Ecological Indicators 64 (2016) 72–84

Contents lists available at ScienceDirect

Ecological Indicators

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

Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques

Zohre Sadat Pourtaghi[a], Hamid Reza Pourghasemi[b,*], Roberta Aretano[c], Teodoro Semeraro[c]

[a] Department of Environment Management Engineering, College of Natural Resources, Yazd University, Yazd, Iran
[b] Department of Natural Resources and Environmental Engineering, College of Agriculture, Shiraz University, Shiraz, Iran
[c] Laboratory of Landscape Ecology, Department of Biological and Environmental Sciences and Technologies, University of Salento, Lecce, Italy

ARTICLE INFO

Article history:
Received 8 July 2015
Received in revised form 8 December 2015
Accepted 18 December 2015

Keywords:
Forest fire susceptibility mapping
Boosted regression tree
Generalized additive model
Random forest
GIS
Iran

ABSTRACT

Forests are living dynamic systems and these unique ecosystems are essential for life on earth. Forest fires are one of the major environmental concerns, economic, and social in the worldwide. The aim of current research is to identify general indicators influencing on forest fire and compare forest fire susceptibility maps based on the boosted regression tree (BRT), generalized additive model (GAM), and random forest (RF) data mining models in the Minudasht Township, Golestan Province, Iran. According to expert opinion and literature review, fifteen condition factors on forest fire have been selected in the study area. These are slope degree, slope aspect, elevation, topographic wetness index (TWI), topographic position index (TPI), plan curvature, wind effect, annual temperature and rainfall, soil texture, distance to roads, rivers, and villages, normalized difference vegetation index (NDVI), and land use. Forest fire locations were identified using MODIS images, historical records, and extensive field checking. 106 (≈70%) locations, out of 151 forest fires identified, were used for models building/training, while the remaining 45 (≈30%) cases were used for the models validation.

BRT, GAM, and RF data mining models were used to distinguish between presence and absence of forest fires and its mapping. These algorithms were used to perform feature selection in order to reveal the variables that contribute more to forest fire occurrence. Finally, for validation of models, the area under the curve (AUC) for forest fire susceptibility maps was calculated. The validation of results showed that AUC for three mentioned models varies from 0.7279 to 0.8770 (AUC_{BRT} = 80.84%, AUC_{GAM} = 87.70%, and AUC_{RF} = 72.79%,). Results indicated that the main drivers of forest fire occurrence were annual rainfall, distance to roads, and land use factors. The results can be applied to primary warning, fire suppression resource planning, and allocation work.

1. Introduction

Forests are living dynamic systems with a biologically diverse and complicated structure (Aleemahmoodi Sarab et al., 2013). Forests in the world are a source of ecosystem services essential for human well-being, playing a crucial role in the persistence of vital processes in the environment, including weather adjustment, carbon preservation (UNEP, 2007) and regulating the climate and the carbon cycle. However, forest ecosystems are increasingly threatened by fires caused by natural and anthropogenic factors (Chen et al., 2012; Naebi, 2003; Ghomi Motazeh et al., 2013). Fires can be a destructive ecological factor causing many negative effects in various aspects of life, including natural environment, economics, and health (Herawati et al., 2006; Ghomi Motazeh et al., 2013), but a good plan can provide suitable tools for ecosystem management. In this content, fire prevention must be paid special attention (Ghomi Motazeh et al., 2013). Fire is a natural power that affects vegetation communities over time and, as a natural process, it has a significant effect on conserving ecosystems health. Since twentieth century, increasing fires due to human inattention turned fires to a main threat for forests (Nasi et al., 2002; Ghomi Motazeh et al., 2013).

It is possible both to control nature and also to provide the fire risk map and thereby to minimize the fire frequency and avert damage (Jaiswal et al., 2002). A precise evaluation of forest fire problems and decision on solutions can be satisfactory when fire risk mapping is available (Jaiswal et al., 2002; Erten

* Corresponding author.
E-mail addresses: hamidreza.pourghasemi@yahoo.com, hr.pourghasemi@shirazu.ac.ir (H.R. Pourghasemi).

http://dx.doi.org/10.1016/j.ecolind.2015.12.030
1470-160X/

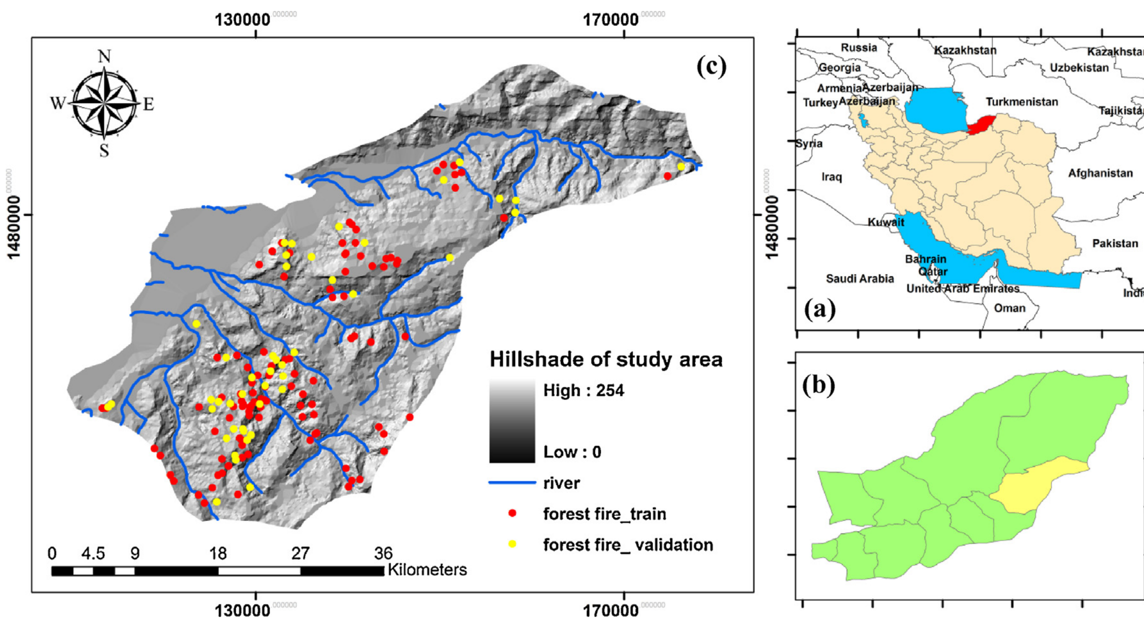

Fig. 1. Location of the study area; (a) Iran map (b) Golestan Province map; (c) Forest fire location map with hill shaded map.

et al., 2004). Various methods and algorithm for fire hazard zoning by Remote Sensing (RS) and Geospatial Information System (GIS) have been presented. Many studies have been executed to produce forest fire risk maps using GIS and RS techniques (Chuvieco and Congalton, 1989; Prosper-Laget et al., 1995; Chuvieco and Sales, 1996; Jaiswal et al., 2002; Erten et al., 2004; Wulder and Franklin, 2006; Pradhan et al., 2007; Razali, 2007; Saklani, 2008; Chuvieco et al., 2010; Adab et al., 2013; Ariapour and Shariff, 2014; Arpacia et al., 2014; Eskandari and Chuvieco, 2015; Salvatia and Ferrara, 2015; Pourghasemi, 2015; Aretano et al., 2015). Vasconcelos et al. (1995) noted that vegetation, topography, climatology, and fire history are considerable components of hazard in order to assess forest fire risk. Moreover, Pradhan et al. (2007) emphasized that normalized differential vegetation index (NDVI), soil, slope, aspect, and land use were the efficient factors to assess fire risk hazard. Janbaz Ghobadi et al. (2012) used topography, vegetation, slope, aspect, NDVI, and meteorology factors to provide forest fire risk map. Recently, new statistical techniques, namely data mining techniques, have been developed to assess fire risk (Chen et al., 1996; Fayyad et al., 1996; Zhu and Davidson, 2007; Gutiérrez et al., 2009). These techniques allow analyzing complex multivariate problems and building predictive models from large datasets. The main advantages of these models are: the estimation of the potential spatial distribution of a phenomenon, the anticipation of future changes on its allocation and the establishment of the importance of each predictor in determining the distribution of the target variable. A vast number of techniques to construct predictive models exist, such as generalized linear model (GLM), logistic multiple regression (LMR), generalized additive model (GAM), classification and regression trees (CART), and artificial neural network (ANN). The random forest (RF) is a development of classification and regression tree (CART) methods (Breiman, 2001; Shataee et al., 2011). Additionally, the boosted regression tree (BRT) is a combination of statistical and machine learning techniques and an extension of CART, as promising technique utilized in ecological modeling (Aertsen et al., 2010; Shataee et al., 2011). Stojanova et al. (2006) used data mining techniques such as logistic regression, and decision trees to predict forest fires in Slovenia. Leuenberger et al. (2013) used random forest analysis to map forest fire occurrences in Swiss Alps. Woolford et al. (2009) used generalized additive model to assess a spatio-temporal model for forest fires caused by people in a portion of boreal forest in northeastern Ontario, Canada.

High risk of wildfire events in forested areas exists in Iran. About seven percent of Iran area is covered by forests/and wooded lands. According to FAO (food and agriculture organization) database on forest/other wooded land fires in Iran (Movaghati et al., 2008), the number of fires per year equals 130 and the average burnt area per year is 5400 ha (Allard, 2003; Movaghati et al., 2008). However, fires are not largely monitored and enough detection facilities are not available. The development of various models based on simulation and mathematics in developing countries such as Iran should be considered, because traditional methods are now slowly given up and other ancient philosophies are less fit to the current world situation (Zeki and Keles, 2005; Ghomi Motazeh et al., 2013). Forest fire in Golestan Province in Iran, is still one of the most natural hazard problems. According to the last decade studies, 9068 ha of the forests have been burnt by fire (Janbaz Ghobadi et al., 2012). Thus, the main purpose of this study is the comparison of BRT, GAM, and RF statistical and decisions tree based regression models for forest fire modeling. In particular, specific aims are: (i) to identify and evaluate the importance of the general indicators involved; (ii) to develop a model capable of predicting forest fire location using BRT, GAM, and RF data mining models in the Minudasht Township of Golestan Province, Iran; (iii) to implement the model into a geographical information system to offer a data mining methodology for forest fire susceptibility mapping in Iran. These maps could be used for early warning, fire suppression resource planning, and allocation works for designing strategies to prevent forest fire or to employ specific measures for controlling occurred forest fire. Furthermore, these maps could provide guidance on how changes in land use or climate will influence the distribution and density of forest fires, especially in Iran County.

2. Study area

The study area, as shown in Fig. 1, lies in the eastern part of Golestan Province, in northern Iran. The geographical location of the study area is between latitudes 37°00′27″ to 37°27′53″ N, and longitudes 55°14′00″ to 56°00′39″ E. It covers an area about 1531 km^2. The elevation of the study area ranges of 100–2500 m above sea level. The climate of Minudasht is moderate and

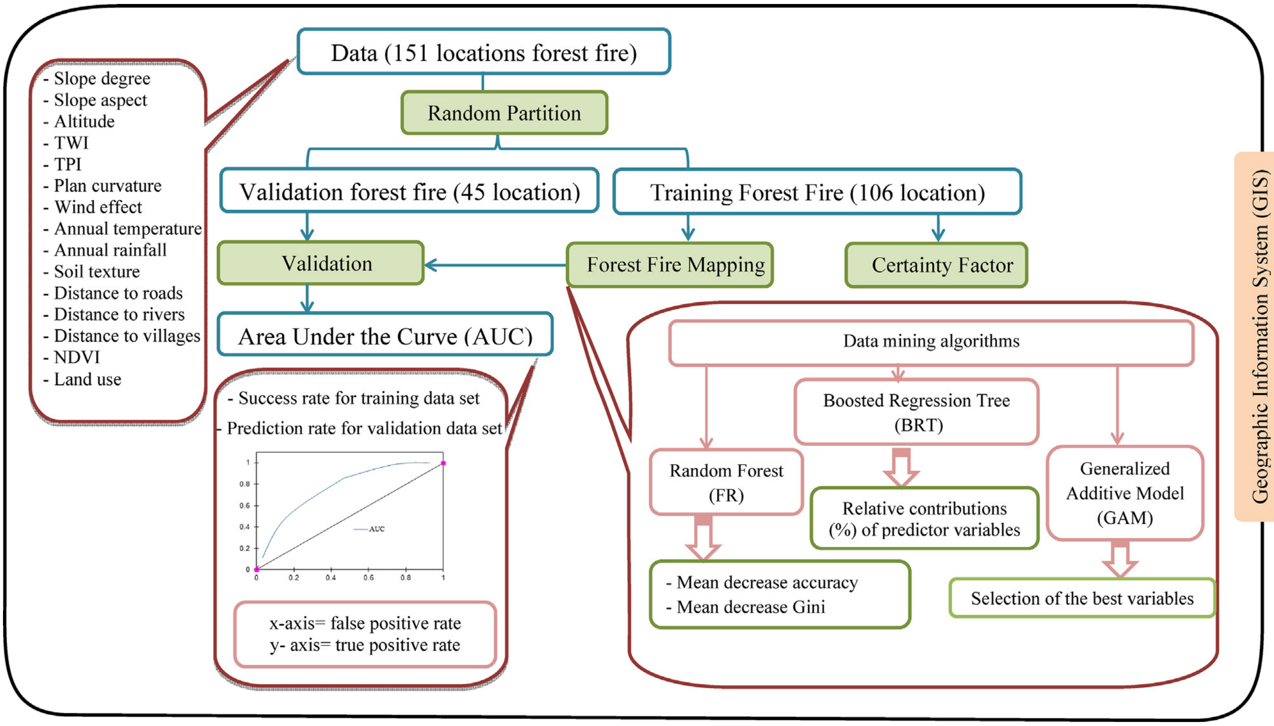

Fig. 2. Flow chart of used methodology in forest fire mapping

mountainous at heights, while in plains it is semi-humid. The mean annual precipitation within the study area ranges from 138 to 335 mm (Shadman Roodposhti et al., 2012). Based on Iranian Meteorological Organization, maximum and minimum temperature was reported as 40 and 5° above and below zero, respectively. The study area consisted of four type soil texture, in which the silt-clay-loams and clay-loam textures are higher than other soil textures in the area. According to the Statistical Center of Iran, the county's population was 135,634 at the 2011 census. The agriculture is the major economic activity of the region. Additionally, a part of Golestan National Park is located within the county, and it has dense rainforests.

3. Methodology

Fig. 2 shows the used methodology in the current study, as a flowchart. This figure shows the factors, and the processes applied in the analysis. The first part of this section concerns the data collection. The following regards the application of BRT, GAM, and RF methods, while the last section deals with the validation of three models and the selection of the best model for the study area.

3.1. Data collection

Generally, in any study data collection and construction of an effective factors database are the most important part of the process. At first, fire occurrences and locations were derived from MODIS (Moderate-Resolution Imaging Spectero Radiometer) satellite images, historical records, national reports, and field checking (Fig. 1). Of 151 forest fire locations, seventy percent (70%) were used in forest fire susceptibility mapping and the remaining thirty percent (30%) were used for validation. In order to produce the forest fire map (FFM), an evaluation of several forest fire-related factors with the forest fire occurrences is necessary (Brown and Davis, 1973; Artsybashev, 1983; Chuvieco and Congalton, 1989; Jaiswal et al., 2002; Erten et al., 2004; Renard et al., 2012; Adab et al., 2013). In this research, fifteen effective factors were considered for the occurred fires in the Minudasht County. These are slope degree, slope aspect, elevation, topographic wetness index (TWI), topographic position index (TPI), plan curvature, wind effect, annual temperature and rainfall, soil texture, distance to roads, rivers and villages, NDVI, and land use (details are shown in Table 1).

Topography data are one of the major factors included in any fire hazard rating system. In the literature, the impacts of slope degree, slope aspect, and elevation in fire behavior have been widely reported (Brown and Davis, 1973; Artsybashev, 1983; Chuvieco and Congalton, 1989; Erten et al., 2004; Renard et al., 2012; Adab et al., 2013). Slope degree, slope aspect, elevation, TWI, TPI, and plan curvature were produced using a digital elevation model (DEM) in 30 m resolution.

The slope degree map was prepared of above DEM and classified as: (1) 0–5°, (2) 5°–15°, (3) 15°–30°, and (4) >30°. Nine classes of slope aspects are observed in the study area such as: flat, north, north-east, east, south-east, south, south-west, west, and north-west. The elevation map was also prepared using DEM and categorized into five classes such as: <500 m, 500–1000 m, 1000–1500 m, 1500–2000 m, and >2000 m, respectively. Another topographic factor is TWI that is proposed by Moore et al. (1991) according to Eq. (1):

$$\mathrm{TWI} = \ln\left(\frac{\alpha}{\tan\beta}\right) \tag{1}$$

where, α is the cumulative up slope area draining through a point and tan β is the slope angle at the point.

Another factor that was used for classification of the landscape into morphological categories is TPI. It shows the difference in elevation between a focal cell and all cells in the neighborhood (Jenness, 2000). In the case of plan curvature, negative curvatures illustrate concave, zero curvature represent flat, whereas, positive curvatures are known as convex.

In the current research, the wind effect was provided in SAGA-GIS 2.8 and sorted into three classes such as (0.75–0.95), (0.95–1.14), and (>1.14). Wind effect map was created based on three input parameters such as DEM in grid format, wind

Table 1
Spatial relationship between each forest fire conditioning factors and forest fire by Certainty Factor model.

Factor	Class	No. of pixel	% pixels	No. of forest fire	% forest fire	PP_b	PP_d	Final weight
Slope degree	<5°	544,326	31.99	16	15.09	0.00003	0.000062	−0.53
	5–15°	400,683	23.55	25	23.58	0.00006	0.000062	0.001
	15–30°	621,420	36.53	49	46.23	0.00008	0.000062	0.21
	>30°	134,856	7.93	16	15.09	0.00012	0.000062	0.47
Slope aspect	Flat	316,740	18.62	12	11.32	0.00004	0.000062	−0.39
	North	209,679	12.32	15	14.15	0.00007	0.000062	0.13
	Northeast	139,591	8.21	9	8.49	0.00006	0.000062	0.03
	East	92,104	5.41	14	13.21	0.00015	0.000062	0.59
	Southeast	100,549	5.91	5	4.72	0.00005	0.000062	−0.20
	South	134,018	7.88	7	6.60	0.00005	0.000062	−0.16
	Southwest	143,326	8.42	13	12.26	0.00009	0.000062	0.31
	West	195,660	11.50	15	14.15	0.00008	0.000062	0.19
	Northwest	369,618	21.73	16	15.09	0.00004	0.000062	−0.31
Altitude (m)	<500	464,718	27.32	6	5.66	0.00001	0.000062	−0.80
	500–1000	534,642	31.43	44	41.51	0.00008	0.000062	0.24
	1000–1500	521,829	30.67	43	40.57	0.00008	0.000062	0.24
	1500–2000	169,287	9.95	11	10.38	0.00006	0.000062	0.04
	>2000	10,809	0.64	2	1.89	0.00019	0.000062	0.66
TPI	Canyons	563,157	33.10	51	48.11	0.00009	0.000062	0.31
	Gentle slopes	364,788	21.44	6	5.66	0.00002	0.000062	−0.74
	Steep slopes	214,049	12.58	13	12.26	0.00006	0.000062	−0.03
	Ridges	559,291	32.87	36	33.96	0.00006	0.000062	0.03
Soil	Silty clay loam and clay loam	1,344,693	79.04	101	95.28	0.00008	0.000062	0.17
	Silty clay loam and sandy clay loam	190,574	11.20	0	0	0	0.000062	−1
	Silty loam and silty clay loam	120,193	7.06	2	1.89	0.00002	0.000062	−0.73
	Silty loam and sandy clay loam	45,825	2.69	3	2.83	0.00007	0.000062	0.05
NDVI	<(−0.001)	453	0.03	0	0	0	0.000062	−1
	(0.00–0.001)	302	0.02	0	0	0	0.000062	−1
	(0.05–0.00)	1527	0.09	0	0	0	0.000062	−1
	(0.05–0.1)	7280	0.43	1	0.94	0.00014	0.000062	0.55
	(0.1–0.5)	1,005,442	59.10	28	26.42	0.00003	0.000062	−0.55
	>(0.5)	686,281	40.34	77	72.64	0.00011	0.000062	0.44
Distance to villages (m)	<1200	526,144	30.93	15	14.15	0.00003	0.000062	−0.54
	1200–2200	538,333	31.64	35	33.02	0.00007	0.000062	0.04
	2200–3200	274,825	16.15	33	31.13	0.00012	0.000062	0.48
	>3200	361,983	21.28	23	21.70	0.00006	0.000062	0.02
Distance to rivers (m)	<150	120,096	7.06	2	1.89	0.00002	0.000062	−0.73
	150–300	105,345	6.19	5	4.72	0.00005	0.000062	−0.24
	300–450	103,077	6.06	10	9.43	0.0001	0.000062	0.36
	450–600	99,918	5.87	3	2.83	0.00003	0.000062	−0.52
	>600	1,272,849	74.82	86	81.13	0.00007	0.000062	0.08
Distance to roads (m)	<150	373,278	21.94	5	4.72	0.00001	0.000062	−0.79
	150–300	242,583	14.26	9	8.49	0.00004	0.000062	−0.40
	300–450	180,898	10.63	8	7.55	0.00004	0.000062	−0.29
	450–600	140,598	8.26	8	7.55	0.00006	0.000062	−0.09
	>600	763,928	44.90	76	71.70	0.0001	0.000062	0.37
TWI	<9	229,533	13.49	19	17.92	0.00008	0.000062	0.25
	9–13	905,137	53.20	64	60.38	0.00007	0.000062	0.12
	>13	566,615	33.31	23	21.70	0.00004	0.000062	−0.35
Plan curvature (100/m)	Concave	557,685	32.78	46	43.40	0.00008	0.000062	0.24
	Flat	554,058	32.57	21	19.81	0.00004	0.000062	−0.39
	Convex	589,542	34.65	39	36.79	0.00007	0.000062	0.06
Wind effect	(0.75–0.95)	721,245	42.39	41	38.70	0.00006	0.000062	−0.09
	(0.95–1.14)	629,205	36.98	30	28.30	0.00005	0.000062	−0.23
	>(1.14)	350,835	20.62	35	33.02	0.0001	0.000062	0.38
Land use	IR	199,564	11.73	0	0	0	0.000062	−1
	DF	1,012,790	59.53	105	99.07	0.0001	0.000062	0.40
	LF	22,381	1.32	1	0.94	0.00004	0.000062	−0.28
	IRMF	95,828	5.63	0	0	0	0.000062	−1
	RF	347,307	20.41	0	0	0	0.000062	−1
	GR	267	0.02	0	0	0	0.000062	−1
	MR	1345	0.08	0	0	0	0.000062	−1
	MF	12,649	0.74	0	0	0	0.000062	−1
	WS	634	0.04	0	0	0	0.000062	−1
	U	8520	0.50	0	0	0	0.000062	−1

Table 1 (*Continued*)

Factor	Class	No. of pixel	% pixels	No. of forest fire	% forest fire	PP_b	PP_d	Final weight
Annual temperature (°C)	<15	40,017	2.35	1	0.94	0.00002	0.000062	−0.60
	15–16	393,766	23.15	13	12.26	0.00003	0.000062	−0.47
	16–17	647,991	38.09	70	66.04	0.00011	0.000062	0.42
	17–18	485,269	28.52	22	20.75	0.00005	0.000062	−0.27
	>18	134,242	7.89	0	0	0	0.000062	−1
Annual rainfall (mm)	<500	89,912	5.28	1	0.94	0.00001	0.000062	−0.82
	500–600	207,900	12.22	4	3.77	0.00002	0.000062	−0.69
	600–700	832,444	48.93	48	45.28	0.00006	0.000062	−0.07
	>700	571,029	33.56	53	50	0.00009	0.000062	0.33

direction (degree) and wind speed (m/s) in SAGA GIS (http://saga.sourcearchive.com).

The annul temperature map was sorted as follows <15, 15–16, 16–17, 17–18 and >18 °C. Annual rainfall map was prepared based on interpolation of measured precipitations in six rain-gages stations recorded in the region (Tangrah, Galikesh, Golitapeh, Lazore, Minudasht, and Pasposhteh) and was sorted into four classes <500 mm, 500–600 mm, 600–700 mm and >700 mm.

Additionally, there are four main types of soil type, they are silt-clay-loam and clay-loam, silt-clay-loam and sandy-clay-loam, silt-loam and silt-clay-loam, and silt-loam and sandy-clay-loam in the study area. Using topographic database in the study area, the distance to roads and rivers was provided. The roads and rivers buffers were calculated in 150 m intervals. Human activity, (especially in summer the presence of camps near riversides and springs) can affect the occurrence of forest fires. So, the increase of the concentration of people along rivers is considered as one of the serious threatening factors (Rajabi et al., 2013). In the study area, the distance to village map was also calculated in 1000 m intervals in ArcGIS 9.3. For the assessment of vegetation cover, the normalized difference vegetation index (NDVI) was used, which is the most commonly index used for assessing live fuel moisture content (Chuvieco, 2003). For calculating NDVI, the Landsat ETM$^+$ imagery was used by path 162 and row 34 obtained on 13 November 2010 and divided into six classes.

Finally, the land use map was created using Landsat-7 images. Ten land use classes were specified as follows: irrigation farming (IF), dense forest (DF), sparse forest (SF), irrigated and rained mixed farming (IRMF), rainfed farming (RF), good range (GR), moderate range (MR), moderate forest (MF), woodlands and shrubbery (WS), and urban (residential) (U). With a supervised classification and maximum likelihood algorithm, this map was created for the study area (Kappa coefficient = 88.7%). To apply BRT, GAM, and RF data mining models, all the mentioned forest fire inducing factors were converted to a raster grid with 30 m × 30 m pixel size.

3.2. Certainty Factor (CF)

In this research, the CF model was applied to illustrate the spatial relationship between distributions of fire occurrences with conditioning factors. The Certainty Factor (CF) method is one of the possible proposed Favorability Functions (FF) to apply the problem of combination of different data layers and the heterogeneity and uncertainty of the input data (Chung and Fabbri, 2003; Binaghi et al., 1998; Pourghasemi et al., 2013). The CF model was originally applied by Shortliffe and Buchanan (1975) and later modified by Heckerman (1986) according to Eq. 2:

$$\mathrm{CF} = \begin{cases} pp_b - pp_d/pp_{b(1-pp_d)} & \text{if} \quad pp_b \geq pp_d \\ pp_b - pp_d/pp_{d(1-pp_b)} & \text{if} \quad pp_b < pp_d \end{cases} \tag{2}$$

where, pp_b is the conditional probability of having a number of forest fire events occurring in category a, and pp_d is the prior probability of having the total number of forest fire events occurring in the study area. The range of variation of the CF is [−1.1] (Kanungo et al., 2011). The CF values are calculated for all condition factors by overlaying the forest fires and different layers according to above equation.

3.3. Boosted regression tree (BRT)

BRT is one of the several techniques that aim to increase the performance of a single model by fitting many models and combining them for prediction (Elith et al., 2008). BRT include important advantages of tree-based methods, handling different types of predictor variables and accommodating missing data. They have no need for prior data transformation or elimination of outliers, and they can fit complex nonlinear relationships, and automatically address interaction effects between predictors. The BRT combines the powers of two algorithms: regression trees and boosting (Elith et al., 2008). A regression tree is a piece-wise linear estimate of a regression function, which is built by the recursive partitioning of the data and the sample space (Loh, 2002; Carty, 2011). Modern decision trees are expressed statistically by Breiman et al. (1984) and Hastie et al. (2001), and were developed for ecological applications (De'ath and Fabricius, 2000; Elith et al., 2008).

Boosting is a technique used to improve the predictive performance of regression trees. Boosting is similar to model averaging, where the results of several competing models are merged. The boosting uses a forward, step-wise procedure, where tree models are fitted iteratively to a subset of the training data. There are two important parameters in determination of the number of trees required for optimal prediction (Elith et al., 2008). They are: learning rate and tree complexity. The learning rate, determines the contribution of each tree to the growing model, and the tree complexity controls whether interactions are fitted: a tree complexity of 1 (single decision stump; two terminal nodes) fits an additive model, a tree complexity of two fits a model with up to two-way interactions, and so on. In the current study, BRT model was fitted in R3.0.2 statistical software (R Development Core Team, 2006) using gbm (Generalized Boosted Regression Models) package (Ridgeway, 2006; Elith et al., 2008).

3.4. Generalized additive model (GAM)

Generalized additive model (GAM) is a semi-parametric regression model (Hastie and Tibshirani, 1990; Hastie, 1992; Maggini et al., 2006). A response curve in GAM is estimated with smooth functions, allowing a wide range of response curves to be fitted (Yee and Mitchell, 1991; Maggini et al., 2006). The model allows for rather flexible specification of the dependence of the response on the covariates, rather than detailed parametric relationships. This flexibility and convenience come at the cost of two new theoretical problems. It is necessary both to represent the smooth functions in some way and to choose how smooth they should be (Wood, 2006). The model fit of the GAM can be simply interpretable, unlike most machine learning algorithms (Brenning, 2008; Goetz et al., 2011; Petschko et al., 2014). Moreover, the GAM is more suitable for representing the nonlinear response to changing site conditions than the GLM (Goetz et al., 2011; Petschko et al., 2014). In general, the model has a structure something like:

$$g(\mu_i) = K_i^{*}\theta + f_1(k_{1i}) + f_2(k_{2i}) + f_3(k_{3i}, k_{4i}) + \cdots \quad (3)$$

where

$\mu_i = E(Y_i)$ and $Y_i \sim$ some exponential family distribution

Y_i is a response variable, K_i^{*} is a row of the model matrix for any strictly parametric model components, θ is the corresponding parameter vector, and the f_j are smooth functions of the covariates, k_n. In this study, GAM method was run in generalized regression analysis and spatial prediction (GRASP) package (Lehmann et al., 2002). For fitting the model, a combined backward and forward stepwise variable selection in R 2.0.7 based on Akaike's Information Criterion (AIC, Akaike, 1974) was used. By comparing the resulting AIC for each model fitted with different variable selections, the GAM decides on the model with the best combination of variables (Petschko et al., 2012).

3.5. Random forest (RF)

Random forests (RFs) are very powerful and flexible ensemble classifiers based upon decision trees, first developed by Breiman (2001) (Breiman, 2001; Catani et al., 2013; Micheletti et al., 2014).

The algorithm applies random binary trees which use a subset of the observations through bootstrapping techniques: from the original data set a random choice of the training data is sampled and used to build the model, the data not included are referred to as "out-of-bag" (OOB) (Breiman, 2001; Catani et al., 2013).

The algorithm estimates the importance of a variable by looking at how much the prediction error goes up when OOB data for that variable is permuted while all others are left unchanged (Liaw and Wiener, 2002; Catani et al., 2013). This capability can be fruitfully applied to study the relative importance of the different explanatory variables, a quite important but often neglected aspect of FFM (forest fire mapping). In the R statistical package implementation of RF, the model output is a membership probability to one of the two possible classes "forest fire" and "no-forest fire". Random forests need two parameters to be tuned by the user: (1) the number of trees T, (2) the number of variables m to be stochastically chosen from the available set of features. It is suggested (Breiman, 2001; Micheletti et al., 2014) to pick a large number of trees and the square root of the dimensionality of the input space for m (Micheletti et al., 2014). So, in this study, the number of trees in RF has been fixed to 1000 after a preliminary analysis and the number m of variables sampled at each node has been selected to be 3. No calibration set is needed to tune the parameters (Micheletti et al., 2014). As well as, two types of error were calculated in this model such as mean decrease in accuracy and mean decrease in node impurity (mean decrease Gini). This different importance measure can be used for ranking variables and for variable selection (Calle and Urrea, 2010).

4. Results

The results of spatial relationship between forest fire and conditioning factors using Certainty Factor model is shown in Table 1. In Table 1, slope degree classes showed that 15–30° and >30° classes have higher CF weight. In the case of slope aspect, most of the forest fire occurred in east and south-west facing. In the case of altitude, both 500–1000 m and 100–1500 m classes have CF values of 0.24 and class >2000 has higher CF value (0.66). Relation between TPI and forest fire probability showed that canyons classes have the highest value of CF (0.31). Investigation of soil conditions showed that class of silty clay loam and clay loam has higher value of CF (0.17). In the case of NDVI, it can be seen that the class of (−0.05)–(0.1) has CF value of 0.55, indicating that the probability of occurrence of forest fire in this NDVI class is very high. In the case of distance to villages, distances between 2200 and 3200 m have weight (CF) of 0.48. Assessment of distance to rivers showed that distance of 300–450 m has high correlation with forest fire occurrence. In the case distance to roads, distances of higher >600 had weight (CF) of 0.37. For topographic wetness index, class <9 has the higher CF value. Assessment of plan curvature revealed that, concave areas have a higher CF value than convex areas. The results of CF for wind effect showed that the class ">1.14" had the highest CF (CF = 0.38). In the case of land use, higher CF value was registered for dense forest (0.40). Considering the case of the relationship between forest fire probability and annual temperature, the CF is 0.42 for 16–17 °C class which show abundant of forest fire in this class. Finally, the results of CF for annual rainfall showed that the class ">700 mm" had the highest CF (CF = 0.33).

Table 2
Summary of the relative contributions (%) of predictor variables for a boosted regression tree model.

Predictor	Relative influence (%)
Rain	22.71
Temperature	10.07
Slope	9.32
Road	9.20
Land use	8.97
NDVI	7.56
Village	6.05
DEM	5.91
River	4.73
TWI	4.41
Curvature	3.16
TPI	3.12
Wind Effect	2.96
Aspect	1.81
Soil	0.02

4.1. Application of BRT

In the BRT model, where learning rate = 0.005, tree complexity = 5, and bag fraction = 0.5, the optimal number of trees was reached/selected at 550. The BRT final model included 58.60% of the mean total deviance (1 − mean residual deviance/mean total deviance = 1 − (0.68/1.39) = 0.23)) (Abeare, 2009). An index of relative influence was calculated summing the contribution of each variable, which is equivalent to summing the branch length for each variable in the regression tree (Abeare, 2009). For the main effects BRT model fitted here, the five most influential variables were annual rainfall (22.71%), annual temperature (10.027%), slope degree (9.32%), distance to roads (9.20%) and land use (8.97) (Table 2).

4.2. Application of GAM

The results of the contribution of different factors to the final GAM model are presented in Table 3 and Fig. 3. Inspection of Table 3 and Fig. 3 showed that only seven of fifteen conditioning factors, including plan curvature, land use, annual rainfall, distance to roads, annual temperature, TWI, and distance to village were selected in the final model. In contrast, other conditioning factors were excluded during the analysis. The percentage of importance or contribution of selected factors showed that the most effective factors are land use (42.69%), annual rainfall (18.66%), plan curvature (13.23%), and annual temperature (12.93%); whereas distance to

Table 3
Model contribution of conditioning factors in GAM model.

Factors	Plan curvature	Land use	Annual rainfall	Distance to roads	Annual temperature	TWI	Distance to village
% Important	13.23	42.69	18.66	2.62	12.93	3.75	6.12

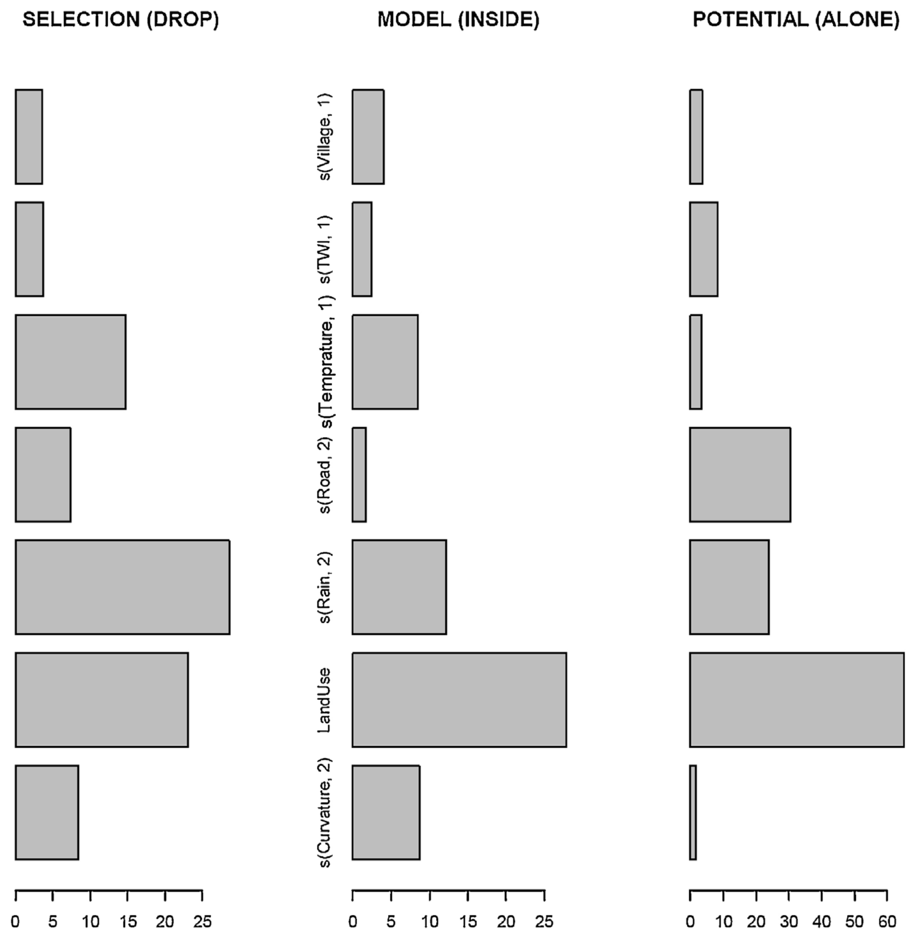

Fig. 3. Cross validation of the GAM model.

roads (2.62%), TWI (3.75%), and distance to villages (6.12%) had the lowest contribution. In addition, other model performance parameters of GAM model are reported in Table 4. The results indicated that the final model has a $r^2 = 0.44$, $r^2_{adjusted} = 0.39$, RMSE = 0.38, rRMSE = 75.16, and AIC = −177.48. By the way, one of the central parts of the interpretation of GAM models was the description of the predictor's partial response curves (Lehmann et al., 2002) (Fig. 4). In this figure, the solid lines are the predicted value of the dependent variable as a function of the x axis. The dotted lines are plus-or-minus two standard errors, and the small lines along the x axis are the "rug", showing the location of the sample plots. The *y*-axis is in the linear units, which in this case is logit and extend to both positive and negative values (Fig. 4). So, according to Fig. 4, forest fire is characterized by a strong positive response to annul rainfall, plan curvature, and distance to village, and a negative influence to TWI and annual temperature factors. Subsequently, the stepwise selection of statistically significant conditioning factors for forest fire susceptibility mapping (FFSM) using GAM resulted in the following model equation (Eq. (4)):

$$\mathrm{FFSM_{GAM}} = s(\text{Plan curvature}, 2) + \text{Landuse} + s(\text{Annual rainfall}, 2) + s(\text{Distance to roads}, 2) + s(\text{annual temperature}, 1) + s(\text{TWI}, 1) + s(\text{Distance to villages}, 1) \quad (4)$$

where, s is the spline smoother, 2 and 1 are the degrees of freedom for the spline smoother.

4.3. Application of RF

The OOB predictions are presented in Fig. 5 and Table 5 (confusion matrix). The OOB suggests that when the resulting model is applied to new observations, the answer will be in error 25% of the time, while 75% will be accurate, which is a reasonably good

Table 4
Model performance parameters in GAM model.

Parameters	r^2	$r^2_{adjusted}$	RMSE	rRMSE	AIC
Value	0.44	0.39	0.38	75.16	−177.48

Table 5
Confusion matrix from RF model (0 = non-forest fire, 1 = forest fire).

	0	1	Class error
0	58	17	0.23
1	20	53	0.27

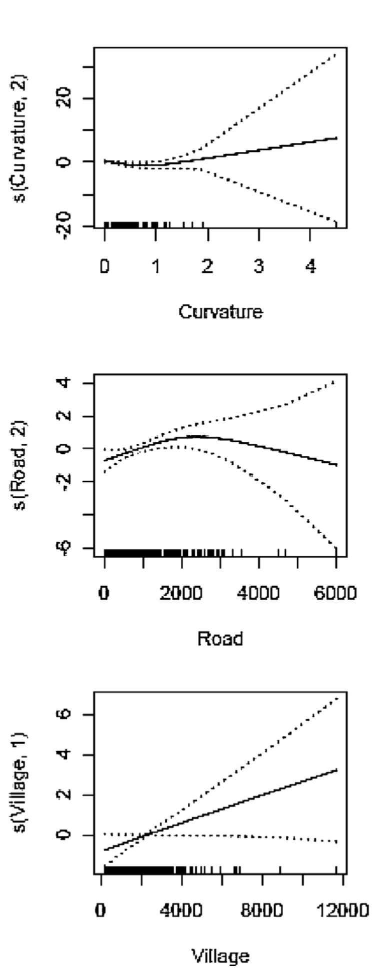

Fig. 4. Partial response curves for forest fire factors of the GAM model.

model. Overall measure of accuracy is then followed by a confusion matrix that records the disagreement between that final model's predictions and the actual outcomes of the training observations. The actual observations are the rows of this Table, whilst the columns corresponded to what the model predicts for observations, and the cell counts the number of observations in each variable (Williams, 2011). The model predicts 1 and the observation was 0 for 20 observations. The finding showed that the model and the training dataset agree that it won't be forest fire for 58 observations. They agree that it will be forest fire for 53 observations. However, there are 20 forest fires for which the model predicts that it is not forest fire. Similarly, the model predicts that it will be forest fire for 17 observations when, in fact, it is not forest fire. Results from variable selection random forest are showed in Table 6. Based on Table 6, the higher values indicate that the variable is relatively more important (Williams, 2011). The accuracy measure (mean decrease) then lists annual rainfall (31.06), land use (24.64), altitude (DEM) (14.03), and distance to roads (12.30), as the most important factors. We also notice that annual rainfall (9.94), distance to roads (6.12) and altitude (5.99) has higher importance according to the Gini measure than with an accuracy measures.

Table 6
Relative influence of effective conditioning factors in RF model (0 = non-forest fire, 1 = forest fire).

Factor	0	1	Mean decrease accuracy	Mean decrease Gini
Slope degree	3.15	9.01	8.59	4.63
Slope-aspect	−1.96	2.14	0.34	2.10
Altitude	2.94	15.92	**14.03**	**5.99**
TPI	3.82	6.28	7.07	1.83
Soil NDVI	1.71	2.91	3.34	0.51
NDVI	3.80	9.75	9.59	5.94
Distance to villages (m)	1.84	5.01	4.98	4.64
Distance to rivers (m)	−3.55	7.21	3.42	4.40
Distance to roads (m)	3.39	13.35	**12.30**	**6.12**
TWI	−2.10	4.18	2.05	4.04
Plan curvature (100/m)	2.81	3.39	4.35	2.41
Wind effect	−1.74	3.63	1.41	3.57
Land use	13.51	24.18	**24.64**	4.74
Annual temperature (°C)	−4.43	14.66	9.43	5.90
Annual rainfall (mm)	11.99	31.18	**31.06**	**9.94**

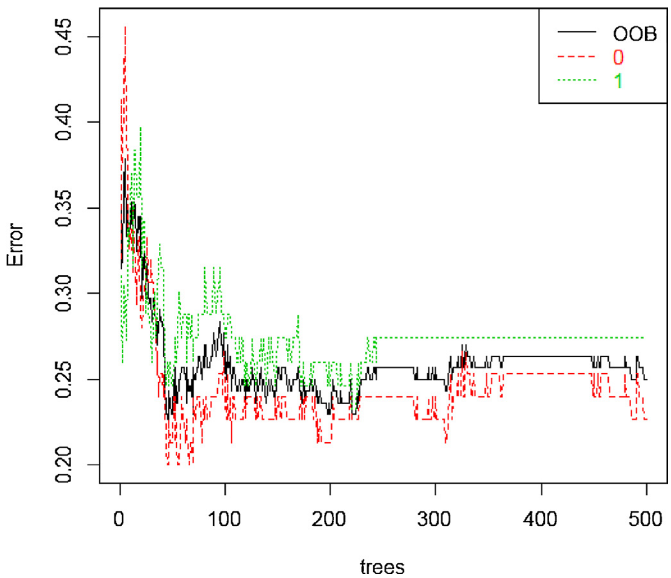

Fig. 5. The error rate of the overall RF model (OOB: out of bag, 0: absent forest fire and 1: present forest fire).

Table 7
The distribution of the forest fire values and areas with respect to the forest fire occurrence potential zones.

Forest fire mapping	Boosted regression tree Area (%)	Random forest Area (%)	Generalized additive model Area (%)
Low	5.43	27.83	2.22
Moderate	24.90	32.68	46.20
High	50.71	26.41	25.47
Very High	18.96	13.07	26.12

4.4. Forest fire susceptibility mapping

Fig. 6(a–c) display forest fire susceptibility maps created using BRT, RF, and GAM models for the Minudasht Township. The obtained pixel values from BRT, RF, and GAM model were then classified based on natural break classification scheme (Ozdemir, 2011; Pourghasemi et al., 2012a,b, 2013; Mohammady et al., 2012a) into low, moderate, high and very high potential groups. The forest fire map achieved from the BRT model, which covered 18.95% of the total area, was designated to be a very high FFM class (Table 7). On the other hand, the areas related to low, moderate and high FFM zones are 5.43%, 24.90%, and 50.71%, respectively (Fig. 6a and

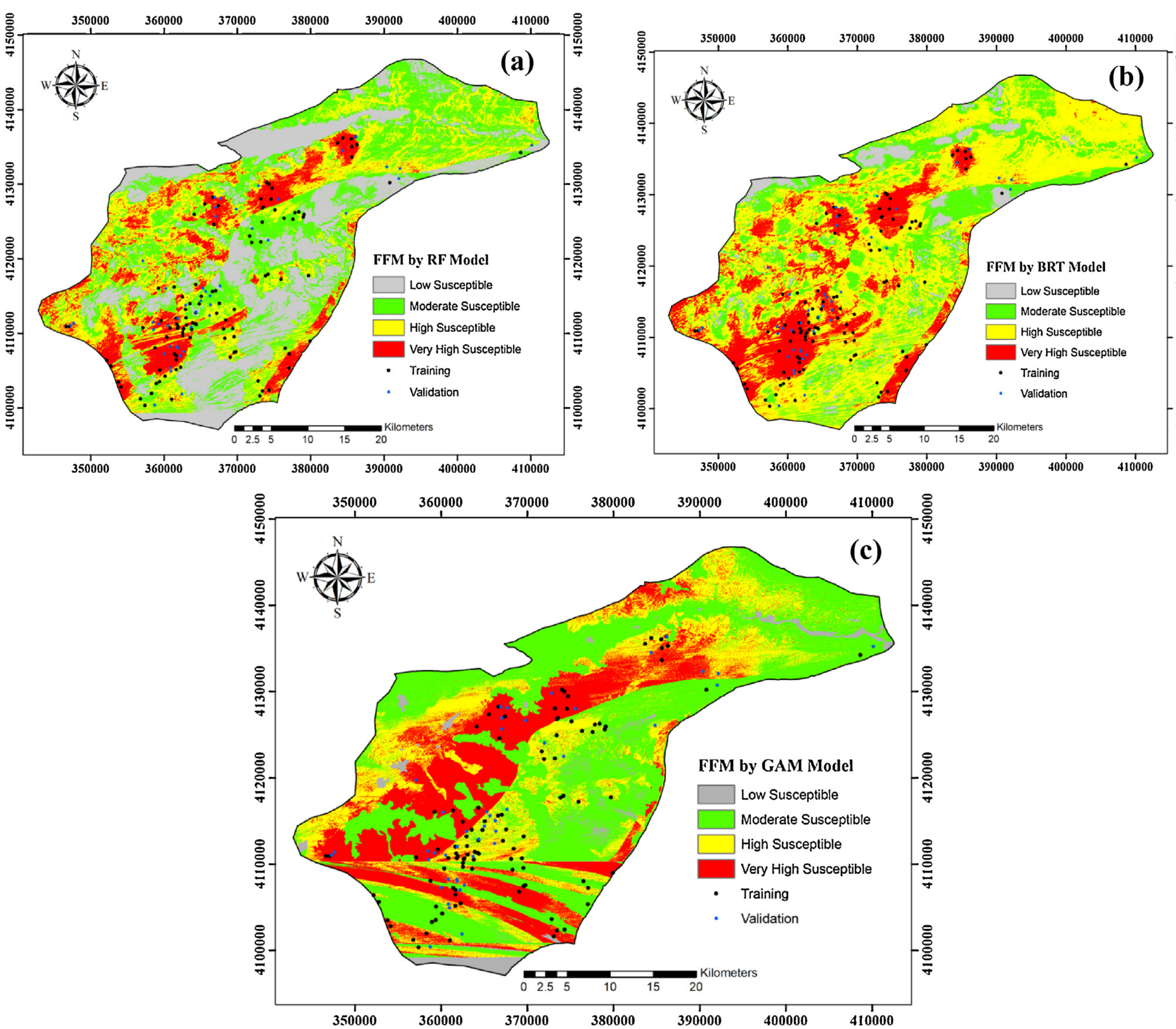

Fig. 6. Forest fire susceptibility map produced by (a) BRT model, (b) RF model, (c) GAM model.

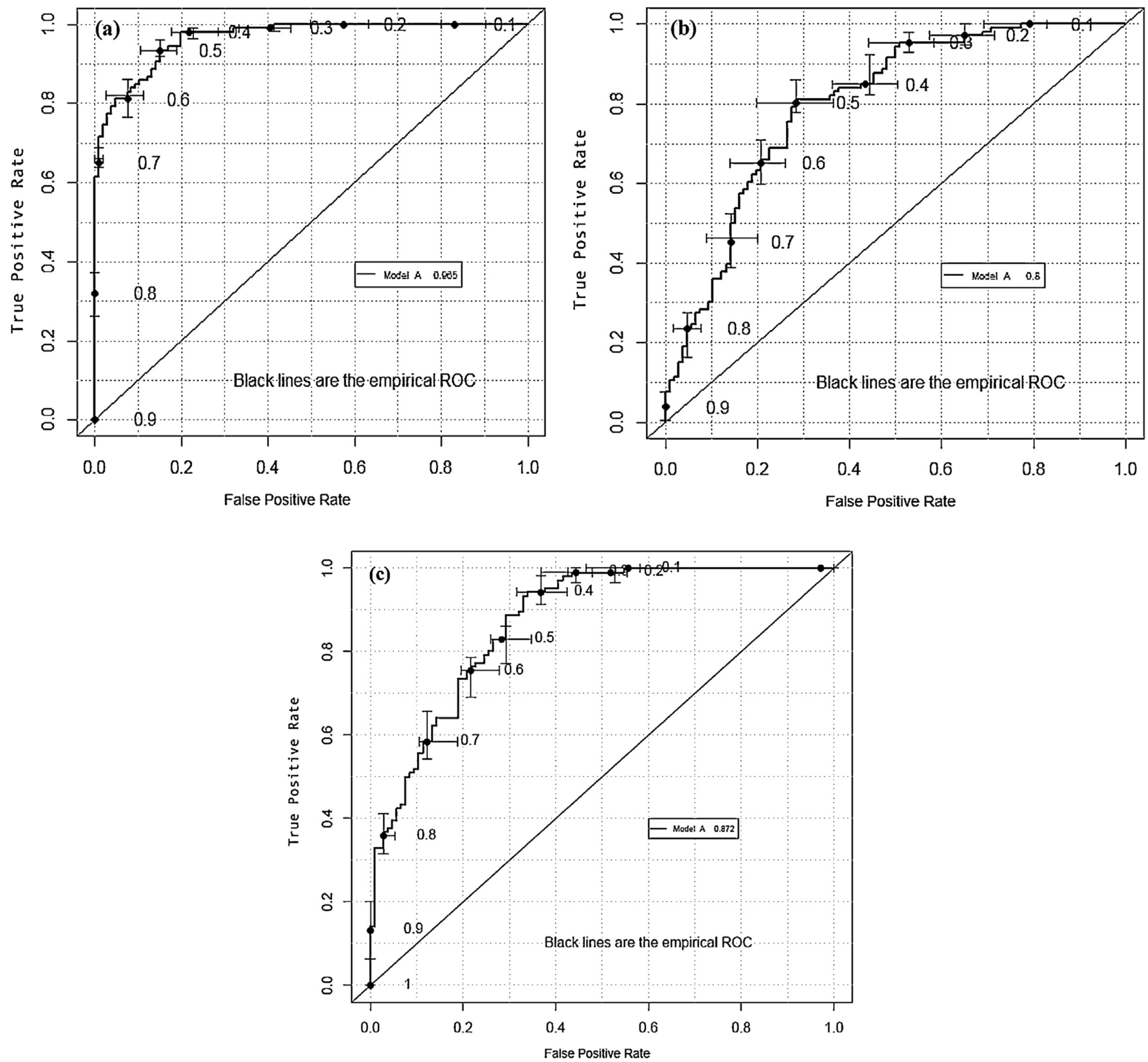

Fig. 7. Success rate curve for the forest fire susceptibility map by BRT (a), RF (b), and GAM (c) models.

Table 7). In the forest fire map based on RF model, 27.83% and 32.68% of the study areas were classified as 'low' and 'moderate' risk, whereas 26.41% and 13.07% of the areas were classified as 'high' and 'very high' risk, respectively (Fig. 6b and Table 7). The forest fire map achieved from the GAM model, which covered 26.12% of the total area, was designated to be a very high FFM class (Fig. 6c and Table 7). On the other hand, the areas related to low, moderate, and high FFM zones are 2.22%, 46.20%, and 25.47%, respectively (Table 7).

4.4.1. *Validation of the forest fire susceptibility models*

Validation is a key step in the development of susceptibility and determination of its quality (Chung and Fabbri, 2003). Additionally, the critical strategy in prediction modeling is the validation of predicted results, so that the results can supply a meaningful interpretation with respect to forest fire susceptibility (Pourghasemi, 2015). The quality of forest fire susceptibility model is usually estimated using independent information that is not available for model building. To apply validation, success rate and prediction rate curves have been used and the area under the curve (AUC) was calculated according to the existing forest fire locations with the three forest fire susceptibility maps.

Fig. 7(a–c) illustrates the success-rate curves of the three forest fire susceptibility maps in this study. It could be seen that BRT model has the highest area under the curve (AUC) values (0. 97) (Fig. 7a) than RF (0.80) (Fig. 7b), and the GAM (0.87) models (Fig. 7c). Also, the results of the prediction rate curves are shown in Fig. 8(a–c). It shows that, BRT model (Fig. 8a) has relatively higher prediction performance than the RF model, meanwhile is a little lower than the GAM. These graphs indicate that in the forest fire susceptibility map produced using BRT, the AUC was 0.8084 with a prediction accuracy of 80.84%. In the forest fire susceptibility map created by RF model, the AUC was 0.7279 and the prediction accuracy was 72.79% (Fig. 8b). Furthermore, the AUC for GAM model was 0.8770, with a prediction accuracy of 87.70% (Fig. 8c).

5. Discussion

Data mining refers to a set of computer-based tools, which permit exploratory data analysis to reveal patterns and relationships in databases (Sweeney et al., 2007). BRT, GAM, and RF have been used in several studies to investigate spatial relationship between environmental predictors and forest fire occurrence (Stojanova et al., 2006; Özkana et al., 2008; Woolford et al., 2009; Leuenberger et al., 2013; Parisien and Moritz, 2009; Bar Massada et al., 2013; Renard et al., 2012; De Angelis et al., 2015). The purpose of this study was the comparison of these statistical and decisions tree based regression models for forest fire mapping in the Minudasht Township,

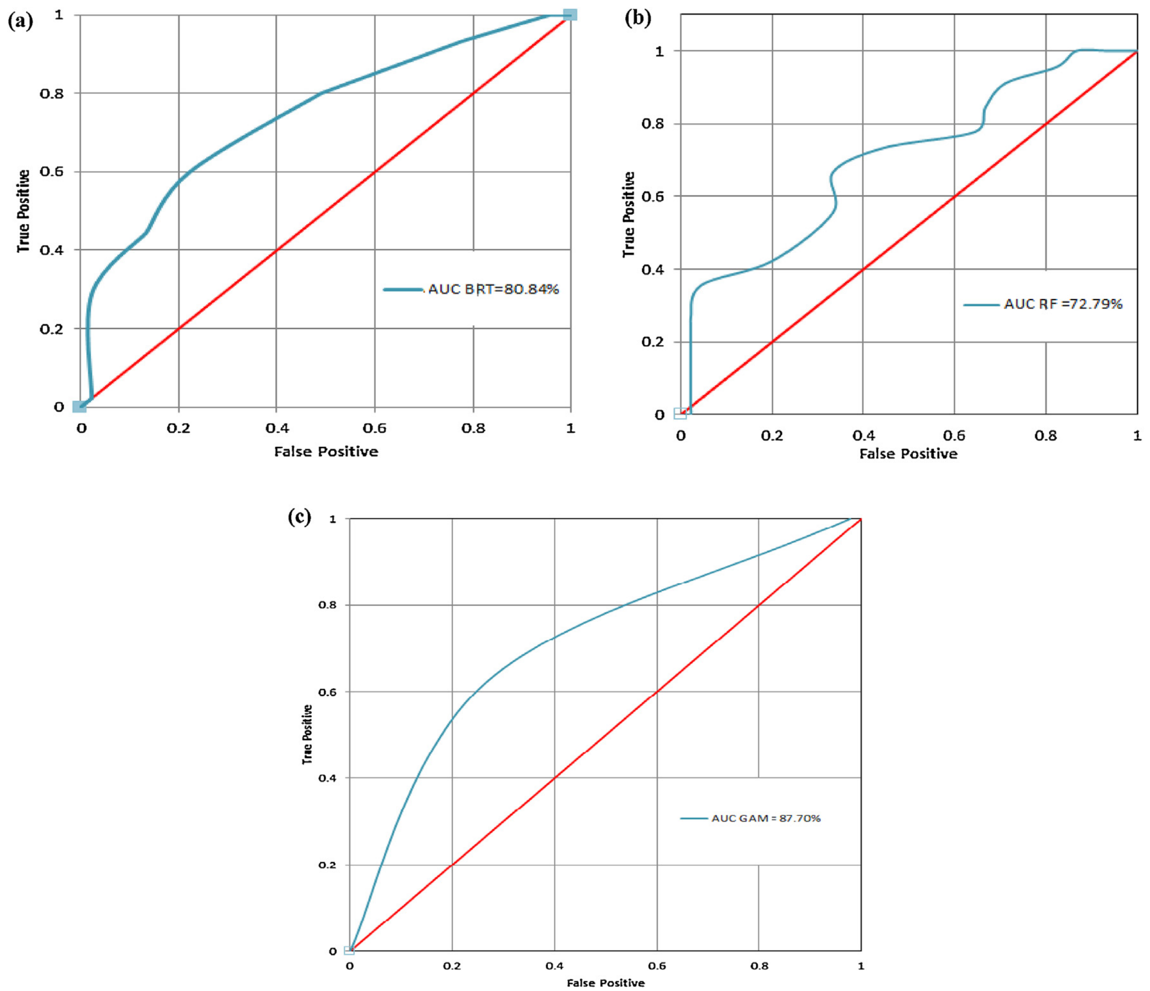

Fig. 8. Prediction rate curve for the forest fire susceptibility map by BRT (a), RF (b), and GAM (c) models.

Iran. The good and high performance of the decision tree based regression models, and their implementations are related to suitable determination of tree model options such as the number of trees, the number of predictors in each node for producing good response. Regulating these parameters in the decision tree models can play significant rules (Shataee et al., 2011). Additionally, statistical modeling of earth surface processes and landforms can be approached in various ways (e.g., prediction or explanation, univariate or multivariate analysis, and spatial or non-spatial modeling) and the utilized techniques vary depending on the focus of the study (e.g., geostatistics, least square regression, and machine learning) (Hjort and Luoto, 2013). In the BRT implementations, to control the best model in terms of sub samples of training data in a bagging bootstrapping sampling, the 70% (106 locations) forest fire occurrence was for training and 30% (45 locations) for test. As well as, in RF and GAM performance, the proportions were the same for training and testing sample data.

Although previous studies have noted superior performance for the RF method (Vorpahl et al., 2012; Oliveira et al., 2012), the results from this study showed that the RF method had the moderate performance in comparison with BRT and GAM models in forest fire susceptibility mapping. Based on results achieved, implementation of BRT had the slightly (8.05 percent (80.84–72.79)) more accurate estimation in terms of AUC than RF. This superiority is due to the BRT mix strengths of two algorithms, including regression trees, models that relate a response to their predictors by recursive binary split and boosting, an adaptive method for combining many simple models to give improved predictive performance (Elith et al., 2008; Shataee et al., 2011). The RF and BRT are similar in using regression trees to estimate a dependent variable and averaging them as final prediction, but the use of a number of boosting repetition and adaptive method and regularization of linear models as shrinkage rate to fit many models in BRT can be superior if compared to RF. Both shrinkage rate and boosting iterations control prediction risk on the training data, model complexity and avoid over fitting through choosing the optimal tree, number of iterations and regularization of the model in which things don't find in the RF algorithm (Shataee et al., 2011). The main difference between BRT and RF methods is that the boosted regression trees do not rely on bootstrap samples or randomized variable selection (Jun, 2013). The RF technique has several benefits with respect to other, more used, multivariate regression or classification methods. First, it does not need assumptions on the distribution of explanatory variables; secondly, it allows for the mixed use of categorical and numerical variables without recurring to the use of indicator (or dummy) variables and, third, it is capable of considering interactions and nonlinearities between variables. These are big advantages that confine the generation of outliers, especially when working with terrain variables with a high frequency of missing data and an intrinsic uncertainty in the assignment to the correct class also in surveyed areas (see e.g. the case that a correctly defining the type of soil for a given areal extent for which only few point locations have been directly surveyed) (Catani et al., 2013). A further advantage, even more so for our study, is the ability of RF tree bagger (RFtb) models to offer information on the statistical weight of each single variable on the overall result. Additionally, in the present study GAM method appears to be the best compromise between model stability and performance among the three methods tested. Hjort and Luoto (2013) expressed GAM model has rather high flexibility, model interpretability, usable in explanation when compared with BRT model. GAMs are beneficial in exploratory analysis or when analysts have weak a priori ideas as to the functional form relating explanatory variables to response variables. GAMs are particularly beneficial to study the shape of the response function. GAMs are more complicated to fit and require greater judgment, and it is possible to over-fit features in the data when compared with linear regression model. Over-fitted models include too many predictors, are extremely complex, and may begin to fit random noise in the data. Thus, the predictive abilities of over-fitted models are often poor, specifically if the models are extrapolated to new data or areas. In the end, even with spatially independent evaluation data, the prediction potential of GAMs is generally higher when compared with parametric techniques (Hjort and Luoto, 2013).

6. Conclusion

The paper studied the application of three machine learning/data mining methods for forest fire susceptibility mapping from a set of topographical, metrological, and geological features. The forest fire occurrence of Minudasht Township, Golestan Province, Iran, was used to prepare and test the difference between the mentioned machine learning models. The BRT, GAM, and RF were utilized to discriminate between absence or presence of forest fire, with performances peaking at AUC of 0.8084, 0.8770, and 0.7279, respectively. The machine learning algorithms were used to perform feature selection in order to reveal the variables which contribute most in determining the spatial distribution of forest fire. Our finding from BRT, GAM, and RF highlighted that annual rainfall, slope degree, distance to roads, land use and annual temperature were the more effective factors in forest fire occurrence. This study concluded that GAM model could be more useful in forest fire occurrence and mapping in comparison with BRT and RF models. Additionally, these methods indicate the most important features to be selected. As a final conclusion, results of this study can be applied to early warning, fire suppression resource planning and allocation of works. The results obtained from this study provide a considerable contribution to the forest fire literature. The used models in this study can be further improved using other forest type, tree composition, and canopy cover percentage factors. In total, we couldn't use the same variables in different regions because forest fire in each part of earth has own characteristics. By the way, the mentioned models could be compared with other data mining models including CART, MARS (Multivariate adaptive regression splines), ANN, SVM (support vector machines), and their results considered in this area and other areas.

Acknowledgement

The authors would like to thank editorial comments and two anonymous reviewers for their helpful comments on the primary version of the manuscript.

References

Abeare, Sh.M., (MSc thesis) 2009. Comparisons of Boosted Regression Tree, GLM and GAM Performance in the Standardization of Yellowfin Tune Catch-Rate Data from the Gulf of Mexico Lonline Fishery. The Department of Oceanography and Coastal Sciences, Pretoria, 94 pp.

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

Aertsen, W., Kint, V., Van Orshoven, J., Özkan, K., Muys, B., 2010. Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. Ecol. Model. 221, 1119–1130.

Aleemahmoodi Sarab, S., Feghhi, J., Goshtasb, H., 2013. Determining the main parameters affecting on forest fire using MLP neural network (forests of western Iran: Izeh). Int. J. Mol. Evol. Biodivers. 3 (4), 15–23.

Allard, G.B., 2003. Fire situation in the Islamic Republic of Iran. Int. Forest Fire News (IFFN) 28, 88–91.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automatic Control 19 (6), 716–723.

Aretano, R., Semeraro, T., Petrosillo, I., De Marco, A., Pasimeni, M.R., Zurlini, G., 2015. Mapping ecological vulnerability to fire for effective conservation management of natural protected areas. Ecol. Model. 295, 163–175.

Ariapour, A., Shariff, A.R.M., 2014. Rangeland fire risk zonation using remote sensing and geographical information system technologies in Boroujerd Rangelands, Lorestan Province, Iran. Ecopersia 2 (4), 805–818.

Arpacia, A., Malowerschnigb, B., Sassb, O., Vacika, H., 2014. Using multi variate data mining techniques for estimating fire susceptibility of Tyrolean forests. Appl. Geogr. 53, 258–270.

Artsybashev, E.S., 1983. Forest Fires and Their Control. Oxonian, New Delhi (Ist ed. in Russian, 1974).

Bar Massada, A., Syphard, A.D., Stewart, S.I., Radeloff, V.C., 2013. Wildfire ignition-distribution modelling: a comparative study in the Huron-Manistee National Forest, Michigan, USA. Int. J. Wildland Fire 22, 174–183.

Binaghi, E., Luzi, L., Madella, P., Pergalani, F., Rampini, A., 1998. Slope instability zonation: a comparison between certainty factor and Fuzzy Dempster-Shafer approaches. Nat. Hazards 17, 77–97.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth International Group, Belmont, CA, USA.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (l), 5–32.

Brenning, A., 2008. Statistical geocomputing combining R and SAGA: the example of landslide susceptibility analysis with generalized additive models. SAGA—seconds out. Hamburger Beiträge Phys. Geogr Landschaftsökol. 19, 23–32, Germany.

Brown, A.A., Davis, K.P., 1973. Forest Fire: Control and Use. McGraw-Hill, New York.

Calle, M.L., Urrea, V., 2010. Letter to the Editor: Stability of random forest importance measures. Brief Bioinf. 12 (1), 86–89.

Carty, D.M., (MSc thesis) 2011. An Analysis of Boosted Regression Trees to Predict the Strength Properties of Wood Composites. The University of Tennessee, Knoxville, 105 pp.

Catani, F., Lagomarsino, D., Segoni, S., Tofani, V., 2013. Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. Nat. Hazards Earth Syst. Sci. 13, 2815–2831.

Chung, C.F., Fabbri, A.G., 2003. Validation of spatial prediction models for landslide hazard mapping. Nat. Hazards 30, 451–472.

Chen, M.S., Han, J., Yu, P.S., 1996. Data mining: an overview from a database perspective. IEEE Trans. Knowl. Data Eng. 8 (6), 866–883.

Chen, X.C., Karpatne, A., Chamber, Y., Mithal, V., Lau, M., Karsten, S., Boriah, Sh., Steinbach, M., Kumar, V., et al., 2012. A new data mining framework for forest fire mapping. In: Conference on Intelligent Data Understanding.

Chuvieco, E., Congalton, R.G., 1989. Application of remote sensing and geographic information systems to forest fire hazard mapping. Remote Sens. Environ. 29, 147–159.

Chuvieco, E., Sales, J., 1996. Mapping the spatial distribution of forest fire danger using GIS. Int. J. Geogr. Inf. Syst. 10, 333–345.

Chuvieco, E., Aguadoa, I., Yebraa, M., 2010. Development of a framework for fire risk assessment using remote sensing and geographic information system technologies. Ecol. Model. 221, 46–58.

Chuvieco, E., 2003. Wildland fire danger estimation and mapping: the role of remote sensing data. In: Series in Remote Sensing. World Scientific, Singapore.

De Angelis, A., Ricotta, C., Conedera, M., Boris Pezzatti, G., 2015. Modelling the meteorological forest fire niche in heterogeneous pyrologic conditions. PLOS ONE, http://dx.doi.org/10.1371/journal.pone.0116875.

De'ath, G., Fabricius, K.E., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81, 3178–3192.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77, 802–813, http://dx.doi.org/10.1111/j.1365-2656.2008.01390.x.

Erten, E., Kurgun, V., Musaoglu, N., 2004. Forest fire risk zone mapping from satellite imagery and GIS: a case study. In: XXth Congress of the International Society for Photogrammetry and Remote Sensing, Istanbul, Turkey, pp. 222–230.

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

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., 1996. From data mining to knowledge discovery in databases. A.I. Mag. 17 (3), 37–53, http://dx.doi.org/10.1609/aimag.v17i3.1230.

Ghomi Motazeh, A., Farahi Ashtiani, E., Baniasadi, R., Masoumpoor Choobar, F., 2013. rating and mapping fire hazard in the hardwood Hyrcanian forests using GIS and expert choice software. For. Idea 19 (2 (46)), 141–150.

Goetz, J.N., Guthrie, R.H., Brenning, A., 2011. Integrating physical and empirical landslide susceptibility models using generalized additive models. Geomorphology 129, 376–386.

Gutiérrez, A.G., Schnabel, S., Lavado Contador, J.F., 2009. Using and comparing two nonparametric methods (CART and MARS) to model the potential distribution of gullies. Ecol. Model. 220, 3630–3637.

Hastie, T., Tibshirani, R., Friedman, J.H., 2001. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Verlag, New York.

Hastie, T.J., 1992. Generalized additive models. In: Chambers, J.M., Hastie, T.J. (Eds.), Statistical Models in S. Wadsworth & Brooks/Cole.

Hastie, T., Tibshirani, R.J., 1990. Generalized Additive Models. Chapman & Hall, London.

Herawati, H., Heru, S., Claudio, F., 2006. Forest Fires and Climate Change in Indonesia. Background Document for the Southeast Asia Kick-Off Meeting of the Project Tropical Forests and Climate Change Adaptation (TroFCCA), 5 pp., Available at: http://www.scribd.com/doc/125463814/Forest-Fire-CC.

Heckerman, D., 1986. Probabilistic interpretations for MYCIN's certainty factors. In: Kanal, L., Lemmer, J. (Eds.), Uncertainty in Artificial Intelligence. North-Holland, New York, pp. 167–196.

Hjort, J., Luoto, M., 2013. Statistical methods for geomorphic distribution modeling. In: Shroder, J., (Editor in Chief), Baas, A.C.W. (Eds.), Treatise on Geomorphology. Vol. 2: Quantitative Modeling of Geomorphology. Academic Press, San Diego, CA, pp. 59–73.

Jaiswal, R.K., Mukherjee, S., Raju, K.D., Saxena, R., 2002. Forest fire risk zone mapping from satellite imagery and GIS. Int. J. Appl. Earth Observ. Geoinform. 4, 1–10.

Janbaz Ghobadi, Gh., Gholizadeh, B., Majidi Dashliburun, O., 2012. Forest fire risk zone mapping from geographic information system in Northern Forests of Iran (Case study, Golestan province). Int. J. Agri. Crop Sci. 4 (12), 818–824.

Jenness, J.S., (MSc thesis) 2000. The Effects of Fire on Mexican Spotted Owls in Arizona and New Mexico. Northern Arizona University, 137 pp.

Jun, S.H., 2013. Boosted Regression Trees and Random Forests. Statistical Consulting Report for Michael Melnychuk, Visiting Post-Doc Fisheries Centre, UBC, 13 pp.

Kanungo, D.P., Sarkar, S., Sharma, S., 2011. Combining neural network with fuzzy, certainty factor and likelihood ratio concepts for spatial prediction of landslides. Nat. Hazards 59, 1491–1512.

Lehmann, A., Overton, J., Leathwick, Mc.C.J.R., 2002. GRASP: generalized regression analysis and spatial prediction. Ecol. Model. 157, 189–207.

Leuenberger, M., Kanevski, M., Vega Orozco, C.D., 2013. Forest fires in a random forest. Eur. Geosci. Union Gen. Assembly 15, 2013–3238.

Liaw, A., Wiener, M., 2002. Classification and regression by random forest. R News 2, 18–22.

Loh, W.Y., 2002. Regression trees with unbiased variable selection and interaction detection. Stat. Sin. 12 (2), 361–386.

Maggini, R., Lehmann, A., Zimmermann, N.E., Guisan, A., 2006. Improving generalized regression analysis for the spatial prediction of forest communities. J. Biogeogr. 33, 1729–1749.

Micheletti, N., Foresti, L., Robert, S., Leuenberger, M., Pedrazzini, A., Jaboyedoff, M., Kanevski, M., 2014. Machine learning feature selection methods for landslide susceptibility mapping. Math Geosci. 46, 33–57, http://dx.doi.org/10.1007/s11004-013-9511-0.

Mohammady, M., Pourghasemi, H.R., Pradhan, B., 2012a. Landslide susceptibility mapping at Golestan Province, Iran: a comparison between frequency ratio, Dempster-Shafer, and weights-of-evidence models. J. Asian Earth Sci. 61, 221–236.

Moore, I.D., Grayson, R.B., Ladson, A., 1991. Digital terrain modeling: a review of hydrological, geomorphological, and biological applications. Hydrol. Process. 5, 3–30.

Movaghati, S., Samadzadegan, F., Azizi, A., 2008. A comparative study of three algorithms for fire detection in Iran. Int. Arch. Photogram. Remote Sens. Spatial Inform. Sci. xxxvii, 379–389.

Naebi, H., 2003. Identifying and Combating of Forest Fires. Publication of Program Planning Promotions Office, Tehran, Iran, 29 pp.

Nasi, R., Dennis, R., Meijaard, E., Applegate, G., Moore, P., 2002. Forest fire and biological diversity. Unasylva 53 (209), 36–40.

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. For. Ecol. Manage. 275, 117–129.

Ozdemir, A., 2011. Using a binary logistic regression method and GIS for evaluating and mapping the groundwater spring potential in the Sultan Mountains (Aksehir, Turkey). J. Hydrol., http://dx.doi.org/10.1016/j.jhydrol.2011.05.015.

Özkana, C., Sunarb, F., Berberoğluc, S., Dönmezc, C., 2008. Effectiveness of boosting algorithms in forest fire classification. Int. Arch. Photogram. Remote Sens. Spat. Inf. Sci. XXXVII (Pt B7), Beijing.

Parisien, M.A., Moritz, M.A., 2009. Environmental controls on the distribution of wildfire at multiple spatial scales. Ecol. Monogr. 79 (1), 127–154.

Petschko, H., Bell, R., Glade, T., 2012. Landslide susceptibility modeling with generalized additive models facing the heterogeneity of large regions. In: Eberhardt, et al. (Eds.), Landslides and Engineered Slopes: Protecting Society Through Improved Understanding. , pp. 769–775.

Petschko, H., Brenning, A., Bell, R., Goetz, J., Glade, T., 2014. Assessing the quality of landslide susceptibility maps – case study Lower Austria. Nat. Hazards Earth Syst. Sci. 14, 95–118.

Pourghasemi, H.R., 2015. GIS-based forest fire susceptibility mapping in Iran: a comparison between evidential belief function and binary logistic regression models. Scand. J. For. Res., 1–40.

Pourghasemi, H.R., Moradi, H.R., Fatemi Aghda, S.M., Gokceoglu, C., Pradhan, B., 2013. GIS-based landslide susceptibility mapping with probabilistic likelihood ratio and spatial multi criteria evaluation models (North of Tehran, Iran). Arab. J. Geosci., http://dx.doi.org/10.1007/s12517-012-0825-x.

Pourghasemi, H.R., Pradhan, B., Gokceoglu, C., Mohammadi, M., Moradi, H.R., 2012a. a. Application of weights-of evidence and certainty factor models and their comparison in landslide susceptibility mapping at Haraz watershed, Iran. Arab. J. Geosci., http://dx.doi.org/10.1007/s12517-012-0532-7.

Pourghasemi, H.R., Mohammady, M., Pradhan, B., 2012b. Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran. Catena 97, 71–84.

Pradhan, B., Suliman, M.D.H.B., Awang, M.A.B., 2007. Forest fire susceptibility and risk mapping using remote sensing and geographical information systems (GIS). Disaster Prevent. Manage. 16 (3), 344–352.

Prosper-Laget, V., Douguedroitl, A., Guinot, J.P., 1995. Mapping the risk of forest fire occurrence using NOAA satellite information. EARseL. Adv. Remote Sens. 4 (3-XII), 30–38.

Razali, S.B.M., (MSc thesis) 2007. Forest Fire Hazard Rating Assessment in Peat Swamp Forest Using Integrated Remote Sensing and Geographical Information System. University Putra Malaysia, 52 pp.

Rajabi, M., Alesheikh, A., Chehreghan, A., Gamzeh, H., 2013. An innovative method for forest fire risk zoning map using fuzzy inference system and GIS. Int. J. Sci. Technol. Res. 2 (12), 57–64.

Renard, Q., Pelissier, R., Ramesh, B.R., Kodandapani, N., 2012. Environmental susceptibility model for predicting forest fire occurrence in the Western Ghats of India. Int. J. Wildland Fire, http://dx.doi.org/10.1071/WF10109, 15 pp.

Ridgeway, G., 2006. Generalized boosted regression models. Documentation on the R Package 'gbm', version 1.5–7, http://www.i-pensieri.com/gregr/gbm.shtml (accessed March 2008).

Salvatia, L., Ferrara, A., 2015. Validation of MEDALUS Fire Risk Index using Forest Fires Statistics through a multivariate approach. Ecol. Indic. 48, 365–369.

Saklani, P., (MSc thesis) 2008. Forest fire risk zonation, A case study Pauri Garhwal, Uttarakhand, India. International Institute for Geo-information Science and Earth Observation Enschede of the Netherlands and Indian Institute of Remote Sensing (NRSA), Dehradun, India, 71 pp.

Shadman Roodposhti, M., Rahimi, S., Jafar Beglou, M., 2012. PROMETHEE II and fuzzy AHP: an enhanced GIS-based landslide susceptibility mapping. Nat. Hazards, http://dx.doi.org/10.1007/s11069-012-0523-8.

Shataee, Sh., Weinaker, H., Babanejad, M., 2011. Plot-level forest volume estimation using airborne laser scanner and TM Data, comparison of boosting and random forest tree regression algorithms. Proc. Environ. Sci. 7, 68–73.

Shortliffe, E.H., Buchanan, B.G., 1975. A model of inexact reasoning in medicine. Math. Biosc. 23 (3), 351–379.

Stojanova, D., Panov, P., Kobler, A., Džeroski, S., Taškova, K., 2006. Learning to predict forest fires with different data mining techniques. In: Proceedings of the 9th International multi conference Information Society IS 2006, 9-3th October 2006. Jožef Stefan Institute, Ljubljana.

Sweeney, A.W., Beebea, N.W., Cooper, R.D., 2007. Analysis of environmental factors influencing the range of anopheline mosquitoes in northern Australia using a genetic algorithm and data mining methods. Ecol. Model. 203, 375–386.

UNEP, 2007. Global Environment Outlook 4. United Nations Environment Programme, Nairobi.

Vasconcelos, M.J.P., Pereira, J.M.C., Zeigler, B.P., 1995. Simulation of fire growth using discrete event hierarchical modular models. EARSeL. Adv. Remote Sens. 4 (3), 54–62.

Vorpahl, P., Elsenbeer, H., Märker, M., Schröder, B., 2012. How can statistical models help to determine driving factors of landslides? Ecol. Model. 239, 27–39.

Williams, G., 2011. Data Mining with Rattle and R (The Art of Excavating Data for Knowledge Discovery Series).

Woolford, D.G., Bellhouse, D.R., Braun, W.J., Dean, Ch.B., Martell, D.L., Sun, J., 2009. A spatio temporal model for people-caused forest fire occurrence in the Romeo Malette forest. J. Environ. Stat. 2 (1), 1–26.

Wood, S.N., 2006. Generalized Additive Models: An Introduction with R.

Wulder, M.A., Franklin, S.E., 2006. Understanding Forest Disturbance and Spatial Pattern: Remote Sensing and GIS Approaches. CRC Press, 246 pp.

Yee, T.W., Mitchell, N.D., 1991. Generalized additive models in plant ecology. J. Veg. Sci. 2, 587–602.

Zhu, X., Davidson, I., 2007. Knowledge Discovery and Data Mining: Challenges and Realities. Information Science Reference, Hershey, PA, 292 pp.

Zeki, E., Keles, S., 2005. Developing alternative wood harvesting strategies with linear programming in preparing forest management plans. Turk. J. Agric. For. 30, 67–79.