Ecological Informatics 43 (2018) 200–211

Contents lists available at ScienceDirect

Ecological Informatics

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

Wildfire spatial pattern analysis in the Zagros Mountains, Iran: A comparative study of decision tree based classifiers

Abolfazl Jaafari[a,*], Eric K. Zenner[b], Binh Thai Pham[c]

[a] Young Researchers and Elite Club, Karaj Branch, Islamic Azad University, Karaj, Iran
[b] Department of Ecosystem Science and Management, The Pennsylvania State University, Forest Resources Building, University Park, PA 16802, USA
[c] Department of Geotechnical Engineering, University of Transport Technology, 54 Trieu Khuc, Thanh Xuan, HaNoi, Viet Nam

ARTICLE INFO

Keywords:
Data mining
Decision tree
GIS
Feature selection
Spatial prediction

ABSTRACT

Knowledge of wildfire behavior is of key importance for planning and allocating resources to fire suppression efforts. In this study, we analyzed the spatial pattern of wildfires with five decision tree based classifiers, including alternating decision tree (ADT), classification and regression tree (CART), functional tree (FT), logistic model tree (LMT), and Naïve Bayes tree (NBT). The classifiers were trained using historical fire locations in the Zagros Mountains (Iran) from the years 2007–2014 and a set of fifteen explanatory variables (i.e., slope degree, aspect, altitude, plan curvature, topographic position index (TPI), topographic roughness index (TRI), topographic wetness index (TWI), mean annual temperature and rainfall, wind effect, soil type, land use, and proximity to settlements, roads, and rivers) that were first optimized with a twostep process using multicollinearity analysis and the Gain Ratio variable selection method. The classifiers were then validated using the Kappa index and several statistical index-based evaluators (i.e., accuracy, sensitivity, specificity, precision, and F-measure). The global performance of the classifiers was measured using the ROC-AUC method. In this comparative study, the ADT classifier demonstrated the highest performance both in terms of goodness-of-fit with the training dataset (accuracy = 99.8%, AUC = 0.991) and the capability to predict future wildfires (accuracy = 75.7%, AUC = 0.903). This study contributes to the suite of research that evaluates data mining methods for the prediction of natural hazards.

1. Introduction

Wildfires are one of the major natural hazards that result in the loss of forest resources and threaten the safety of human life and property (Flannigan et al., 2000; Guo et al., 2016). Rapid urbanization and land use changes, extensive conversion from natural vegetation to farmland, and direct human impact have resulted in ongoing climatic changes that have, in turn, increased the intensity and frequency of wildland fires across the globe at an alarming rate (Bedia et al., 2015; Tien Bui et al., 2017). The substantial impact of wildfires on terrestrial ecosystems, socio-economic conditions, and land use policies have encouraged governments and many researchers to invest great effort in the prevention and suppression of wildfires (Guo et al., 2016). Rational wildfire defense and suppression plans require first and foremost accurate estimates of the differential susceptibility of the land to wildfires in relation to the characteristics of the territory.

An accurate estimate of wildfire susceptibility provides an analytical framework in which the underlying patterns of fire ignitions and their causative factors in diverse and complex landscapes can be more fully understood (Jaafari et al., 2017; Jahdi et al., 2015). Understanding the susceptibility of a landscape can then be used to develop maps that show different relative levels of susceptibility, hazard, and risk zones.

In addition to the quality of the available data, an accurate estimate of wildfire susceptibility for a given landscape also depends to a large extent also on the employed modeling approach (Boubeta et al., 2015; Jaafari et al., 2017; Pourtaghi et al., 2016; Teodoro and Duarte, 2013; Tien Bui et al., 2017). Given the complex nature of wildland fires and the multitude of causative factors, the contribution of advanced analytical techniques, such as data mining tools, for adequately investigating spatial patterns and drivers of fire occurrence has increasingly been recognized. Data mining, as an interdisciplinary subfield of computer science, is the process of discovering patterns and extracting hidden knowledge from databases (Fayyad et al., 1996). In the domain of environmental research, the term 'data mining' has typically been substituted by other terms such as dynamic, empirical, agent-based, rule-based or machine learning models (Tayyebi et al., 2014).

The objective of machine learning models is to develop a set of rules or to provide an equation to estimate and map the variation in the

* Corresponding author.
E-mail address: a.jaafari@modares.ac.ir (A. Jaafari).

https://doi.org/10.1016/j.ecoinf.2017.12.006
Received 22 June 2017; Received in revised form 10 December 2017; Accepted 19 December 2017
Available online 19 December 2017
1574-9541/

susceptibility to wildfire occurrences across landscapes. Common models include logistic regression (LR), maximum entropy (ME), artificial neural networks (ANN), random forest (RF), support vector machine (SVM), and classification and regression tree (CART) (Catry et al., 2009; Oliveira et al., 2012; Pan et al., 2016; Pourtaghi et al., 2016; Tien Bui et al., 2016a, 2017). Despite recent advances in wildfire susceptibility modeling and publications of several comparative studies that sought to identify the most suitable approach for modeling susceptibility to wildfire occurrence, it is still unclear which modeling approach is best for identifying and delineating susceptible, wildfire-prone areas when the risk for the occurrence of wildfire is differentially distributed across the landscape. For instance, Oliveira et al. (2012) and Guo et al. (2016) found that RF had a higher predictive capability of wildfire occurrence than LR. Rodrigues and de la Riva (2014) found that RF was the most successful method for wildfire prediction than boosted regression tree (BRT), SVM, and, LR. Arpaci et al. (2014) showed that RF was slightly better than ME in wildfire prediction. In contrast, Pourtaghi et al. (2016) demonstrated that a generalized additive model was superior to the RF and BRT, and Tien Bui et al. (2016a, 2016b) showed that Kernel LR had a better predictive capability than SVM for wildfire modeling.

While large areas of terrestrial ecosystems of Iran have been burned by wildfires over the recent decade (Adab et al., 2013; Jaafari et al., 2017; Jahdi et al., 2015), relatively little research has been carried out to estimate the susceptibility of different ecosystems to wildfire occurrences. Further, our experience with data-mining modeling in Iran is limited to using the frequency ratio and index of entropy (Beygi et al., 2014; Jaafari and Mafi Gholami, 2017; Pourtaghi et al., 2015), Bayesian theory (Jaafari et al., 2017; Nami et al., 2017; Pourghasemi, 2016), Analytic hierarchy process (Beygi et al., 2014; Eskandari and Miesel, 2017) and LR (Mohammadi et al., 2014; Eskandari and Chuvieco, 2015; Pourghasemi, 2016). However, ongoing improvements in data mining methods and geographic information systems (GIS) have led to new techniques for modeling natural hazards such as landslides and floods that have typically outperformed the more conventional approaches (e.g., Pham et al., 2016), but have not yet been used in the context of wildfire studies. In this study, we seek to fill this gap in research by investigating the suitability and performance of five decision tree based classifiers, i.e., alternating decision tree (ADT), functional tree (FT), logistic model tree (LMT), Naïve Bayes tree (NBT), and CART, in the context of wildfire susceptibility modeling. The main advantage of these classifiers is that the prediction problem can be divided into several similar sub-problems with clear decision rules to which the same strategy can be applied to solve the entire prediction problem. With the exception of CART (e.g., Amatulli et al., 2006), this study is the first to apply the remaining four classifiers to model wildfire susceptibility. Further, we applied a two-step process that uses multi-collinearity analysis and the Gain Ratio method with a 10-fold cross-validation to select a smaller, optimal subset of predictor variables to minimize the potential bias inherent in over-fitted models. We used the receiver operating characteristics (ROC), the Kappa index, and several other statistical index-based measures for the assessment, validation, and comparison of the five decision tree based classifiers for predicting wildfire occurrence in a portion of the Zagros Mountains in south-western Iran.

2. Study area and data used

2.1. Study area

This study was conducted in the Chaharmahal-Bakhtiari Province, located on the eastern slope of the Zagros Mountains, Iran (Fig. 1). This province covers an area of about 16,532 km^2 and is characterized by several land cover types, i.e., forest (24%), rangeland (53%), farmland (21%), and orchard (0.5%) that exhibit a range of vegetation conditions (Jaafari et al., 2017). The terrain of this province is complex and steep, with slope gradients between 0° (flat) and > 80° and altitudes between 783 and 4178 m. The terrain strongly affects the local climate as temperature decreases and precipitation increases markedly with elevation. The mean annual temperatures vary between 5 °C in the central parts and 16 °C in the western parts of the province, with a provincial average of 10 °C (Meteorological Organization of Iran (MOI), 2014). The annual rainfall amounts range from 1400 mm in the northwestern to 250 mm in the eastern and southeastern portion, with an average of 560 mm. Rainfall mainly occurs during the winter and spring months and summers are typically dry. Winds are predominantly from the west and southwest, but easterly winds dominate during the summer months.

The estimated human population of this province in the year 2011 was 895,263 inhabitants (Statistical Centre of Iran (SCI), 2011). While much of the western lands remain undeveloped, the eastern lands have been experiencing increased development pressure due to their proximity to national roads and the Isfahan Province. The area has suffered from some of the most serious and frequent fire activity in recent years that also stands out in a historical context (Jaafari et al., 2017). Although the fire season typically extends from June until October, fire activity has a single modal seasonal distribution that peaks in July and August. Examples of wildfire events in the study area are shown in Fig. 2.

2.2. Data used

Preparing an inventory map of historical fire events is the first step in wildfire susceptibility modeling. This inventory map is typically constructed using different data sources, such as satellite image interpretation, aerial photography, fieldwork, and historical archives (Adab et al., 2015; Pourtaghi et al., 2016). In this study, the wildfire inventory map for the study area was obtained from the administrative office of natural resources of the Chaharmahal-Bakhtiari Province as well as from national reports that was enhanced by multiple field surveys and screening processes. The screening and cleaning processes resulted in the removal of records that contained illogical information (e.g., records with zero as the fire detection time or occurrences of crown fires with no associated wooded burn areas), records with insufficient or no information (blanks), and records that indicated small (< 0.3 ha) fires. After screening 268 fire records from the period of 2003 to 2014, 132 fire records from the period of 2007 to 2014 were retained for the modeling purpose (Fig. 1).

Fire ignitions are affected by a variety of interrelated factors (Adab et al., 2013, 2015; Catry et al., 2009; Chen et al., 2015; Jaafari and Mafi Gholami, 2017; Mhawej et al., 2015, 2017; Nami et al., 2017; Oliveira et al., 2012; Pourtaghi et al., 2016; Tien Bui et al., 2016a, 2016b, 2017). Based on the review of the literature, characteristics of fire locations, and available data, the following 15 factors were selected as explanatory variables for modeling the wildfire susceptibility in this study: slope degree, aspect, altitude, plan curvature, topographic position index (TPI), topographic roughness index (TRI), topographic wetness index (TWI), mean annual temperature and rainfall, wind effect, soil type, land use, and proximity to settlements, roads, and rivers. The significance of these variables in wildfire occurrence has been shown in Stambaugh and Guyette (2008), Adab et al. (2013, 2015) and Pourtaghi et al. (2016). In this study, slope degree, slope aspect, altitude, plan curvature, TPI, TRI, and TWI layers were created from a Digital Elevation Model (DEM) with 20 × 20 m grid size using ArcGIS and SAGA GIS. The mean annual temperature and rainfall maps were prepared using Kriging toolbox of the ArcMap software based on the data obtained from the Meteorological Organization of Iran for the period of 2003–2014. The wind effect variable was computed using data on wind speed and direction in the SAGA GIS workbench. The soil type map was obtained from National Cartographic Centre of Iran on a 1:100,000 scale. The land use map that contained eight land use classes (see below) was extracted from Landsat-7 images using the maximum

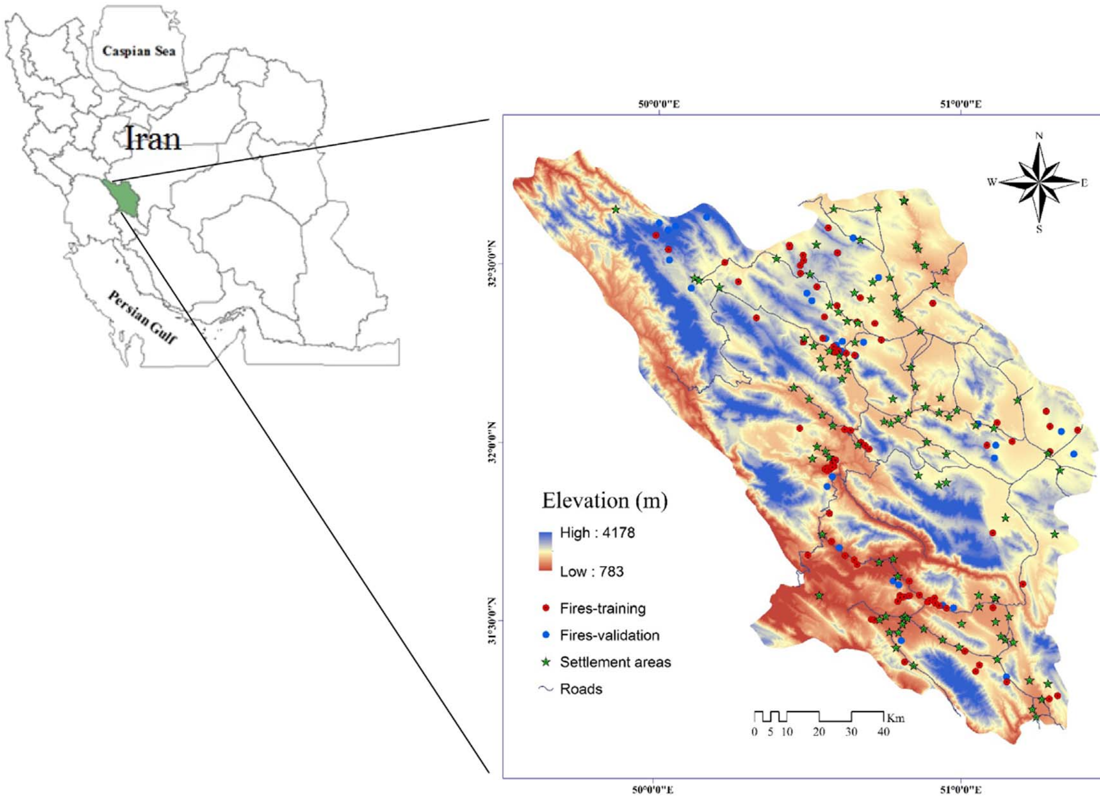

Fig. 1. Location of study area with fire event locations.

likelihood algorithm provided in the PCI Geomatica software at an accuracy of 91%. The proximity layers were computed using the spatial analyst tools of ArcGIS. All of these maps were constructed in raster format with a grid cell size of 50 × 50 m, which was small enough to capture the spatial characteristics of wildfire locations and large enough to reduce computing complexity.

Fig. 2. Photographs showing wildfire events in the Zagros Mountains (Photos by: Amirhosein Zolfaghari and Sajad Zolfaghari).

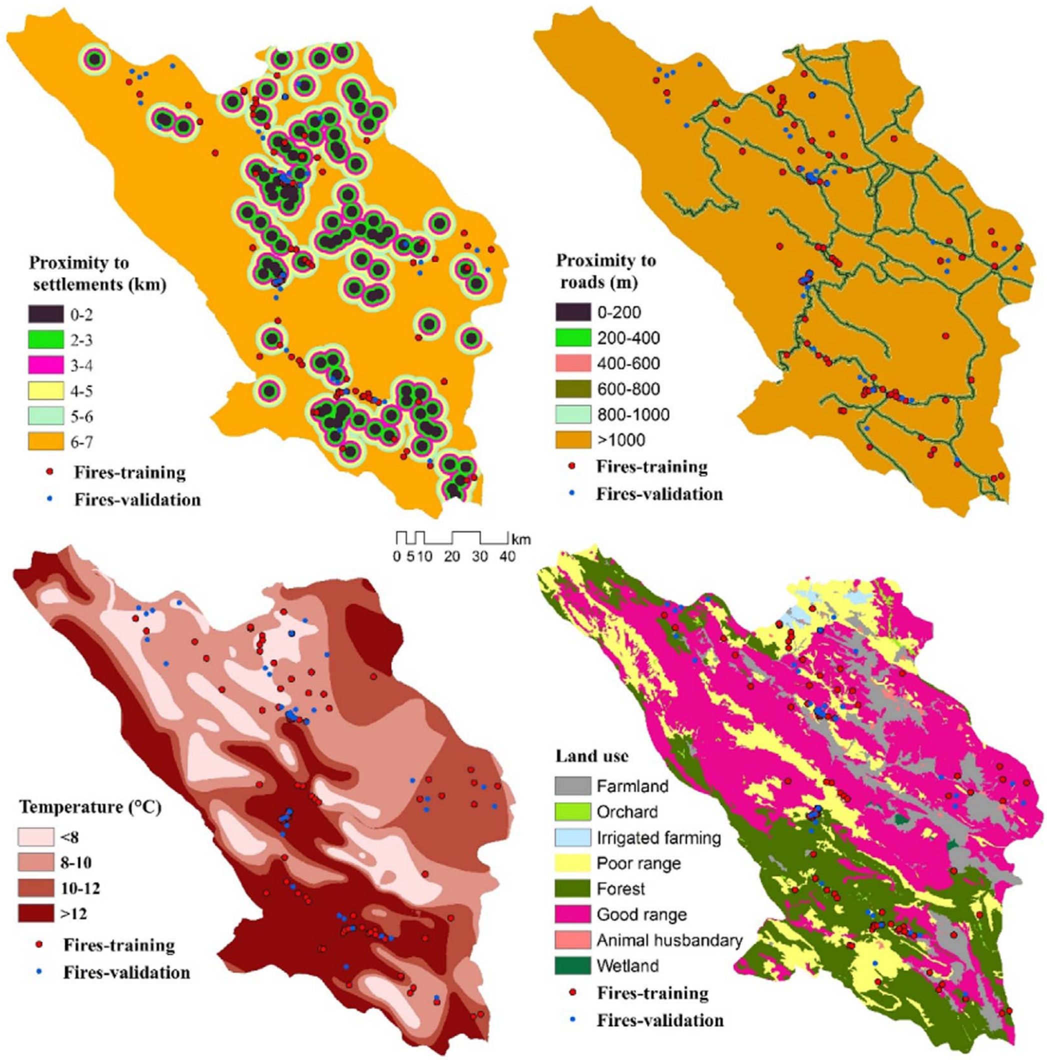

Fig. 3. Relation of the four most prominent predictor/explanatory variables for wildfire susceptibility in the Zagros Mountains.

In the final step, each of the 15 explanatory values observed on the grid cells was divided into categorical classes and assigned to the appropriate class in thematic maps. To do so, we first adopted classes as suggested in the literature (e.g., Nami et al., 2017; Pourghasemi, 2016; Pourtaghi et al., 2016), but found that we needed to adapt the classes to fit local conditions of soils types present, land use in the area, and actually observed distances of burned areas to roads and rivers. Thus, we ended up with six soil type classes (i.e., entisol, inceptisol, miscellaneous, badland, lake, pedocal) and eight land use classes (i.e., farmland, irrigated farmland, orchard, poor range, good range, animal husbandry, forests, and wetlands). Fig. 3 illustrates the spatial distribution of the classes of the most prominent predictor variables (see results) of wildfire susceptibility. The classes for each explanatory variable and the frequencies of occurrence in the study area are seen in Fig. 4.

3. Methodology

The workflow of the methodology used in this study (Fig. 5) is as follows:

3.1. Preparation of training and validation datasets

To build and then independently evaluate the wildfire susceptibility models, two different datasets are required for training and testing (Chen et al., 2015; Pourghasemi, 2016; Pourtaghi et al., 2016). We therefore randomly partitioned the fire locations in our inventory map into two portions and used 70% of the fire locations (92 fires comprising 1096 fire grid cells) for training the models and the remaining 30% (40 fires comprising 470 fire grid cells) for validating the models.

Since wildfire susceptibility modeling is done as a binary classification, in which the susceptibility index is classified into two classes (i.e., 'fire' and 'non-fire'), all of the 1566 grid cells denoting the presence of a fire were assigned the value of "1". We then randomly sampled the same number of grid cells from those cells without any evidence of wildfire and assigned a value of "-1" to those cells. Thus, a total of 2192 grid cells were used for the training and 940 grid cells for the validation model. Data-processing and calculations were carried out using the spatial analyst tools of the ArcGIS software.

3.2. Variable selection

It is very likely that some of the 15 explanatory variables are not relevant or that some might even introduce bias into the final model, which can actually decrease the prediction quality and enhance computational time. As a remedy, finding a subset of predictor variables that exhibit lower correlations with each other will actually result in a more accurate and robust estimate of wildfire probability. We thus

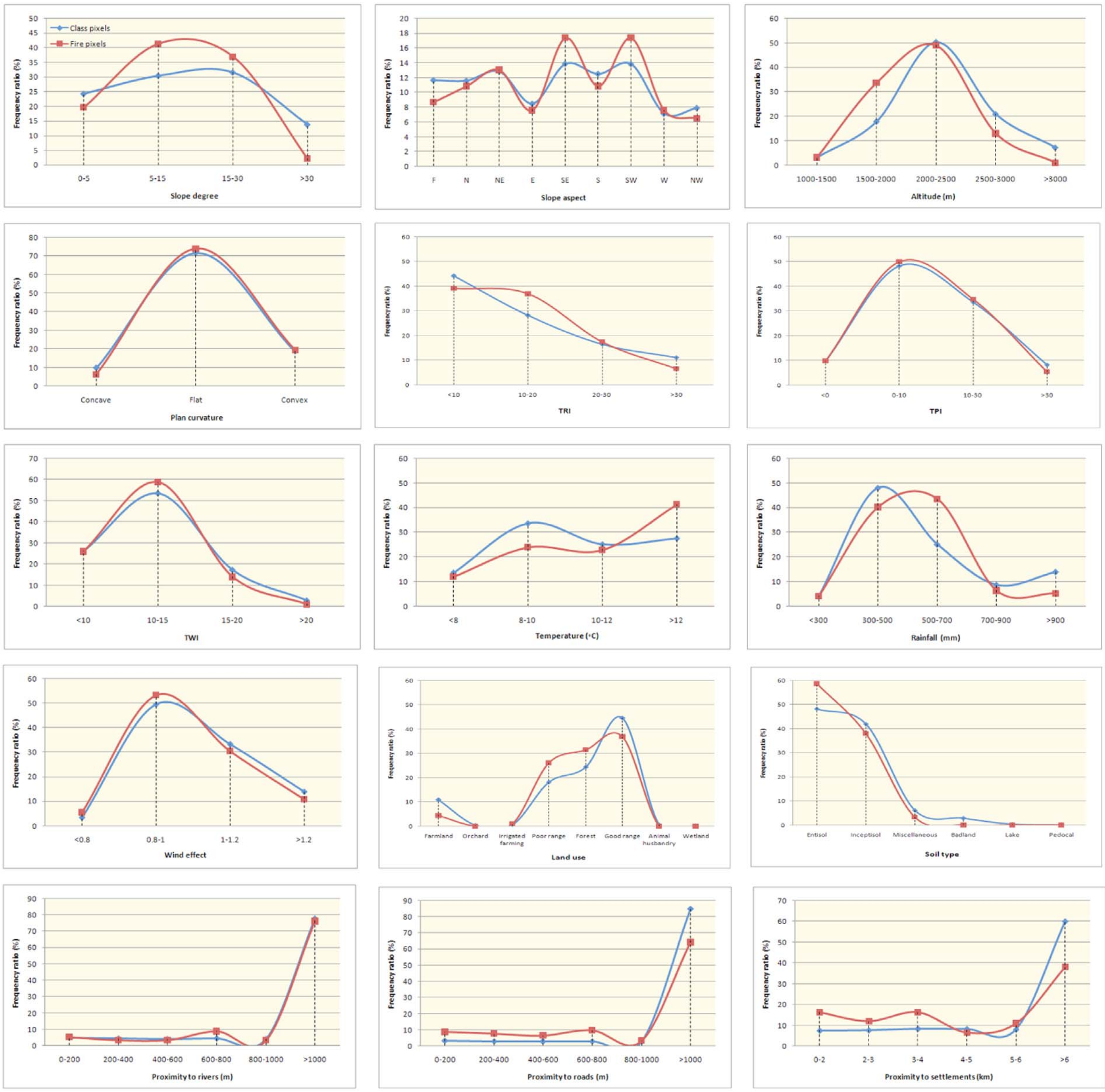

Fig. 4. Frequency ratio of pixels on wildfire explanatory variables map.

applied a two-step process, using multicollinearity analysis and the Gain Ratio with a 10-fold cross validation technique, to find the optimal set of explanatory variables that contribute the most to wildfire occurrence in our study area. This two-step process first eliminated some highly correlated predictor variables and then ranked the predictive capabilities of the remaining variables. Variables with null predictive usefulness were thus excluded from the modeling process. Two-step variable selection processes have been widely used in the context of natural hazard predictions and have produced good results (e.g., Pham et al., 2016; Tien Bui et al., 2016b).

3.2.1. *Multicollinearity analysis*

Multicollinearity exists when two or more predictor variables are highly correlated, which can cause a less accurate estimate of the effect of an independent variable on the dependent variable than when the independent variables are uncorrelated with one another (Dormann et al., 2013). In this study, we used the variance inflation factors (VIF) and tolerances (Liao and Valliant, 2012) to check for multicollinearity among wildfire explanatory variables. Critical values of VIF > 5 and tolerance < 0.2 indicate a potential problem with multicollinearity (Liao and Valliant, 2012). We used the SPSS statistical software to perform collinearity diagnostic procedures and excluded variables with VIF > 5 and/or tolerance < 0.2 from further analyses.

3.2.2. *Gain ratio method*

The Gain Ratio method (Quinlan, 1993) is an effective approach for the selection of an optimal subset of variables (features) capable of representing the whole dataset (Dash and Liu, 1997). This method has been used to identify and remove as much irrelevant and unnecessary information as possible (Martínez-Álvarez et al., 2013; Tien Bui et al., 2016b). The Gain Ratio for a given training data set f, regarding the class attribute C, is calculated as:

$$Entropy = -\sum_{i=1}^{k} \mathrm{P}(\mathrm{c_i})\log_2(\mathrm{P}(\mathrm{c_i})) \tag{1}$$

$$\mathrm{GainRatio}(f, \mathrm{C}) = \frac{\mathrm{Gain}(f, \mathrm{C})}{\mathrm{SplitInfo}(f, \mathrm{C})} \tag{2}$$

where k is the number of independent variables (features), entropy corresponds to the uncertainty about the value of the class attribute C, P

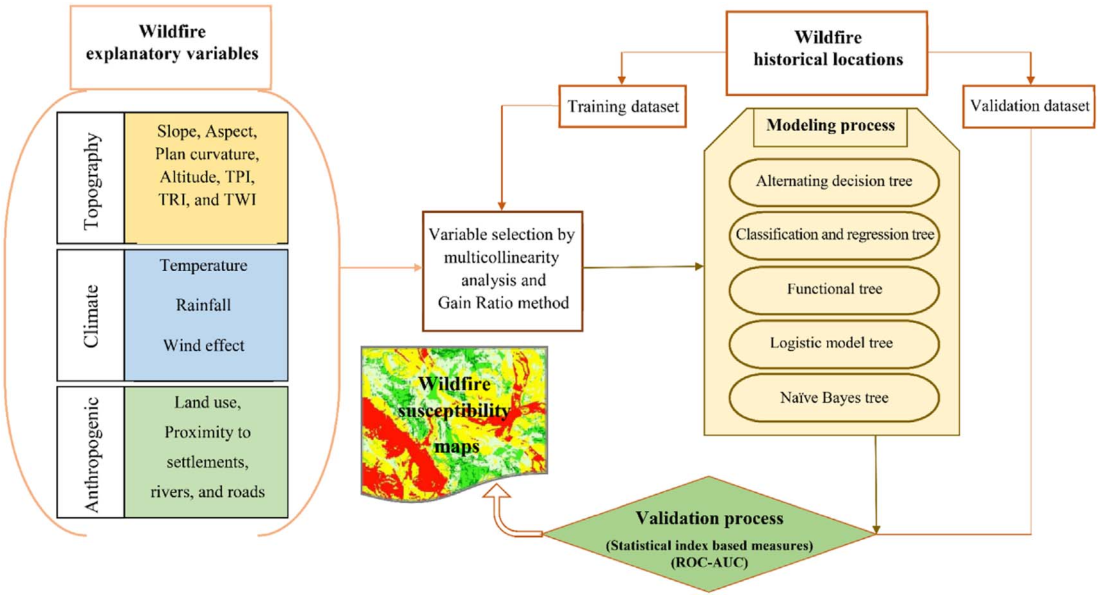

Fig. 5. Methodology adopted in this study.

(c_i) is the probability that C = c_i, and SplitInfo represents the potential information obtained by dividing attribute C of the training data set f into m subsets. SplitInfo is estimated as:

$$\mathrm{SplitInfo}(f, \mathrm{C}) = -\sum_{j=1}^{m} \frac{|f_j|}{|f|} \log_2 \frac{|f_j|}{|f|} \tag{3}$$

Variables with a "0" Gain Ratio (average merit) were excluded from the next steps in the modeling process. The Gain Ratio associated with each explanatory variable was assessed using the open source WEKA software Ver. 3.7 (Hall et al., 2009).

3.3. Overview of decision tree based classifiers

Within this subsection we briefly describe the five classifiers that were used in this study for modeling wildfire susceptibility and refer the interested reader to Freund and Mason (1999), Breiman et al. (1984), Gama (2004), Landwehr et al. (2005), and (Kohavi, 1996) for a full description of each technique.

Alternating Decision Tree (ADT) is a combination of boosting techniques and decision tree algorithms intended to improve the predictive accuracy of binary classification issues. In the ADT classifier, the number of boosting iterations needs to be manually tuned. More boosting iterations will result in larger (potentially more accurate) decision trees and a slower learning process. A small number of boosting iterations will result in a small decision tree that often performs poorly. In this study, 2200 boosting iterations resulted in the best performance of the algorithm.

Classification and Regression Tree (CART) encompasses a non-parametric regression method that grows a decision tree based on a binary partitioning algorithm that recursively splits the data into homogenous domains (Aertsen et al., 2011; Myles et al., 2004), forming a root, leaves, branches and nodes that mimic natural trees. For best performance, we tuned three of the four CART parameters, and set the number of folding for pruning at 100, seed at 90, and the minimum number of objects at 1.

Functional Trees (FT) represents a hierarchical model framework to construct multivariate trees for regression and classification problems. The most important FT feature is the logistic regression function at the inner nodes and/or leaves, used for splitting the inner nodes and for prediction at the leaves, instead of dividing the inputs at a tree node by comparing the value of some input attributes with a constant value (Pham et al., 2015). The combination of functional leaves and functional inner nodes with 4 instances per leaf and 50 boosting iterations resulted in the best performance of FT in this study.

Logistic Model trees (LMT) are an addition to decision trees that replace the terminal nodes of a decision tree with logistic regression functions (Dancey et al., 2007). LMT has two parameters that affect the performance of the classifier, i.e., the minimum number of instances per leaf and the number of bootstrap iterations, which were found to be 15 and -1, respectively, for best performance in this study.

Naïve Bayes Tree (NBT) is a hybrid of decision tree classifiers and Naïve Bayes. The decision-tree nodes contain univariate splits as do regular decision trees, whereas the leaves contain Naïve Bayes classifiers (Kohavi, 1996). The intuition is that Naïve Bayes classifiers work better than decision trees when the sample data set is small (Liang and Yan, 2006). NBT is a suitable approach for learning scenarios when many attributes are relevant for a classification.

Application of these five classifiers was hosted in the WEKA 3.7 software. The WEKA classification module was used to explore different values for the various parameters of each classifier to maximize the performance of each classifier.

3.4. Accuracy assessment and comparison

3.4.1. Statistical index-based measures

We used several statistical index-based measures, i.e., accuracy, sensitivity, specificity, precision, F-measure, and the Kappa index to evaluate the predictive abilities of each classifier. Accuracy is defined as the overall efficiency of a classifier and is computed as the proportion of pixels that were correctly classified (Eq. 4). Sensitivity (recall) is the proportion of correctly classified fire pixels out all pixels that were correctly classified as fire pixels plus those incorrectly classified as non-fire pixels (Eq. 5). Specificity is the proportion of correctly classified non-fire pixels out of all pixels that were correctly classified as non-fire pixels plus those incorrectly classified as fire pixels (Eq. 6). Precision is the proportion of correctly classified fire pixels out of all pixels classified as fire pixels (Eq. 7). The F-measure is the harmonic mean of precision and sensitivity (Eq. 8). All of these measures were calculated using the confusion matrices that are intermediate results of the ADT, CART, FT, LMT, and NBT analyses performed in the WEKA workbench.

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{4}$$

Table 1
The multicollinearity diagnosis statistics for the explanatory variables.

Variable	Collinearity statistics	
	Tolerance	VIF
Slope	0.242	4.128
Aspect	0.987	1.013
Altitude	0.282	3.551
Plan curvature	0.894	1.178
TPI	0.793	1.261
TRI	0.252	3.962
TWI	0.708	1.413
Temperature	0.347	2.879
Rainfall	0.875	1.143
Wind effect	0.505	1.979
Land use	0.996	1.004
Soil type	**0.004**	**21.819**
Proximity to rivers	0.804	1.244
Proximity to roads	0.873	1.146
Proximity to settlements	0.872	1.146

The excluded variable is shown in bold.

$$Sensitivity = \frac{TP}{TP + FN} \quad (5)$$

$$Specificity = \frac{TN}{TN + FP} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$F - measure = \frac{2TP}{2TP + FP + FN} \quad (8)$$

where TN (true negative) and TP (true positive) are the number of pixels that are properly assigned as fire occurrences and FP (false positive) and FN (false negative) are the numbers of pixels erroneously assigned (Pham et al., 2015).

We used the Kappa index to evaluate the reliability of the five classifiers. The Kappa index (K) is estimated as:

$$K = \frac{P_{obs} - P_{exp}}{1 - P_{exp}} \quad (9)$$

$$P_{obs} = (TP + TN) \quad (10)$$

$$P_{exp} = ((TP + FN) \times (TP + FP) + (FP + TN) \times (FN + TN)) \quad (11)$$

where P_{obs} is the proportion of pixels that were correctly classified as fire or non-fire, and P_{exp} represents the proportion of pixels for which the agreement is expected by chance alone (Hoehler, 2000). A Kappa index value equal to zero donates a random agreement between the predicted and the observed events and a Kappa value equal to one indicates perfect agreement and perfect model performance (Van Den Eeckhaut et al., 2009). Landis and Koch (1977) characterized values < 0 as indicating no agreement and 0–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1 as almost perfect agreement.

3.4.2. Receiver operating characteristic curve

The global performance of the classifiers was measured using the ROC curve method. The ROC is a common analytic method used to assess the accuracy of prediction models (Jaafari et al., 2015) and graphically represents the trade-off between the 1–specificity and sensitivity rates for every possible cutoff value (Egan, 1975). The plot shows 1–specificity on the X axis (Eq. 12) and sensitivity on the Y axis (Eq. 13):

$$X = 1 - Specificity = 1 - \left(\frac{TN}{TN + FP}\right) \quad (12)$$

$$Y = Sensitivity = \left(\frac{TP}{TP + FN}\right) \quad (13)$$

The best possible ROC curve passes through the point of (0, 1), where the area under curve (AUC) = 1 and represents 100% specificity (no false positives; all non-fire pixels correctly predicted) and 100% sensitivity (no false negatives; all fire pixels correctly predicted). AUC values of 0.5–0.6 indicate a poor, 0.6–0.7 a moderate, 0.7–0.8 a good, 0.8–0.9 a very good, and > 0.9 an excellent model performance (Hosmer Jr et al., 2013).

3.5. Development of wildfire susceptibility maps

After successfully training and validating the models, wildfire susceptibility maps were produced. To do so, we first computed the

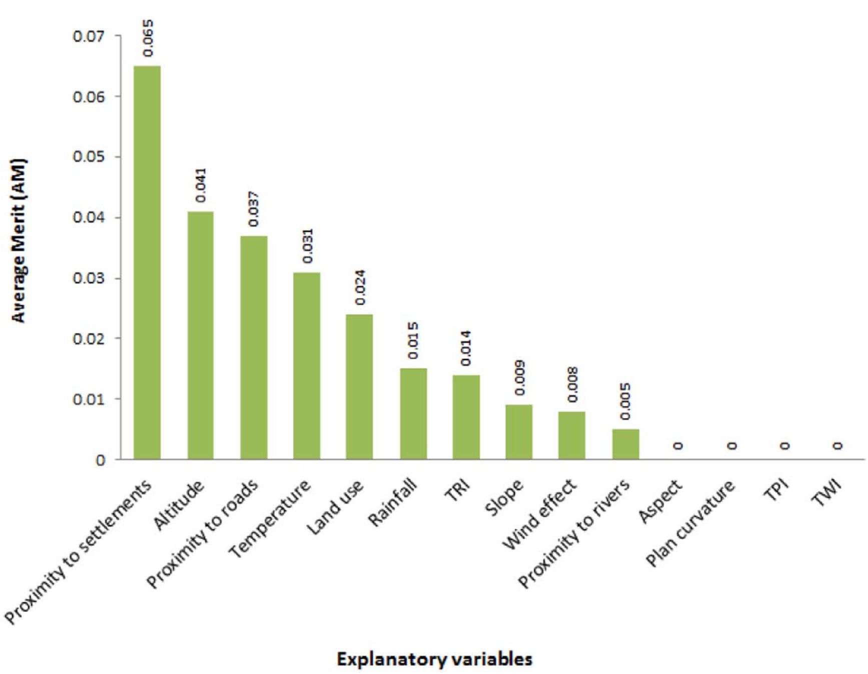

Fig. 6. Rank variables by importance using the Gain Ratio method.

```
|  (1)Prox2settlements < 5.5: 0.294
|  |  (22)Slope < 1.5: -0.165
|  |  |  (25)Prox2roads < 5.5: -0.217
|  |  |  |  (388)Prox2settlements < 3.5: -0.116
|  |  |  |  (388)Prox2settlements >= 3.5: 0.405
|  |  |  |  (552)Altitude < 3.5: -0.367
|  |  |  |  (552)Altitude >= 3.5: 0.044
|  |  |  (25)Prox2roads >= 5.5: 0.132
|  |  |  |  (126)Temperature < 2.5: -0.555
|  |  |  |  |  (263)Rainfall < 2.5: -0.424
|  |  |  |  |  |  (699)Landuse < 1.5: 0.258
|  |  |  |  |  |  (699)Landuse >= 1.5: -0.552
|  |  |  |  |  (263)Rainfall >= 2.5: 0.088
|  |  |  |  (126)Temperature >= 2.5: 0.084
|  |  |  |  |  (406)Rainfall < 2.5: -0.019
|  |  |  |  |  (406)Rainfall >= 2.5: 0.332
|  |  |  |  |  (482)Prox2settlements < 4.5: -0.051
|  |  |  |  |  (482)Prox2settlements >= 4.5: 0.2
|  |  |  (920)Prox2settlements < 4.5: -0.035
|  |  |  (920)Prox2settlements >= 4.5: 0.153
|  |  |  |  (1172)Wind eff. < 2.5: 0.26
|  |  |  |  (1172)Wind eff. >= 2.5: -0.077
|  |  |  |  (1255)Landuse < 3.5: -0.097
|  |  |  |  (1255)Landuse >= 3.5: 0.157
|  |  |  (1628)Temperature < 2.5: -0.126
|  |  |  (1628)Temperature >= 2.5: 0.032
|  |  (22)Slope >= 1.5: 0.038
|  |  |  (99)Prox2roads < 1.5: 0.672
|  |  |  |  (281)TRI < 1.5: -0.413
|  |  |  |  (281)TRI >= 1.5: 0.461
|  |  |  |  (604)Prox2settlements < 3.5: 0.389
|  |  |  |  (604)Prox2settlements >= 3.5: -0.241
|  |  |  (99)Prox2roads >= 1.5: -0.013
|  |  |  |  (275)Prox2roads < 2.5: -0.25
|  |  |  |  (275)Prox2roads >= 2.5: -0.002
|  |  |  |  |  (635)Prox2roads < 5.5: 0.139
|  |  |  |  |  (635)Prox2roads >= 5.5: -0.012
|  |  |  (361)Wind eff. < 2.5: 0.045
```

Fig. 7. A part of the rules defined by the ADT classifier to predict wildfire occurrence.

susceptibility to wildfire values for each pixels in the study area and then classified the values using the Natural Breaks (Jenks) method (Jenks and Caspall, 1971) by grouping them into relative levels that depict very low, low, moderate, high, and very high susceptibility to wildfire occurrence across the study area. The reliability of the susceptibility maps was investigated using fire density analysis. Fire density is defined as the ratio of the number of fire pixels within each susceptibility class to the total number of pixels in the study landscape. It is expected that the fire density index increases from the very low to the very high susceptibility levels (Nami et al., 2017; Razavizadeh et al., 2017).

Table 3
Kappa index in the training and validation datasets.

Classifier	Kappa index	
	Training dataset	Validation dataset
ADT	0.886	0.757
CART	0.843	0.734
LMT	0.808	0.711
FT	0.800	0.695
NBT	0.684	0.609

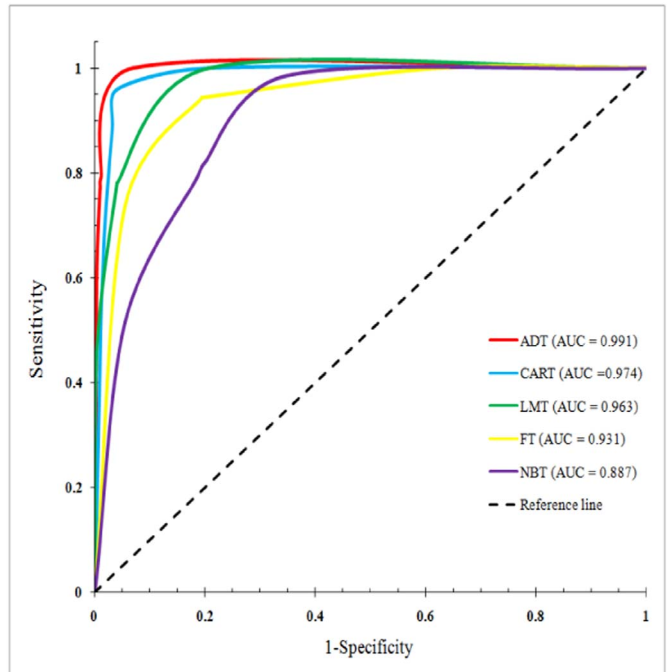

Fig. 8. The ROC-AUC for the training dataset.

4. Results

4.1. Variable analysis

The results of the multicollinearity diagnostic procedure indicate that the soil type variable did not satisfy the critical values of VIF ≤ 5 and tolerance ≥ 0.2 (Table 1), suggesting that this variable should be excluded from further analyses. The average merit based on the Gain Ratio method shows that four explanatory variables (i.e., aspect, plan curvature, TPI, and, TWI) represent "0" average merit, which indicates "null" usefulness to wildfire modeling (Fig. 6). Since the inclusion of these variables may in fact negatively influence the resulting models, they were also excluded from further analyses. After these two preliminary analyses, 10 out of the original 15 explanatory variables were retained for further wildfire modeling.

Table 2
Model performance in the training and validation datasets.

Measure (%)	Training dataset					Validation dataset				
	ADT	CART	LMT	FT	NBT	ADT	CART	LMT	FT	NBT
Accuracy	94.30	92.02	90.42	90.01	84.22	87.87	86.70	85.53	82.23	77.96
Sensitivity	94.58	92.61	90.35	90.60	85.26	89.24	87.11	87.52	83.59	79.13
Specificity	94.02	91.42	90.49	89.42	83.27	86.62	86.31	83.81	81.03	77.90
Precision	93.98	91.52	90.51	89.54	82.17	85.53	85.96	81.91	80.36	76.30
F-measure	94.28	92.06	90.43	90.07	83.68	87.34	86.53	84.61	81.41	78.16

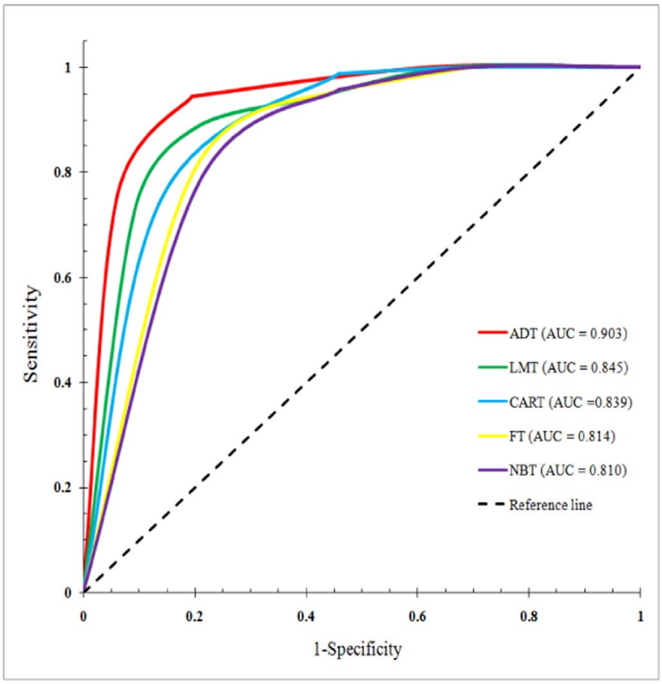

Fig. 9. The ROC-AUC for the validation dataset.

Table 4
Performance of the ADT classifier using fifteen variables.

Measure	ADT	
	Training dataset	Validation dataset
Accuracy (%)	97.77	74.68
Sensitivity (%)	97.73	72.55
Specificity (%)	97.82	76.81
Precision (%)	97.82	75.78
F-measure (%)	97.77	74.13
AUC	0.990	0.825
Kappa	0.905	0.494

4.2. Training and validating the wildfire models

Classification rules for predicting wildfire occurrences derived from the training dataset (Fig. 7) resulted in values of the different statistical index-based performance measures that ranged from ~94 to ~82% for the different classifiers (Table 2). All of the performance measures revealed the same ranking of classifiers, with the highest values consistently achieved by the ADT classifier, followed by CART, LMT, FT, and NBT, respectively. The ADT classifier correctly classified 94.30% of all training dataset pixels (= accuracy), with 94.58% of properly classified fire pixels into the fire class (that also included false negatives) (= sensitivity), 94.02% of properly classified non-fire pixels in the non-fire class (that also included false positives) (= specificity), 93.98% of fire pixels correctly categorized in the fire class (that also included false positives) (= precision), and an overall greatest F-measure value of 94.28%.

The statistical index-based performance measures for the validation dataset revealed that all models predicted future wildfires quite well (Table 2). Again, ADT resulted in the model with the greatest predictive ability, followed by the CART, LMT, FT, and NBT models, respectively.

The Kappa indices for the various classifiers ranged from 0.886 (ADT) to 0.684 (NBT) for the training dataset and from 0.757 (ADT) to 0.609 (NBT) for the validation dataset (Table 3), indicating a substantial (Kappa > 0.6) to almost perfect (Kappa > 0.8) agreement between predicted and observed fire events.

Analyses of the global performances of the five classifiers, measured using the ROC-AUC, showed that the greatest goodness-of-fit was achieved by the ADT classifier in both the training dataset (AUC = 0.991; Fig. 8) and in the validation dataset (AUC = 0.903; Fig. 9), indicating that the agreement of prediction of the classifier with reality was 90.30%. The LMT was ranked as the second most powerful classifier for wildfire prediction ($AUC_{validation}$ = 0.845), followed by CART ($AUC_{validation}$ = 0.838), FT ($AUC_{validation}$ = 0.826), and NBT ($AUC_{validation}$ = 0.804).

An evaluation of the efficiency of the ADT classifier using all 15 variables revealed that the resulting model performed slightly better in classifying the training dataset than the model with 10 variables, but that its performance decreased significantly in the validation phase (i.e., for predicting future ignitions; Table 4).

4.3. Susceptibility maps

Whereas wildfire susceptibility maps based on the models derived from the five classification methods consistently depicted the presence of all five (very low to very high) wildfire susceptibility classes across the entire study region (Fig. 10), the relative ranking of the likelihood of wildfire occurrence depended somewhat on the chosen classification method. For example, zones of high and very high susceptibility to wildfire encompassed between 23 and 8% of the study area (Fig. 11). The ADT, LMT, and CART models showed consistently greater values than the FT and NBT models (i.e., ~25% vs. 18%), particularly in the high susceptibility class. The results of the reliability analysis of the susceptibility maps showed that the value of fire density varied among the classes and ranged from 0.37 in the very low susceptibility level of the ADT and LMT models to 4.10 in the very high susceptibility level of the NBT model (Fig. 12). In each map, the highest value belongs to very high susceptibility level which is followed by high class, moderate class, and low level, respectively, indicating that the models are capable of properly delimiting the study area into different susceptibility levels of historic fire occurrences.

5. Discussion

Developing a reliable estimate of wildfire susceptibility is challenging because of geological, topographical and environmental complexities over the landscape. Although several methods for the prediction of natural hazards have been proposed, the evaluation of capabilities of these methods for wildfire prediction still lags. In this study, we evaluated five decision tree based classifiers in the context of wildfire susceptibility modeling and compared the predictive abilities of the models to assign differential, spatially explicit wildfire susceptibilities across the study landscape. Utilizing a wide array of explanatory variables for prediction, it was first necessary to perform a two-step process using multicollinearity analysis and the Gain Ratio method to remove irrelevant and unnecessary predictor/explanatory variables, reduce potential bias, and achieve a higher prediction quality (Martínez-Álvarez et al., 2013; Tien Bui et al., 2016b). Indeed, there were very clear signals that a single application of a multicollinearity diagnostic as proposed by many authors (e.g., Chen et al., 2015; Oliveira et al., 2014; Pourtaghi et al., 2015) may not be sufficient to exclude marginally useful variables from that model that can actually reduce the predictive ability for forecasting future wildfire events. However, after following-up the multicollinearity analysis with the Gain Ratio method, we were able to eliminate all of the variables that introduced null usefulness to model performance and identified ten variables that were most relevant and decisive for the occurrence of wildfire within our study area. The utility of the variable selection process was demonstrated when a full predictive model with all 15 explanatory variables and a reduced predictive model with ten explanatory variables were built for the validation dataset with the ADT classifier, which exhibited superior performance over other classifiers

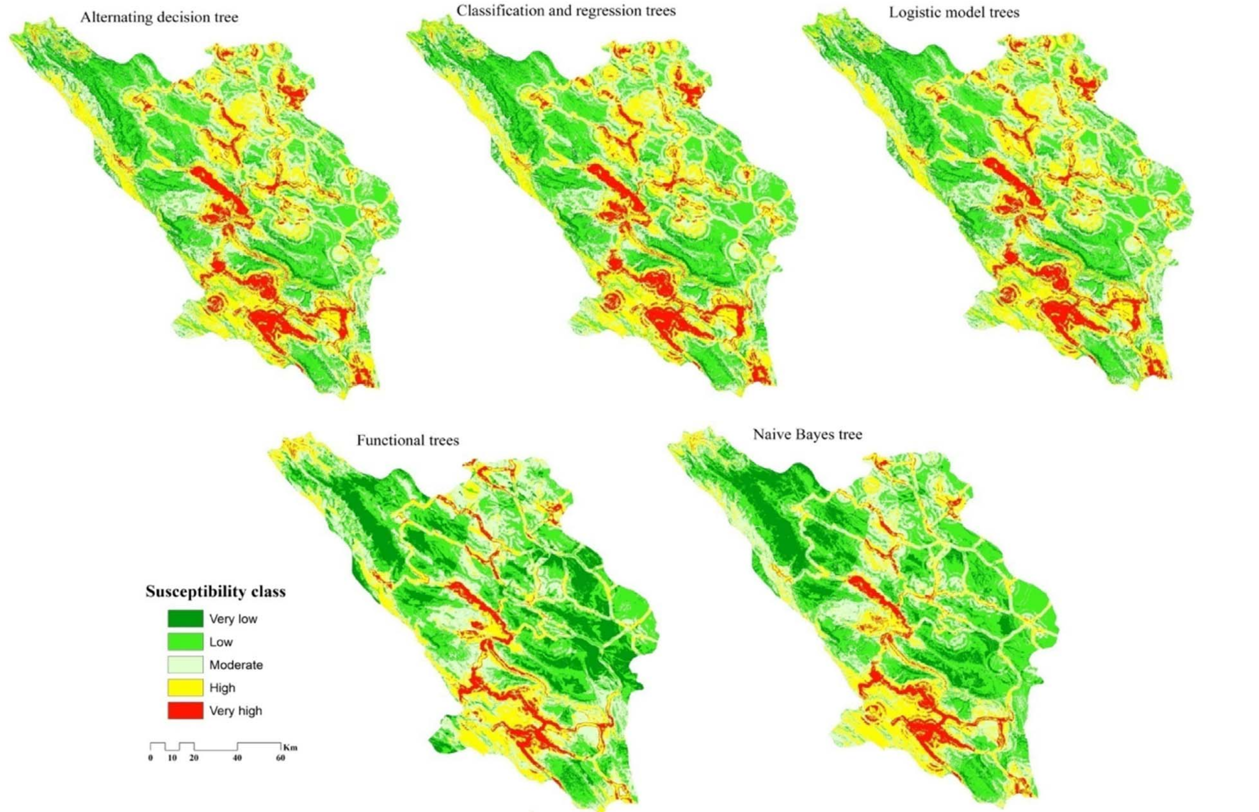

Fig. 10. Wildfire susceptibility maps.

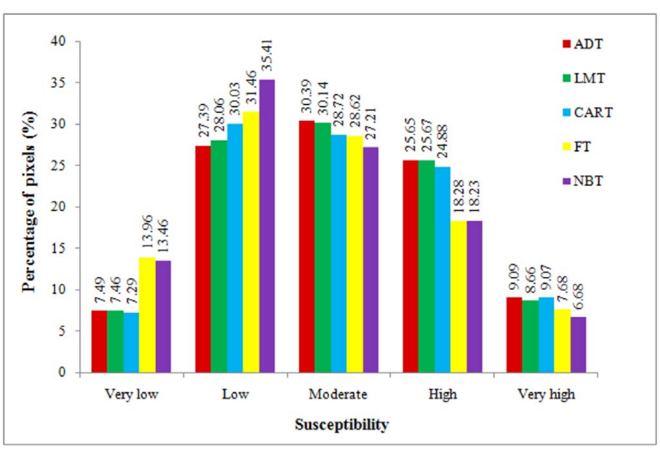

Fig. 11. Distribution of susceptibility levels in five wildfire susceptibility maps.

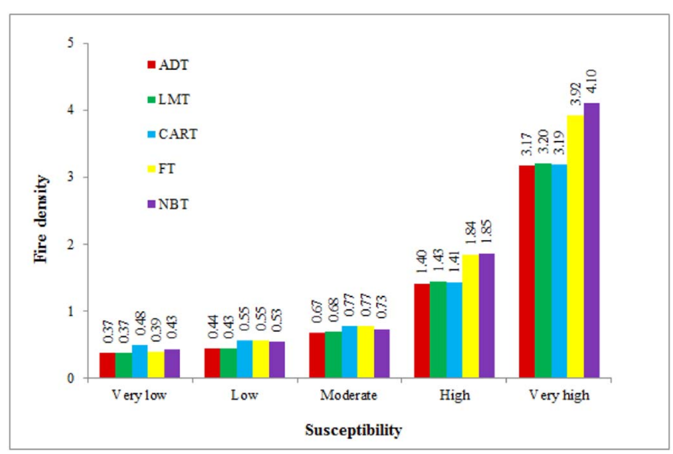

Fig. 12. Distribution of fire density across the five wildfire susceptibility classes.

in this study. The validation results clearly revealed that the capability of the ADT for predicting future wildfires decreased precipitously, even to lower levels than those obtained using the NBT classifier (i.e., the weakest classifier) with the original set of fifteen variables. These results are consistent with previous findings that data reduction often results in improved performances of decision trees, both in terms of the resulting tree-size and predictive capability (Piramuthu, 2008).

The modeling process using the remaining reduced set of ten explanatory variables clearly identified the ADT classifier as the classifier that yielded the best overall performance results during the model training and, more importantly, during the validation process. Although this is the first application of the ADT classifier in wildfire modeling, these positive results are mirrored by other modeling studies where the ADT classifier was employed to model landslide hazards (e.g., Hong et al., 2015), confirming the utility of this classifier over some other state-of-the art data mining methods. The particular advantage of the ADT over other classifiers is that it is quite simple to build, because only the boosting iteration needs to be tuned. In contrast to some other data mining models, such as SVM, the ADT also portrays a more transparent model structure with classification rules that are easy to interpret (Hong et al., 2015).

The CART classifier, which has been regularly applied for predicting wildfires (Amatulli et al., 2006; Lozano et al., 2008; McKenzie et al., 2000), exhibited a lower global performance (ROC-AUC) than the ADT and LMT classifiers. Although we did not find a study in the literature that compared the performance of CART to other classification methods in wildfire modeling, the results of the application of the CART classifier in other modeling studies (e.g., Aertsen et al., 2011; Felicísimo et al., 2013; Gong and Ordieres-Meré, 2016; Naghibi et al., 2016) support our findings that CART represents a data mining method of only moderate predictive capability. Whereas these other researchers concluded that the CART methodology was easy and straightforward to interpret, there was shared concern that its analytical outcomes may somewhat over-simplify many real-world problems. Nonetheless, despite superior assessment statistics and predictive capabilities of the ADT classifier, all five classifiers resulted in AUC values > 80% and exhibited very high levels of accuracy of wildfire predictions (cf. Arpaci et al., 2014; Pourtaghi et al., 2016). Thus, all five classifiers appear to be suitable for modeling wildfire susceptibility across fire-prone landscapes. It appears that ADT, LMT, and CART came up with strikingly similar distributions of wildfire susceptibility across the landscape that differed from FT and NBT that were very similar to one another (Fig. 10). Indeed, the NBT classifier exhibited the lowest prediction ability among all classifiers in this study, which may likely be attributed to its algorithm that is based on the assumption of independence among all input variables (Mujalli et al., 2016). Although the independence assumption may be a viable assumption in many other fields of study (Shirzadi et al., 2017; Zhang and Su, 2004), it seems that this assumption is not met when using topological, climatic, and land-use landscape-level variables for modeling wildfire susceptibility.

Overall, the results of reliability analysis of the five maps revealed that the zones of moderate, high, and very high susceptibility agreed well with the density of historical fire events. Thus, all five classifiers can be recommended for modeling wildfire susceptibility across fire-prone landscapes, and modelers can easily tailor these classifiers to their particular question of interest.

6. Conclusion

Faced with ongoing climate and land use changes, and increasing multifarious human impacts on terrestrial ecosystems, wildfire managers tasked with the allocation of fire prevention and suppression resources have now begun to adopt planning horizons of several years. Such longer-term management plans need to be informed by realistic model predictions that are increasingly made feasible by improved data-processing techniques and new predictive modeling approaches that have opened up new opportunities for making accurate and spatially reliable predictions of wildfire susceptibility across larger spaces. Using a robust approach to estimate wildfire susceptibility across a fire-prone landscape allows managers and decision makers to think through decisions when accurate and detailed forecasts are not possible. The approach suggested in this study was successful at identifying a small subset of landscape-level variables that contribute to the likelihood of wildfire occurrence and delimiting areas of different levels of fire susceptibility across the landscape. The modeling results revealed that all of the five classifiers investigated in this study performed reasonably well, with AUC values over 80%. Although we limited our analysis to five decision tree-based classifiers, there are many more data mining methods that have not yet been sufficiently investigated for wildfire susceptibility modeling. Future work could extend the current analysis to a comparison of the predictive capabilities of other data mining models in predicting the likely occurrence of wildland fires across landscapes. The results of these comparative analyses might indeed greatly contribute to adequately customizing data mining models for different conditions and environmental settings.

Finally, it should be remembered that there is always a trade-off between the quality of the data, the resources involved, and the reliability of the wildfire susceptibility prediction. To achieve the best quality, it is very important to invest in accurate wildfire records and databases that include important explanatory variables.

Acknowledgement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We thank Dr. Davood Mafi-Gholami and administrative office of natural resources of the Chaharmahal-Bakhtiari Province, Iran for providing us with the data used in this study.

References

Adab, H., Kanniah, K.D., Solaimani, K., 2013. Modeling forest fire risk in the northeast of Iran using remote sensing and GIS techniques. Nat. Hazards 65 (3), 1723–1743.

Adab, H., Kanniah, K.D., Solaimani, K., Sallehuddin, R., 2015. Modelling static fire hazard in a semi-arid region using frequency analysis. Int. J. Wildland Fire 24 (6), 763–777.

Aertsen, W., Kint, V., Van Orshoven, J., Muys, B., 2011. Evaluation of modelling techniques for forest site productivity prediction in contrasting ecoregions using stochastic multicriteria acceptability analysis (SMAA). Environ. Model Softw. 26 (7), 929–937.

Amatulli, G., Rodrigues, M.J., Trombetti, M., Lovreglio, R., 2006. Assessing long-term fire risk at local scale by means of decision tree technique. J. Geophys. Res. Biogeosci. 111 (G4).

Arpaci, A., Malowerschnig, B., Sass, O., Vacik, H., 2014. Using multivariate data mining techniques for estimating fire susceptibility of Tyrolean forests. Appl. Geogr. 53, 258–270.

Bedia, J., Herrera, S., Gutiérrez, J.M., Benali, A., Brands, S., Mota, B., Moreno, J.M., 2015. Global patterns in the sensitivity of burned area to fire-weather: implications for climate change. Agric. For. Meteorol. 214, 369–379.

Beygi, H.H., Banj, S.A., Erfanian, M., 2014. Forest fire risk mapping using analytical hierarchy process technique and frequency ratio method (case study: Sardasht forests, NW Iran). Iranian J. Forest Pop. Res. 22 (4), 559–573.

Boubeta, M., Lombardía, M.J., Marey-Pérez, M.F., Morales, D., 2015. Prediction of forest fires occurrences with area-level Poisson mixed models. J. Environ. Manag. 154, 151–158.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees. CRC press.

Catry, F.X., Rego, F.C., Bação, F.L., Moreira, F., 2009. Modeling and mapping wildfire ignition risk in Portugal. Int. J. Wildland Fire 18 (8), 921–931.

Chen, F., Du, Y., Niu, S., Zhao, J., 2015. Modeling forest lightning fire occurrence in the Daxinganling Mountains of northeastern China with MAXENT. Forest 6 (5), 1422–1438.

Dancey, D., Bandar, Z.A., McLean, D., 2007. Logistic model tree extraction from artificial neural networks. IEEE Trans. Syst. Man Cybern. B Cybern. 37 (4), 794–802.

Dash, M., Liu, H., 1997. Feature selection for classification. Intell. Data Anal. 1 (1-4), 131–156.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., ... Münkemüller, T., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36 (1), 27–46.

Egan, J.P., 1975. Signal detection theory and {ROC} analysis. Trans. NY Acad. Sci. 195, 266–268.

Eskandari, S., Chuvieco, E., 2015. Fire danger assessment in Iran based on geospatial information. Int. J. Appl. Earth Obs. Geoinf. 42, 57–64.

Eskandari, S., Miesel, J.R., 2017. Comparison of the fuzzy AHP method, the spatial correlation method, and the Dong model to predict the fire high-risk areas in Hyrcanian forests of Iran. Geomat. Nat. Haz. Risk 1–17.

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., 1996. From data mining to knowledge discovery in databases. AI Mag. 17 (3), 37.

Felicísimo, Á.M., Cuartero, A., Remondo, J., Quirós, E., 2013. Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. Landslides 10 (2), 175–189.

Flannigan, M.D., Stocks, B.J., Wotton, B.M., 2000. Climate change and forest fires. Sci. Total Environ. 262 (3), 221–229.

Freund, Y., Mason, L., 1999. The alternating decision tree learning algorithm. ICML 99, 124–133.

Gama, J., 2004. Functional trees. Mach. Learn. 55 (3), 219–250.

Gong, B., Ordieres-Meré, J., 2016. Prediction of daily maximum ozone threshold exceedances by preprocessing and ensemble artificial intelligence techniques: case study of Hong Kong. Environ. Model Softw. 84, 290–303.

Guo, F., Wang, G., Su, Z., Liang, H., Wang, W., Lin, F., Liu, A., 2016. What drives forest fire in Fujian, China? Evidence from logistic regression and random forests. Int. J. Wildland Fire 25 (5), 505–519.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: an update. ACM SIGKDD Explorations Newsl. 11 (1), 10–18.

Hoehler, F.K., 2000. Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. J. Clin. Epidemiol. 53 (5), 499–503.

Hong, H., Pradhan, B., Xu, C., Bui, D.T., 2015. Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. Catena 133, 266–281.

Hosmer Jr., D.W., Lemeshow, S., Sturdivant, R.X., 2013. Applied Logistic Regression. vol. 398 John Wiley & Sons.

Jaafari, A., Mafi Gholami, D., 2017. Wildfire hazard mapping using an ensemble method of frequency ratio with Shannon's entropy. Iranian J. Forest Pop. Res. 25 (2), 232–243.

Jaafari, A., Najafi, A., Rezaeian, J., Sattarian, A., Ghajar, I., 2015. Planning road networks in landslide-prone areas: a case study from the northern forests of Iran. Land Use Policy 47, 198–208.

Jaafari, A., Mafi Gholami, D., Zenner, E.K., 2017. A Bayesian modeling of wildfire probability in the Zagros Mountains, Iran. Eco. Inform. 39, 32–44.

Jahdi, R., Salis, M., Darvishsefat, A.A., Alcasena, F., Mostafavi, M.A., Etemad, V., ... Spano, D., 2015. Evaluating fire modelling systems in recent wildfires of the Golestan National Park, Iran. Forestry 89 (2), 136–149.
Jenks, G.F., Caspall, F.C., 1971. Error on choroplethic maps: definition, measurement, reduction. Ann. Assoc. Am. Geogr. 61 (2), 217–244.
Kohavi, R., 1996. Scaling up the accuracy of Naive-Bayes classifiers: a decision-tree hybrid. In: KDD. vol. 96. pp. 202–207.
Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33 (1), 159–174.
Landwehr, N., Hall, M., Frank, E., 2005. Logistic model trees. Mach. Learn. 59 (1–2), 161–205.
Liang, H., Yan, Y., 2006. Learning Naive Bayes tree for conditional probability estimation. In: Conference of the Canadian Society for Computational Studies of Intelligence. Springer, Berlin Heidelberg, pp. 7–466.
Liao, D., Valliant, R., 2012. Variance inflation factors in the analysis of complex survey data. Survey Methodol. 38 (1), 53–62.
Lozano, F.J., Suarez-Seoane, S., Luis, E., 2008. A multi-scale approach for modeling fire occurrence probability using satellite data and classification trees: a case study in a mountainous Mediterranean region. Remote Sens. Environ. 112 (3), 708–719.
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. Knowl.-Based Syst. 50, 198–210.
McKenzie, D., Peterson, D.L., Agee, J.K., 2000. Fire frequency in the interior Columbia River basin: building regional models from fire history data. Ecol. Appl. 10 (5), 1497–1516.
Meteorological Organization of Iran (MOI), 2014. Annual report of climatic statistics. http://www.irimo.ir/english/.2012.
Mhawej, M., Faour, G., Adjizian-Gerard, J., 2015. Wildfire likelihood's elements: a literature review. Challenges 6 (2), 282–293.
Mhawej, M., Faour, G., Adjizian-Gerard, J., 2017. A novel method to identify likely causes of wildfire. Clim. Risk Manag. 16, 120–132.
Mohammadi, F., Bavaghar, M.P., Shabanian, N., 2014. Forest fire risk zone modeling using logistic regression and GIS: an Iranian case study. Small-scale Forestry 13 (1), 117–125.
Mujalli, R.O., López, G., Garach, L., 2016. Bayes classifiers for imbalanced traffic accidents datasets. Accid. Anal. Prev. 88, 37–51.
Myles, A.J., Feudale, R.N., Liu, Y., Woody, N.A., Brown, S.D., 2004. An introduction to decision tree modeling. J. Chemom. 18 (6), 275–285.
Naghibi, S.A., Pourghasemi, H.R., Dixon, B., 2016. GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. Environ. Monit. Assess. 188 (1), 1–27.
Nami, M.H., Jaafari, A., Fallah, M., Nabiuni, S., 2017. Spatial prediction of wildfire probability in the Hyrcanian ecoregion using evidential belief function model and GIS. Int. J. Environ. Sci. Technol. 1–12. http://dx.doi.org/10.1007/s13762-017-1371-6.
Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A., Pereira, J.M., 2012. Modeling spatial patterns of fire occurrence in Mediterranean Europe using multiple regression and random Forest. For. Ecol. Manag. 275, 117–129.
Oliveira, S., Pereira, J.M., San-Miguel-Ayanz, J., Lourenço, L., 2014. Exploring the spatial patterns of fire density in southern Europe using geographically weighted regression. Appl. Geogr. 51, 143–157.
Pan, J., Wang, W., Li, J., 2016. Building probabilistic models of fire occurrence and fire risk zoning using logistic regression in Shanxi Province, China. Nat. Hazards 81 (3), 1879–1899.
Pham, B.T., Bui, D.T., Pourghasemi, H.R., Indra, P., Dholakia, M.B., 2015. Landslide susceptibility assessment in the Uttarakhand area (India) using GIS: a comparison study of prediction capability of naïve bayes, multilayer perceptron neural networks, and functional trees methods. Theor. Appl. Climatol. 1–19.
Pham, B.T., Pradhan, B., Bui, D.T., Prakash, I., 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 Softw. 84, 240–250.
Piramuthu, S., 2008. Input data for decision trees. Expert Syst. Appl. 34 (2), 1220–1226.
Pourghasemi, H.R., 2016. GIS-based forest fire susceptibility mapping in Iran: a comparison between evidential belief function and binary logistic regression models. Scand. J. For. Res. 31 (1), 80–98.
Pourtaghi, Z.S., Pourghasemi, H.R., Rossi, M., 2015. Forest fire susceptibility mapping in the Minudasht forests, Golestan province, Iran. Environ. Earth Sci. 73 (4), 1515–1533.
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.
Quinlan, J.R., 1993. C45: Programs for Machine Learning. Morgan Kaufmann, San Mateo.
Razavizadeh, S., Solaimani, K., Massironi, M., Kavian, A., 2017. Mapping landslide susceptibility with frequency ratio, statistical index, and weights of evidence models: a case study in northern Iran. Environ. Earth Sci. 76 (14), 499.
Rodrigues, M., de la Riva, J., 2014. An insight into machine-learning algorithms to model human-caused wildfire occurrence. Environ. Model Softw. 57, 192–201.
Shirzadi, A., Bui, D.T., Pham, B.T., Solaimani, K., Chapi, K., Kavian, A., ... Revhaug, I., 2017. Shallow landslide susceptibility assessment using a novel hybrid intelligence approach. Environ. Earth Sci. 76 (2), 60.
Stambaugh, M.C., Guyette, R.P., 2008. Predicting spatio-temporal variability in fire return intervals using a topographic roughness index. For. Ecol. Manag. 254 (3), 463–473.
Statistical Centre of Iran (SCI), 2011. Survey of Chaharmahal-Bakhtiari. Publications Information Base Statistical Centre of Iran, Tehran.
Tayyebi, A., Pijanowski, B.C., Linderman, M., Gratton, C., 2014. Comparing three global parametric and local non-parametric models to simulate land use change in diverse areas of the world. Environ. Model Softw. 59, 202–221.
Teodoro, A.C., 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.
Tien Bui, D., Le, K.T.T., Nguyen, V.C., Le, H.D., 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.
Tien Bui, D., Tuan, T.A., Klempe, H., Pradhan, B., Revhaug, I., 2016b. 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 (2), 361–378.
Tien Bui, D., Bui, Q.T., Nguyen, Q.P., Pradhan, B., Nampak, H., Trinh, P.T., 2017. 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.
Van Den Eeckhaut, M., Reichenbach, P., Guzzetti, F., Rossi, M., Poesen, J., 2009. Combined landslide inventory and susceptibility assessment based on different mapping units: an example from the Flemish Ardennes, Belgium. Nat. Hazards Earth Syst. Sci. 9 (2), 507–521.
Zhang, H., Su, J., 2004. Naïve Bayesian classifiers for ranking. In: European Conference on Machine Learning. Springer, Berlin Heidelberg, pp. 501–512.