibility classes obtained from the SVMC model, the RF model, and the MLP-Net model for the Thuan Chau district.

No.	Forest fire danger (%)	Verbal expression	Areas (km^2)	Forest fire density		
				SVMC	RF	MLP-Net
1	90 – 100	Very high	106.3	4.3	3.6	5.1
2	80 – 90	High	106.3	1.4	2.0	0.4
3	65 – 80	Moderate	159.5	1.1	0.1	0.3
4	40 – 65	Low	265.8	0.6	1.3	0.9
5	0 – 40	Very low	425.4	0.2	0.1	0.4

6. Concluding remarks

This research investigated and compared the three advanced machine learning algorithms for analyzing the spatial pattern of fire danger for the tropical forest of Thuan Chau (Vietnam). Though this work, a high quality forest fire danger map can be obtained that supports forest managers and local authority in forest management and fire suppression. For this task, a GIS database with 564 forest fire locations occurred in 2016 and ten forest fire influencing variables have been established and used for training and validating the forest fire susceptibility prediction models as well as producing the forest fire danger maps.

Accordingly, three supervised machine learning techniques, Support Vector Machine classifier (SVMC), Random Forests (RF), and Multilayer Perceptron Neural Network (MLP-Net) were systematically investigated. They are the most effective machine learning techniques that have been successfully used in various fields and proven to outperform conventional methods. However, to the best of our knowledge, no previous study on forest fire danger modeling has been carried out for the systematic comparison of the aforementioned three methods. In addition, various statically measures have been used for assessing and comparing the performance of the model. The high performance results of the three models indicate that the coding, processing, and modeling processes proposed in this study have been successfully carried out. Furthermore, small differences between the training and testing performance of these models demonstrate that the problem of overfitting has been diminished.

One of the most critical issues in modeling forest fire danger is the selection of forest fire variables. In this research, ten forest fire variables were evaluated the predictive ability using the Pearson technique. This ensures the detection of irrelevant forest fire affecting variables and guarantees that the model performance will not be deteriorated by irrelevant variables. As a result, the high performances of the constructed models indicate that the ten forest fire influencing variables have been properly selected.

Among the three models, the MLP-Net model showed not only the highest performance in overall but also a better balance between the predictions of the forest and non-forest locations. Based on experimental outcome, it is able to conclude that the MLP-Net is the best method for this study area. Accordingly, the model has been used to produce the forest fire danger map for the study area. Visual interpretation of these maps show that the two classes of high and very high have a clear separation with the others classes. High probabilities of forest fire are for Chieng Bom, Co Ma, and Muong E areas. These are reasonable results compared to the forest fire locations and our fieldwork checking. Therefore, these areas should be received special attentions in developing measures for preventing forest fires. In contrast, the other areas i.e. Liep Te, Muong Khiem, Co Tong, and Chieng Li areas have low probability of forest fire. In fact, these areas contain mainly sparse forest and newly planted forests where fuel loads are low; therefore, the probability of forest fire is low.

The main limitation of this research is that only the grid search was used to find the best values of C and γ for the SVMC model and the backpropagation algorithm was used to train the MLP-Net model. Therefore, the performance of these models for forest fire danger could be enhanced if optimization techniques i.e. Particle Swarm Optimization or ensemble framework i.e. Rotation Forest are considered. Despite these limitations, the forest fire danger models and maps produced in this study are valid and useful tools for the local authority and forest manager in forest management and fire suppression. Future directions of the current study may include the investigation of other novel machine learning models in the task of forest fire modeling and the integration of advanced optimization/ensemble methods to meliorate the model predictive accuracy.

Conflict of Interest

The authors declare that there is no conflict of interest.

Acknowledgement

The research was funded by the Project KHCN-TB.13C /13 -18 with title ″Research on modeling and system of sub-regional weather forecasting and warning of flood, forest fire and agricultural pests at district level in the North West Vietnam″ belongs to the National-funded Program for

Science and Technology for the period 2013-2018 titled ″Science and Technology for Sustainable Development in the North-West Region of Vietnam″.

Reference

Abdel-Rahman, E.M., Mutanga, O., Adam, E., Ismail, R., 2014. Detecting Sirex noctilio grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machines classifiers. ISPRS Journal of Photogrammetry and Remote Sensing 88, 48-59.
Arndt, N., Vacik, H., Koch, V., Arpaci, A., Gossow, H., 2013. Modeling human-caused forest fire ignition for assessing forest fire danger in Austria. iForest-Biogeosciences and Forestry 6, 315.
Arpaci, A., Malowerschnig, B., Sass, O., Vacik, H., 2014. Using multi variate data mining techniques for estimating fire susceptibility of Tyrolean forests. Applied Geography 53, 258-270.
Ascione, F., Bianco, N., De Stasio, C., Mauro, G.M., Vanoli, G.P., 2017. Artificial neural networks to predict energy performance and retrofit scenarios for any member of a building category: A novel approach. Energy 118, 999-1017.
Asencio-Cortés, G., Martínez-Álvarez, F., Troncoso, A., Morales-Esteban, A., 2017. Medium–large earthquake magnitude prediction in Tokyo with artificial neural networks. Neural Computing and Applications 28, 1043-1055.
Bajocco, S., Dragoz, E., Gitas, I., Smiraglia, D., Salvati, L., Ricotta, C., 2015. Mapping Forest Fuels through Vegetation Phenology: The Role of Coarse-Resolution Satellite Time-Series. PloS one 10, e0119811.
Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing 114, 24-31.
Bermudez, P.d.Z., Mendes, J., Pereira, J., Turkman, K., Vasconcelos, M., 2009. Spatial and temporal extremes of wildfire sizes in Portugal (1984–2004). International journal of wildland fire 18, 983-991.
Birch, D.S., Morgan, P., Kolden, C.A., Abatzoglou, J.T., Dillon, G.K., Hudak, A.T., Smith, A., 2015. Vegetation, topography and daily weather influenced burn severity in central Idaho and western Montana forests. Ecosphere 6, 1-23.
Breiman, L., 2001. Random forests. Machine learning 45, 5-32.
Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and regression trees. CRC press.
Brown, K.J., Giesecke, T., 2014. Holocene fire disturbance in the boreal forest of central Sweden. Boreas 43, 639-651.
Brown, K.J., Hebda, N.J., Conder, N., Golinski, K.G., Hawkes, B., Schoups, G., Hebda, R.J., 2017. Changing climate, vegetation, and fire disturbance in a sub-boreal pine-dominated forest, British Columbia, Canada. Canadian Journal of Forest Research 47, 615-627.
Camp, A., Oliver, C., Hessburg, P., Everett, R., 1997. Predicting late-successional fire refugia pre-dating European settlement in the Wenatchee Mountains. Forest Ecology and Management 95, 63-77.
Chandrashekar, G., Sahin, F., 2014. A survey on feature selection methods. Computers & Electrical Engineering 40, 16-28.
Chang, Y., Zhu, Z., Bu, R., Chen, H., Feng, Y., Li, Y., Hu, Y., Wang, Z., 2013. Predicting fire occurrence patterns with logistic regression in Heilongjiang Province, China. Landscape ecology 28, 1989-2004.
Cheng, T., Wang, J., 2008. Integrated Spatio-temporal Data Mining for Forest Fire Prediction. Transactions in GIS 12, 591-611.
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.
Conedera, M., Torriani, D., Neff, C., Ricotta, C., Bajocco, S., Pezzatti, G.B., 2011. Using Monte Carlo simulations to estimate relative fire ignition danger in a low-to-medium fire-prone region. Forest Ecology and Management 261, 2179-2187.
Cruz, M.G., Alexander, M.E., 2017. Modelling the rate of fire spread and uncertainty associated with the onset and propagation of crown fires in conifer forest stands. International Journal of Wildland Fire 26, 413-426.

Fernandes, A.M., Utkin, A.B., Lavrov, A.V., Vilar, R.M., 2004. Development of neural network committee machines for automatic forest fire detection using lidar. Pattern Recognition 37, 2039-2047.
Fox, D., Laaroussi, Y., Malkinson, L., Maselli, F., Andrieu, J., Bottai, L., Wittenberg, L., 2016. POSTFIRE: A model to map forest fire burn scar and estimate runoff and soil erosion risks. Remote Sensing Applications: Society and Environment 4, 83-91.
Ghosh, A., Fassnacht, F.E., Joshi, P., Koch, B., 2014. A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales. International Journal of Applied Earth Observation and Geoinformation 26, 49-63.
Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R., 2006. Random forests for land cover classification. Pattern Recognition Letters 27, 294-300.
Goldfarb, B., King, A.A., 2016. Scientific apophenia in strategic management research: Significance tests & mistaken inference. Strategic Management Journal 37, 167-176.
Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. Journal of machine learning research 3, 1157-1182.
Haykin, S., 1998. Neural Networks: A Comprehensive Foundation (2nd Edition). Prentice Hall, Upper Saddle River, NJ, USA.
Hilton, J., Miller, C., Sharples, J., Sullivan, A., 2017. Curvature effects in the dynamic propagation of wildfires. International Journal of Wildland Fire 25, 1238-1251.
Hirose, Y., Yamashita, K., Hijiya, S., 1991. Back-propagation algorithm which varies the number of hidden units. Neural Networks 4, 61-66.
Hoang, N.-D., Bui, D.T., Liao, K.-W., 2016. Groutability estimation of grouting processes with cement grouts using Differential Flower Pollination Optimized Support Vector Machine. Applied Soft Computing 45, 173-186.
Hoang, N.-D., Tien Bui, D., 2016. A Novel Relevance Vector Machine Classifier with Cuckoo Search Optimization for Spatial Prediction of Landslides. Journal of Computing in Civil Engineering 30, 04016001.
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.
Hsu, C.-W., Chang, C.-C., Lin, C.-J., 2003. A practical guide to support vector classification.
Huesca, M., Litago, J., Palacios-Orueta, A., Montes, F., Sebastian-Lopez, A., Escribano, P., 2009. Assessment of forest fire seasonality using MODIS fire potential: A time series approach. Agricultural and Forest Meteorology 149, 1946-1955.
Jenks, G., 1977. Optimal data classification for choropleth maps, Dept. of Geography, University of Kansas, USA.
Kantardzic, M., 2011. Data mining: concepts, models, methods, and algorithms. John Wiley & Sons, Hoboken, New Jersey.
Kate, R.J., 2016. Using dynamic time warping distances as features for improved time series classification. Data Mining and Knowledge Discovery 30, 283-312.
Kavzoglu, T., Colkesen, I., 2009. A kernel functions analysis for support vector machines for land cover classification. International Journal of Applied Earth Observation and Geoinformation 11, 352-359.
Koutsias, N., Martínez-Fernández, J., Allgöwer, B., 2010. Do factors causing wildfires vary in space? Evidence from geographically weighted regression. GIScience & Remote Sensing 47, 221-240.
Lautenberger, C., 2013. Wildland fire modeling with an Eulerian level set method and automated calibration. Fire Safety Journal 62, 289-298.
Lawrence, R.L., Wood, S.D., Sheley, R.L., 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). Remote Sensing of Environment 100, 356-362.
Le, T.H., Thanh Nguyen, T.N., Lasko, K., Ilavajhala, S., Vadrevu, K.P., Justice, C., 2014. Vegetation fires and air pollution in Vietnam. Environmental Pollution 195, 267-275.

Mann, M.L., Batllori, E., Moritz, M.A., Waller, E.K., Berck, P., Flint, A.L., Flint, L.E., Dolfi, E., 2016. Incorporating anthropogenic influences into fire probability models: effects of human activity and climate change on fire activity in California. PLoS One 11, e0153589.

Martínez-Álvarez, F., Reyes, J., Morales-Esteban, A., Rubio-Escudero, C., 2013. Determining the best set of seismicity indicators to predict earthquakes. Two case studies: Chile and the Iberian Peninsula. Knowledge-Based Systems 50, 198-210.

Mason, S.A., Hamlington, P.E., Hamlington, B.D., Matthew Jolly, W., Hoffman, C.M., 2017. Effects of Climate Oscillations on Wildland Fire Potential in the Continental United States. Geophysical Research Letters.

Massada, A.B., 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. International journal of wildland fire 22, 174-183.

Moritz, M.A., Parisien, M.-A., Batllori, E., Krawchuk, M.A., Van Dorn, J., Ganz, D.J., Hayhoe, K., 2012. Climate change and disruptions to global fire activity. Ecosphere 3, art49.

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing 66, 247-259.

Nepstad, D.C., Stickler, C.M., Soares-Filho, B., Merry, F., 2008. Interactions among Amazon land use, forests and climate: prospects for a near-term forest tipping point. Philosophical Transactions of the Royal Society of London B: Biological Sciences 363, 1737-1746.

Nguyen, A.-T., Reiter, S., 2014. A climate analysis tool for passive heating and cooling strategies in hot humid climate based on Typical Meteorological Year data sets. Energy and buildings 68, 756-763.

North, M.A., 2009. A method for implementing a statistically significant number of data classes in the Jenks algorithm, Fuzzy Systems and Knowledge Discovery, 2009. FSKD'09. Sixth International Conference on. IEEE, pp. 35-38.

Nyman, P., Metzen, D., Noske, P.J., Lane, P.N., Sheridan, G.J., 2015. Quantifying the effects of topographic aspect on water content and temperature in fine surface fuel. International Journal of Wildland Fire 24, 1129-1142.

Odion, D.C., Hanson, C.T., Arsenault, A., Baker, W.L., DellaSala, D.A., Hutto, R.L., Klenner, W., Moritz, M.A., Sherriff, R.L., Veblen, T.T., 2014. Examining historical and current mixed-severity fire regimes in ponderosa pine and mixed-conifer forests of western North America. PloS one 9, e87852.

Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A., Pereira, J.M.C., 2012. Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. Forest Ecology and Management 275, 117-129.

Ozcift, A., Gulten, A., 2011. Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms. Computer Methods and Programs in Biomedicine 104, 443-451.

Pettinari, M.L., Chuvieco, E., 2017. Fire Behavior Simulation from Global Fuel and Climatic Information. Forests 8, 179.

Pham, B.T., Pradhan, B., Tien Bui, D., Prakash, I., Dholakia, M.B., 2016a. A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India). Environmental Modelling and Software 84, 240–250.

Pham, B.T., Tien Bui, D., Dholakia, M.B., Prakash, I., Pham, H.V., 2016b. A Comparative Study of Least Square Support Vector Machines and Multiclass Alternating Decision Trees for Spatial Prediction of Rainfall-Induced Landslides in a Tropical Cyclones Area. Geotechnical and Geological Engineering, 1-18.

Pham, T.D., Yoshino, K., Bui, D.T., 2017a. Biomass estimation of Sonneratia caseolaris (l.) Engler at a coastal area of Hai Phong city (Vietnam) using ALOS-2 PALSAR imagery and GIS-based multi-layer perceptron neural networks. GIScience & Remote Sensing 54, 329-353.

Pham, T.N., Hoang, V.T., Pham, V.P., 2017b. Impact of forest fire on diversity of hymenopteran insects–a study at Copia species-used forest, Son La Province. Journal of Vietnamese Environment 8, 4-8.

Pimont, F., Parsons, R., Rigolot, E., de Coligny, F., Dupuy, J.-L., Dreyfus, P., Linn, R.R., 2016. Modeling fuels and fire effects in 3D: Model description and applications. Environmental Modelling & Software 80, 225-244.
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. Indic. 64, 72-84.
Pradhan, B., Lee, S., 2010. Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. Environmental Modelling & Software 25, 747-759.
Reed, B.C., Brown, J.F., VanderZee, D., Loveland, T.R., Merchant, J.W., Ohlen, D.O., 1994. Measuring phenological variability from satellite imagery. Journal of vegetation science 5, 703-714.
Rolstad, J., Blanck, Y.l., Storaunet, K.O., 2017. Fire history in a western Fennoscandian boreal forest as influenced by human land use and climate. Ecological Monographs 87, 219-245.
Sakr, G.E., Elhajj, I.H., Mitri, G., 2011. Efficient forest fire occurrence prediction for developing countries using two weather parameters. Engineering Applications of Artificial Intelligence 24, 888-894.
Satir, O., Berberoglu, S., Donmez, C., 2015. Mapping regional forest fire probability using artificial neural network model in a Mediterranean forest ecosystem. Geomatics, Natural Hazards and Risk, 1-14.
Sileshi, G.W., 2014. A critical review of forest biomass estimation models, common mistakes and corrective measures. Forest Ecology and Management 329, 237-254.
Smola, A., Vapnik, V., 1997. Support vector regression machines. Advances in neural information processing systems 9, 155-161.
Stevens, F.R., Gaughan, A.E., Linard, C., Tatem, A.J., 2015. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. PLoS One 10, e0107042.
Šturm, T., Podobnikar, T., 2017. A probability model for long-term forest fire occurrence in the Karst forest management area of Slovenia. International Journal of Wildland Fire 26, 399-412.
Sumarga, E., 2017. Spatial Indicators for Human Activities May Explain the 2015 Fire Hotspot Distribution in Central Kalimantan Indonesia. Tropical Conservation Science 10, 1940082917706168.
Teodoro, A., Amaral, A., 2017. Evaluation of forest fires in Portugal Mainland during 2016 summer considering different satellite datasets, Remote Sensing for Agriculture, Ecosystems, and Hydrology XIX. International Society for Optics and Photonics, p. 104211R.
Teodoro, A.C., Duarte, L., 2013. Forest fire risk maps: a GIS open source application–a case study in Norwest of Portugal. International Journal of Geographical Information Science 27, 699-720.
Teodoro, A.C., Veloso-Gomes, F., Goncalves, H., 2007. Retrieving TSM concentration from multispectral satellite data by multiple regression and artificial neural networks. IEEE Transactions on Geoscience and Remote Sensing 45, 1342-1350.
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. Agricultural and Forest Meteorology 233, 32-44.
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. Remote Sensing 8, 347.
Tien Bui, D., Pradhan, B., Lofman, O., Revhaug, I., 2012a. Landslide susceptibility assessment in Vietnam using Support vector machines, Decision tree and Naïve Bayes models. Mathematical Problems in Engineering 2012, 1-26.
Tien Bui, D., Pradhan, B., Lofman, O., Revhaug, I., Dick, O.B., 2012b. Landslide susceptibility assessment in the Hoa Binh province of Vietnam: A comparison of the Levenberg-Marquardt and Bayesian regularized neural networks. Geomorphology 171–172, 12–29.

Tien Bui, D., Pradhan, B., Nampak, H., Quang Bui, T., Tran, Q.-A., Nguyen, Q.P., 2016b. 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. Journal of Hydrology 540, 317-330.

Tien Bui, D., Tuan, T.A., Klempe, H., Pradhan, B., Revhaug, I., 2016c. 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.

Tien Bui, D., Tuan, T.A., Klempe, H., Pradhan, B., Revhaug, I., 2016d. 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.

Tue, D.T., 2015. Overview of the potential of the Thuan Chau district for the development of agricultural and forestry products in combination with processing facilities The People Committe of the Thuan Chau district, Thuan Chau.

UNCT, 2016. Viet Nam: Drought and Saltwater Intrusion Situation Report No. 6.

Vapnik, V., 1998. Statistical learning theory. Wiley New York.

Vasconcelos, M.P.d., Silva, S., Tome, M., Alvim, M., Pereira, J.C., 2001. Spatial prediction of fire ignition probabilities: comparing logistic regression and neural networks. Photogrammetric engineering and remote sensing 67, 73-81.

Verde, J., Zêzere, J., 2010. Assessment and validation of wildfire susceptibility and hazard in Portugal. Natural Hazards and Earth System Science 10, 485-497.

Viegas, D.X., Pita, L.P., 2004. Fire spread in canyons. International Journal of Wildland Fire 13, 253-274.

Viegas, D.X., Simeoni, A., 2011. Eruptive Behaviour of Forest Fires. Fire Technology 47, 303-320.

VNA, 2016. Fires ravage over 2,400 ha of forest in 2016. Vietnamnet, Hanoi, pp. http://english.vietnamnet.vn/fms/environment/166497/fires-ravage-over-166492-166400-ha-of-forest-in-162016.html.

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. Ecological Processes 6, 5.

Wenhua, L., 2004. Degradation and restoration of forest ecosystems in China. Forest Ecology and Management 201, 33-41.

Were, K., Tien Bui, D., Dick, Ø.B., Singh, B.R., 2015. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. Ecol. Indic. 52, 394-403.

Whitburn, S., Van Damme, M., Clarisse, L., Turquety, S., Clerbaux, C., Coheur, P.F., 2016. Doubling of annual ammonia emissions from the peat fires in Indonesia during the 2015 El Niño. Geophysical Research Letters 43.

Wilcoxon, F., Wilcox, R.A., 1964. Some rapid approximate statistical procedures. Lederle Laboratories.

Witten, I.H., Frank, E., Mark, A.H., 2011. Data Mining: Practical Machine Learning Tools and Techniques (Third Edition). Morgan Kaufmann, Burlington, USA.

Woo, H., Chung, W., Graham, J.M., Lee, B., 2017. Forest fire risk assessment using point process modelling of fire occurrence and Monte Carlo fire simulation. International Journal of Wildland Fire 26, 789-805.

Wotton, B., Martell, D., Logan, K., 2003. Climate change and people-caused forest fire occurrence in Ontario. Climatic Change 60, 275-295.

Wu, C., Tao, H., Zhai, M., Lin, Y., Wang, K., Deng, J., Shen, A., Gan, M., Li, J., Yang, H., 2017. Using nonparametric modeling approaches and remote sensing imagery to estimate ecological welfare forest biomass. Journal of Forestry Research.

Wu, Z., He, H.S., Yang, J., Liu, Z., Liang, Y., 2014. Relative effects of climatic and local factors on fire occurrence in boreal forest landscapes of northeastern China. Science of the Total Environment 493, 472-480.

Yu, P.-S., Yang, T.-C., Chen, S.-Y., Kuo, C.-M., Tseng, H.-W., 2017. Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting. Journal of Hydrology 552, 92-104.

Zhang, Y., Lim, S., Sharples, J.J., 2017. Wildfire occurrence patterns in ecoregions of New South Wales and Australian Capital Territory, Australia. Natural Hazards 87, 415-435.

Zweig, M.H., Campbell, G., 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical chemistry 39, 561-577.

Highlights

- Spatial pattern assessment of tropical forest fire using advanced machine learning, SVMC, RF, and MLP-Net.
- Predictive ability of forest fire variables was analyzed using Pearson correlation.
- MLP-Net is the best, providing > 85% prediction accuracy for future forest fire.