



Application of remote sensing and machine learning algorithms for forest fire mapping in a Mediterranean area

Meriam Mohajane^{a,b,1}, Romulus Costache^{c,1}, Firoozeh Karimi^d, Quoc Bao Pham^{e,*},
Ali Essahlaoui^b, Hoang Nguyen^{f,g}, Giovanni Laneve^h, Fatiha Oudija^a

^a Soil and Environment Microbiology Team, Department of Biology, Faculty of Sciences, Moulay Ismail University, Zitoune, Meknès BP 11201, Morocco

^b Geo-Engineering and Environment Laboratory, Water Sciences and Environment Engineering Team, Department of Geology, Faculty of Sciences, Moulay Ismail University, Zitoune, Meknès BP 11201, Morocco

^c Department of Civil Engineering, Transilvania University of Brasov, 5, Turnului Str, 500152 Brasov, Romania

^d Department of Geography, Environment, and Sustainability, University of North Carolina-Greensboro, Greensboro, NC 27402, USA

^e Institute of Applied Technology, Thu Dau Mot University, Binh Duong Province, Viet Nam

^f Department of Surface Mining, Mining Faculty, Hanoi University of Mining and Geology, 18 Pho Vien, Duc Thang Ward, Bac Tu Liem District, Hanoi 100000, Viet Nam

^g Center for Mining, Electro-Mechanical Research, Hanoi University of Mining and Geology, 18 Pho Vien, Duc Thang Ward, Bac Tu Liem District, Hanoi 100000, Viet Nam

^h University of Rome 'La Sapienza', Scuola di Ingegneria Aerospaziale, Via Salaria 851, 00138 Rome, Italy

ARTICLE INFO

Keywords:

Forest fire
Hybrid machine learning algorithm
Remote sensing
Mediterranean area

ABSTRACT

Forest fire disaster is currently the subject of intense research worldwide. The development of accurate strategies to prevent potential impacts and minimize the occurrence of disastrous events as much as possible requires modeling and forecasting severe conditions. In this study, we developed five new hybrid machine learning algorithms namely, Frequency Ratio-Multilayer Perceptron (FR-MLP), Frequency Ratio-Logistic Regression (FR-LR), Frequency Ratio-Classification and Regression Tree (FR-CART), Frequency Ratio-Support Vector Machine (FR-SVM), and Frequency Ratio-Random Forest (FR-RF), for mapping forest fire susceptibility in the north of Morocco. To this end, a total of 510 points of historic forest fires as the forest fire inventory map and 10 independent causal factors including elevation, slope, aspect, distance to roads, distance to residential areas, land use, normalized difference vegetation index (NDVI), rainfall, temperature, and wind speed were used. The area under the receiver operating characteristics (ROC) curves (AUC) was computed to assess the effectiveness of the models. The results of conducting proposed models indicated that RF-FR achieved the highest performance (AUC = 0.989), followed by SVM-FR (AUC = 0.959), MLP-FR (AUC = 0.858), CART-FR (AUC = 0.847), LR-FR (AUC = 0.809) in the forecasting of the forest fire. The outcome of this research as a prediction map of forest fire risk areas can provide crucial support for the management of Mediterranean forest ecosystems. Moreover, the results demonstrate that these novel developed hybrid models can increase the accuracy and performance of forest fire susceptibility studies and the approach can be applied to other areas.

1. Introduction

Forest fire disaster is considered as one of the main causes of dramatic depletion of the forest ecosystems worldwide (Venkatesh et al., 2020) due to both anthropogenic or natural processes (Sachdeva et al., 2018). Fire regimes are becoming even more pronounced in many regions due to an increasing effect of recent global warming, with increasing impacts on human well-being resources and ecosystem

function processes (Hantson et al., 2016; Li et al., 2019; Zema et al., 2020). The impacts include deterioration of soil ecology and water hydrology (Santana et al., 2014; Zema et al., 2020), and land degradation and soil erosion (Zema et al., 2020). Consequently, this disaster makes the ecosystem out of balance as an ultimate stage. Despite its dramatic impacts, forest fire plays essential roles in many forest processes, for example, forest fire influences the composition and successional stages (Amit Parashar and Sas Biswas, n.d.; Huebner et al., 2012), and it acts as

* Corresponding author.

E-mail address: phambaoquoc@tdmu.edu.vn (Q. Bao Pham).

¹ The authors contributed equally to this work

a selective factor of the traits of plants (Escudero et al., 2000; Fernández-García et al., 2019; Keeley et al., 2011; Pausas and Vallejo, 1999). In the Mediterranean region, despite their ecological roles and the services they provide, forest ecosystems have been heavily impacted by fragmentation, degradation, and deforestation (Belinchón et al., 2009; Scarascia-Mugnozza et al., 2000). Furthermore, climate change exacerbates the deterioration of these ecosystems, due to rainfall decreasing and temperature increasing (Simioni et al., 2020). These phenomena, together with anthropogenic disturbance, place the Mediterranean forest ecosystem in an alarming situation. Not only these factors but also forest fire is one of the most issues which affect the Mediterranean forest (Oliveira et al., 2018). Olivella et al. (2006) demonstrated that more than 50,000 fires burn an average of 600,000–800,000 ha of Mediterranean forest annually. Thus, in order to protect the functionalities of the forest ecosystems as well as their precious services and benefits for people's well-being, forest fire susceptibility maps remain important to fully understand and predict possible hazards that may occur in this environment. Moreover, establishing a forest fire susceptibility map as early as possible is of the utmost importance of land use planning (Bax and Francesconi, 2018). Such a map can also serve as a valuable reference to reduce vulnerability and could help to improve decision-making planning in ecological risk prevention (Hong et al., 2018).

Creating forest fire susceptibility maps requires the establishment of forest fire inventory as the target variable and some explanatory variables. Various studies used a set of data composed of climate data (rainfall, temperature, humidity, and wind), topography data (elevation, slope, and aspect), and land use data as explanatory variables (Tien Bui et al., 2017; Díaz-Avalos et al., 2001; Ganteaume et al., 2013; Krawchuk et al., 2006; Vecín-Arias et al., 2016; Verdú et al., 2012). Indeed, there is no common agreement to follow for the selection of appropriate ignition variables. The processing of explanatory data is the most challenging task in the elaboration of forest fire maps in large study areas due to various data sources and the modelling complexity (Bonham-Carter, 2014; Tien Bui et al., 2017). In an effort to process such datasets and to produce forest fire susceptibility maps, geographic information system (GIS) have proven to be useful for processing and overlaying geospatial data (El Hafyani et al., 2020; Teodoro and Duarte, 2013). Furthermore, the remote sensing approach and satellite images as a popular means of data collection play a crucial role in constructing forest fire susceptibility maps.

Several statistical bivariate and multivariate methods have been used widely for forest fire modeling such as multiple linear regression (Oliveira et al., 2012), Poisson regression (Wotton et al., 2003), and Monte Carlo simulations (Conedera et al., 2011). In recent years, an obvious trend in the use of machine learning algorithms in a wide range of natural hazard assessments disciplines such as landslide (Chen et al., 2017, 2019; Hong et al., 2016; Pham et al., 2018a), flood (Tehrany et al., 2015b), rainfall modelling (Pour et al., 2020; Pham et al., 2019), streamflow modelling (Pham et al., 2020), water quality modelling (Pham et al., 2021) and forest fire modeling (Arpaci et al., 2014), has been observed. Machine learning algorithms can handle non-linear and highly dimensional data and solve serious problems (Chen et al., 2019; Knudby et al., 2010; Recknagel, 2001). Pourtaghi et al. (2016) compared boosted regression tree (BRT), generalized additive model (GAM), and random forest (RF) for forest fire mapping of the Minudasht Township, Golestan Province, Iran. The results indicated that BRT had better accuracy than GAM and RF. Jaafari et al. (2018) applied 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). Their results indicated that the ADT classifier performed the best accuracy. However, despite the wide use of these models in different fields such as environmental and ecological science, the selection of an appropriate model is still challenging because each algorithm exhibits some drawbacks leading to implications in its outcomes (Jaafari et al., 2019b; Tien Bui et al., 2019).

Given that the use of ensemble / hybrid models built by a

combination of Machine learning algorithms and statistical methods increases the modeling performance (Tien Bui et al., 2014, 2016, 2019; Truong et al., 2018; Akay, 2021), various studies have been extensively studied this approach in related researches. For example, for flood susceptibility assessment in the Haraz watershed, Mazandaran province, Iran, Shafizadeh-Moghadam et al. (2018) applied eight ML models including artificial neural networks (ANNs), classification and regression trees (CART), flexible discriminant analysis (FDA), generalized linear model (GLM), generalized additive model (GAM), boosted regression trees (BRT), multivariate adaptive regression splines (MARS), and maximum entropy (MaxEnt), and seven statistical ensemble models including Ensemble Model committee averaging (EMca), Ensemble Model confidence interval Inferior (EMciInf), Ensemble Model confidence interval Superior (EMciSup), Ensemble Model to estimate the coefficient of variation (EMcv), Ensemble Model to estimate the mean (EMmean), Ensemble Model to estimate the median (EMmedian), and Ensemble Model based on the weighted mean (EMwmean). Their results demonstrate that BRT presents the highest performance among individual models and the ensemble model EMmedian achieved the highest performance among all models. Based on this, the authors highlighted that the ensemble approaches reduce uncertainty, improve generalization performance, and they are more stable and less sensitive models. In a study on gully erosion in Aghemam watershed, Iran, Pourghasemi et al. (2017) used different ML models including ANNs, support vector machine (SVM), maximum entropy (ME), and the ensemble ones including ANN-SVM, ANN-ME, and SVM-ME. The results showed that the ANN-SVM model had the highest performance. Costache and Tien Bui (2019) proposed six new hybrid models including, Multilayer Perception neural network-Frequency Ratio (MLP-FR), Multilayer Perception neural network-Weights of Evidence (MLP-WOE), Rotation Forest-Frequency Ratio (RF-FR), Rotation Forest-Weights of Evidence (RF-WOE), Classification and Regression Tree-Frequency Ratio (CART-FR), and Classification and Regression Tree-Weights of Evidence (CART-WOE), to predict flood in a study area in Romania. It was concluded that the used hybrid models have high prediction performance (Costache et al., 2020a) to assess flash-flood potential. They used a bivariate method based on Statistical Index (SI) and its combination with machine learning models including LR, CART, MLP, RF, and SVM. They found that the MLP-SI hybrid model achieved the highest accuracy in comparison to other models. Tehrany et al. (2019) used Logitboost ensemble-based decision tree (LEDT) and benchmarked it with SVM, RF, and Kernel logistic regression (KLR) and found that the LEDT ensemble model can achieve the highest predictive accuracy for the susceptibility of the forest fire. Nevertheless, few publications have been considered the use of the ensemble models for forest fire susceptibility modeling (Moayedi et al., 2020). In this study, five novel hybrid models by coupling the statistic approach with machine learning algorithms are used. Frequency Ratio was integrated into five machine learning methods, namely MLP, LR, CART, SVM, and RF for forecasting forest fire susceptibility. The research was conducted for the north of Morocco as the case study which is most affected by the fire at the national scale. To the best of our knowledge, no study so far has been conducted for forecasting and assessing of forest fire susceptibility in this area based on machine learning algorithms and GIS tools. The main goals of this paper are: (i) exploring the effectiveness of the hybrid models for forest fire forecasting in this study area (ii) producing spatial susceptibility maps using the proposed approaches to identify the critical areas which require an emergency intervention. Thus, the findings of this paper are expected to be a useful tool for providing a crucial direction for management of Mediterranean forest ecosystem.

2. Materials

2.1. Study area

The study area in this research is located in Tanger-Tétouan-Al

Hoceima region, North of Morocco (Fig. 1), covering Fahs-Anjra province and Tanger-Asilah prefecture with a total area of 1685 km², which has a typical Mediterranean climate with oceanic influence (“DRATT”, 2015). In this area, the temperature varies between 10 °C and 14 °C in winter and between 20 °C and 24 °C in summer, the annual precipitation can reach 635 mm, and the dry season is from May to mid-September (“DRATT”, 2015), based on the Walter-lieth climate diagram, drawn by data weather from Climate Chart application, <https://climatecharts.net/> (Harris et al., 2016) for station Tangier. The study area is characterized by different types of forest composed of mixed oak forests of evergreen and semideciduous forests (Ajbilou et al., 2006) which are characterized by an extend of invasive and flammable species diversity

(Chebli et al., 2018). This diversity is a catalyst for forest fire disaster. Generally, forests in northern Morocco have the highest level and the greatest fragility due to fires (Chebli et al., 2018; Zidane et al., 2019). According to (Chebli et al., 2018), about 1185 ha of forest ecosystem in the north part of Morocco is destroyed by fires annually (43% of the total burned forest in Morocco). Though restoration efforts led by the National Forestry Conservation Agency may offset these losses, forests have been also destroyed by various human activities, for example, forest resources are exploited for other purposes, like cannabis cultivation, collection of fuelwood, and goat grazing (Chebli et al., 2020; Stambouli et al., 2005). Also, the north of Morocco experiencing rapid population growth, with a total population of approximately 3.2 million in 2014

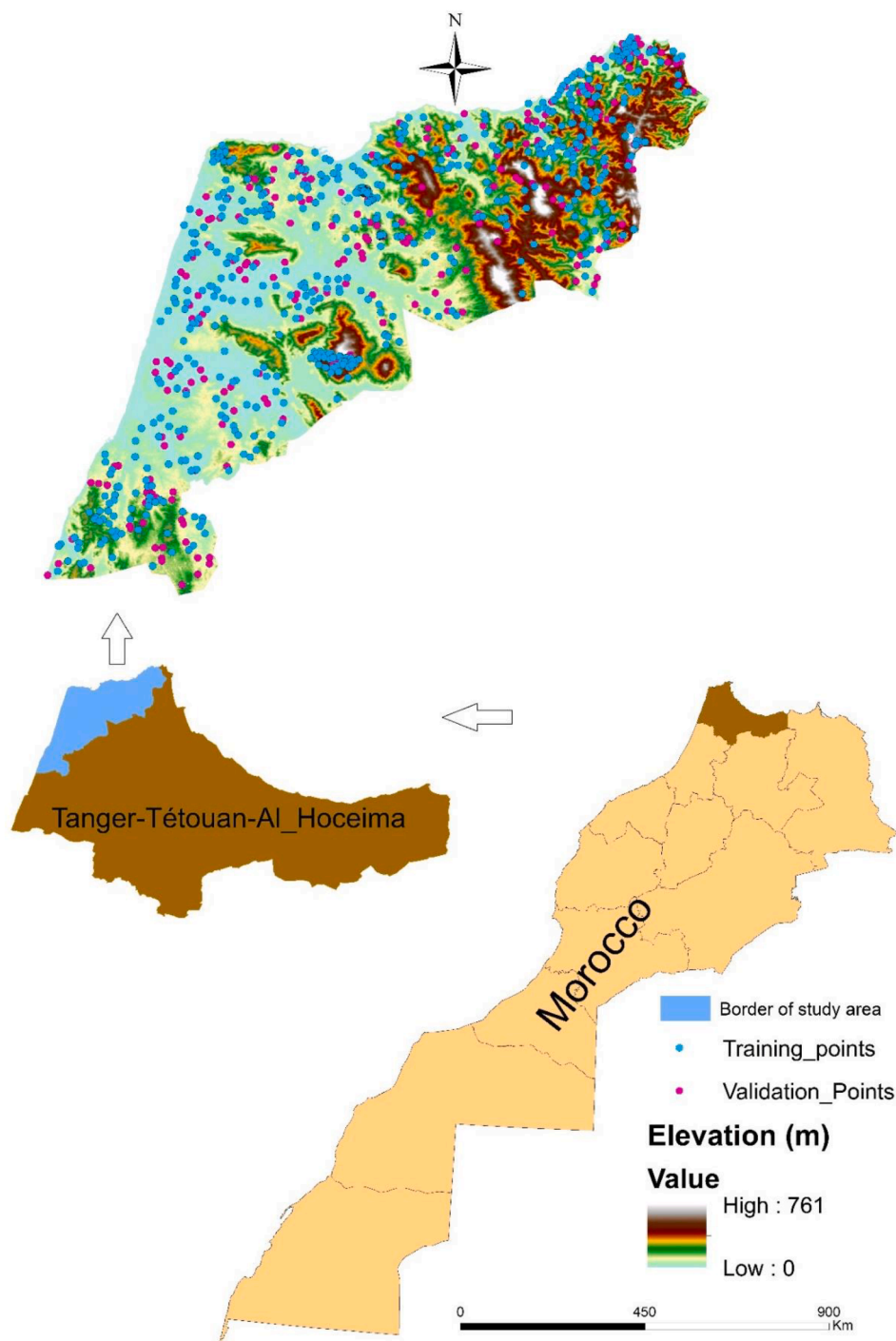


Fig. 1. Location of the study area and forest fire inventory points.

which is characterized by a highly urbanized population with an increase of 2,45% per year. Whereas the rural population increased by only 0,21% per year (HCP, 2014), Most of this population development has occurred in and around forest areas. These forests have also suffered from soil erosion coupled with a decrease of natural vegetation which has led to degradation of the forest and pastoral cover and an increase of cultivated land (Chebli et al., 2018). Therefore, researches on forest ecosystem degradation in this region are necessary to accelerate the process of the restoration project and to provide a scientific contribution to forest fire modeling.

2.2. Data

In this study, and as it is well discussed by several studies (Jaafari et al., 2018; Pourtaghi et al., 2016; Tien Bui et al., 2017), the first step to model forest fire is the establishment of the forest fire inventory map which corresponds to the past-occurred events presenting the target variable. In this case, historical data provided by (HCEFLCD, 2011) and previous reports, were used for preparing the forest fire inventory map of the study area. Moreover, data on fire occurrences can be obtained from MODIS (Moderate-Resolution Imaging Spectroradiometer) (Giglio et al., 2003). Users can use an existing freely-available hotspot data fire occurrences through the exploitation of the Fire Information for Resource Management System (FIRMS), <https://firms.modaps.eosdis.nasa.gov/>, to improve the quality of forest fire historical data. The pre-processing of the data was performed to exclude anomalous points based on the extent of fire occurrence and the presence of severe collateral damage. Moreover, forest fires mostly occurred during March, April, May, June, July, August, September, and October during the period 2005–2015, for this, the remaining months were removed from the analysis. A total number of 510 fire event locations were selected for this study. Forest fire points were randomly split into a 70% training dataset and 30% for validation. For training and testing sets the same number of non-forest fire locations were randomly selected using the random point tool in ArcGIS 10.4, across the study area. And randomly split into 70% training dataset and 30% for validation (Fig. 1). Table 1 summarizes the total number of forest fire occurred during each month for the period 2005–2015. Based on this, all explanatory variables were selected for the same period to match fire occurrence period.

Based on previous studies, the explanatory variables for forest fire modeling are composed of four groups of data including topography, anthropogenic activities, climate, and vegetation (Jaafari et al., 2018; Nami et al., 2018; Pourghasemi, 2016; Pourghasemi et al., 2016; Pourtaghi et al., 2016). These include; slope angle, aspect, elevation, distance to roads and residential, rainfall, temperature, wind speed, Land use, and NDVI (Table 2). All these variables were reclassified following the method outlined by (Tsangaratos et al., 2017).

Topography plays an important role as it can control the distribution of vegetation and wind speed as well as it has an important role in the velocity of rainfall and soil moisture (Chuvieco and Congalton, 1989; Nami et al., 2018; Schmidt et al., 2008; Tehrani et al., 2019). In this study, the Digital Elevation Model (DEM) obtained from Shuttle Radar Topographic Mission (SRTM) with a pixel size of 30 m, was used to

Table 1
Total number of forest fires that occurred during the period 2005–2015.

Month	Total number of forest fire
March	14
April	33
May	27
June	67
July	105
August	109
September	102
October	53

Table 2
Explanatory factors were used in this study.

Conditioning factor	Unit	Source
Slope	Degrees (°)	DEM 30 m from, https://earthexplorer.usgs.gov/
Elevation	Meters (m)	
Aspect	–	
Land cover	–	Two Landsat8-OLI images from, https://earthexplorer.usgs.gov/
NDVI	Ratio	
Rainfall	(mm)	ERA-Interim, https://apps.ecmwf.int/datasets
Temperature	Degree Celsius	
Wind	m/s	
Road distance	km	Road maps
Residential distance	km	Land use map of the study area

derive topography data including slope, aspect, and elevation. Slope (Fig. 2a) was calculated and divided into five groups including (1) 0–3°, (2) 3.1°–7°, (3) 7.1°–15°, (4) 15.1°–25°, and (5) 25.1°–62.2°. Up-slope areas are more affected by intensive fires whereas down slopes areas are less affected (Lentile et al., 2006). Aspect (Fig. 2b) was calculated and classified into nine classes such as flat zones, north, north-east, east, south-east, south, south-west, west, and north-west. The elevation (Fig. 2c) was composed of five classes as 0–62 m, 62.1–142 m, 142.1–243 m, 243.1–370 m, and 370.1–761 m, respectively. All topographic characteristics were calculated using Surface tool in *spatial analyst tools* available in ArcGIS 10.4 software.

Fire is more pronounced in forest ecosystems close to roads and populated areas due to the traffic accidents and the impact of human beings on the natural ecosystems uncontrolled (Tien Bui et al., 2019, 2016). The distance to the roads (Fig. 3a) was calculated and classified into five groups including (0–120 m), (120.1–240 m), (240.1–480 m), (480.1–840 m), and (840.1–8988 m). The distance to the residential area (Fig. 3b) was calculated and divided into five groups including (0–1000 m), (1000–2000 m), (2000–3000 m), (3000–4000 m), and (4000–10835 m). The *Euclidean distance* in the Spatial Analyst tool available in ArcGIS 10.4, was used to calculate the distance maps.

Land cover information is widely recognized as a proxy of fuel (Tien Bui et al., 2016). Also, the Normalized Difference Vegetation Index (NDVI), is used to monitor photosynthetic activity and provides information on vegetation biomass and phenology (Jiang et al., 2006; Mohajane et al., 2018). In this study, two satellite Landsat8-OLI (Operational Land Imager) images, covering the study area (paths/rows 201/035 and 201/036) were acquired from the United States Geological Survey (USGS) website used to extract the land cover information. These two images were mosaiced. First, we applied radiometric and atmospheric corrections for both images using ENVI 5.2 software based on Dark Object Subtraction (DOS) algorithm (Chavez, 1996), then we extracted the study area. based on Ground-based survey data and high-resolution imagery from Google Earth, a total of 250 ground truth points were collected and the Maximum Likelihood supervised classification method using ArcGIS 10.4 was applied to classify the image into four class categories: water bodies, forests, croplands, and build-up/Bare lands. The produced land cover achieved an overall accuracy of 93% (Fig. 3c).

The NDVI factor was calculated using the near-infrared (NIR) and visible red bands through the following Equation (1) (Freden et al., 1973):

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

The NDVI index ranges from – 1 to 1 with the higher values represent healthy vegetation, and the lower values indicate non-vegetative cover. In our case, the NIR and RED bands are band 5 (0.85–0.88 μm) and band 4 (0.64–0.67 μm), respectively, of the Landsat 8 OLI image. We applied an NDVI thresholding to classify NDVI image (Fig. 3d) and we generated six groups: (-1 – 0.05), (-0.05 – 0), (0 – 0.05), (0.05 – 0.1), (0.1 – 0.5), and

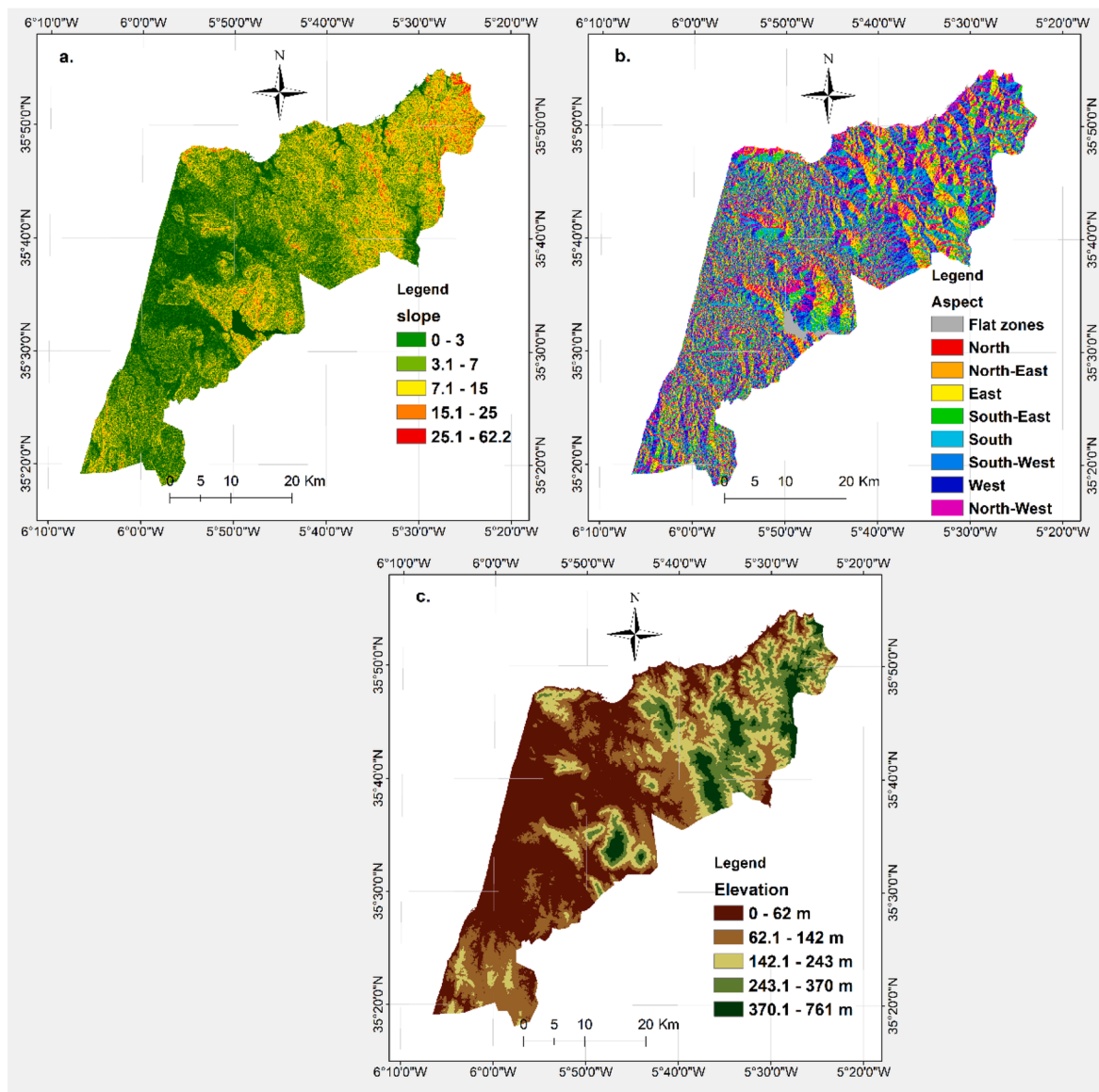


Fig. 2. Fire conditioning factors (a. Slope angle; b. Aspect; c. elevation).

(0.5 – 0.73).

Meteorological data downloaded from <https://apps.ecmwf.int/datasets>, with a spatial resolution of $0.25^\circ \times 0.25^\circ$ and resampled to 30 m spatial resolution were used in this study. The wind speed map (Fig. 4a) was classified into five classes including: (3.00–3.25 m/s), (3.25–3.5 m/s), (3.5–3.75 m/s), (3.75– 4 m/s), and (4 –4.5 m/s). The rainfall map (Fig. 4b) was created with six classes as (629–700 mm), (700–750 mm), (750–800 mm), (800–850 mm), (850–900 mm), and (900–943 mm). The Mean temperature map (Fig. 4c) was created and classified as (15.9–16.5 °C), (16.6– 17 °C), (17.1–17.5 °C), (17.6– 18 °C), (18.1– 18.5 °C).

3. Background of the models

3.1. Frequency Ratio (FR)

FR model is one of the bivariate statistical methods, it is used to describe the importance of classes of each explanatory factor on forest fire occurrence, it is used to define the ratio of the probability of forest fire occurrence to the probability of a non-occurrence for given attributes (Bonham-Carter, 2014; Lee and Talib, 2005). It has been used in different natural environmental hazard studies, for its advantages of

ease of use and understanding (Costache, 2019a; Pradhan and Lee, 2009), such as landslide mapping (Nsengiyumva et al., 2019; Zhou et al., 2016), forest fire mapping (Hong et al., 2017) and groundwater potential mapping. The FR was calculate using the following Equation (2) (Lee and Talib, 2005):

$$FR = \frac{\sum_{i=1}^m \frac{Np(LXi)}{Np(Xj)}}{\sum_{j=1}^m \frac{Np(Xj)}{Np(LXi)}} \quad (2)$$

where: FR is the Frequency Ratio of class i of factor j , $Np(LXi)$ is the number of fire locations within class i of factor variable X , $Np(Xj)$ is the number of pixels within factor variable X_j , m is the number of classes in the factor variable X_i , and n is the number of factors in the study area.

3.2. Multi-Layer perceptron neural network (MLP NN)

Artificial neural networks (ANN), inspired by biological neurons, are among artificial intelligence methods that recognize convoluted patterns in data (Skapura, 1996). ANN consists of various nonlinear computational elements which are operating in parallel and organized

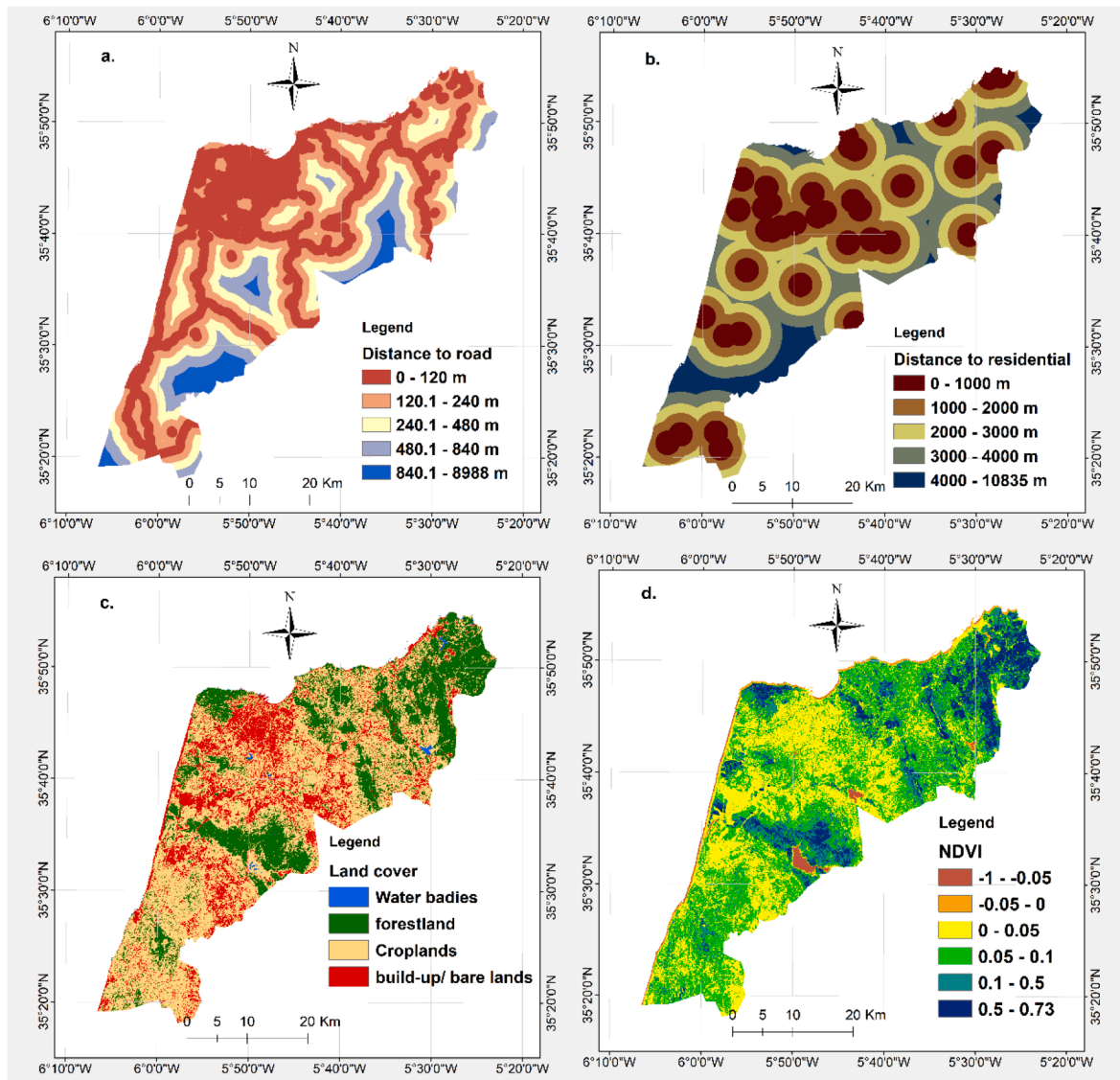


Fig. 3. Fire conditioning factors (a. Distance to roads; b. Distance to residential; c. Land use; d. NDVI).

in patterns (Lippmann, 1987). ANN has traditionally been designed as a structure composed of an input layer and an output layer called a multilayer perceptron (Rosenblatt, 1958) which are connected by one or more hidden layers. This structure can identify non-linear relationships (Basheer and Hajmeer, 2000; Rumelhart et al., 1985). MLP NN is trained in three phases including the feedforward, the backpropagation, and the adjustment of the weights (Basheer and Hajmeer, 2000; Lippmann, 1987). In the feedforward phase, the data broadcasts to each of the hidden layers with multiple weighted summations, then by reaching the output layer an activation function computes the output value. In backpropagation phase, the initial weights are selected randomly, then the predicted output for a given observation and the expected output for that observation are compared. When all the observations are given to the network, the total calculated error is distributed among the nodes in the network, then the weights are updated according to a generalized delta rule. The feed-forward and back propagating process is repeated iteratively until we achieve a minimal error.

In this study, we used Adam optimizer (Adaptive Moment Optimization) optimizer, which is developed by (Kingma and Ba, 2017), the first and second moments are obtained using the following equations:

$$X = v_1 n_{i-1} + (1 - v_1) f \quad (3)$$

$$Y = v_2 u_{i-1} + (1 - v_2) f^2 \quad (4)$$

These two moments X and Y can be corrected using the following equations:

$$X_b = \frac{X}{1 - v_1} \quad (5)$$

$$Y_b = \frac{Y}{1 - v_2} \quad (6)$$

The update the weights U is then expressed as follows:

$$U_{i-U_{i-1}} = \delta - \frac{X_b}{\sqrt{Y_b} + \sigma} \quad (7)$$

With δ is the initial learning rate, by default is set to 0.001.

The main advantages of MLP NN are, its ability to process faster a large volume of data, pattern recognition, and it is distinguished by its ability of learning and parallel processing, data fusion, and handling noise (Ni, 2008).

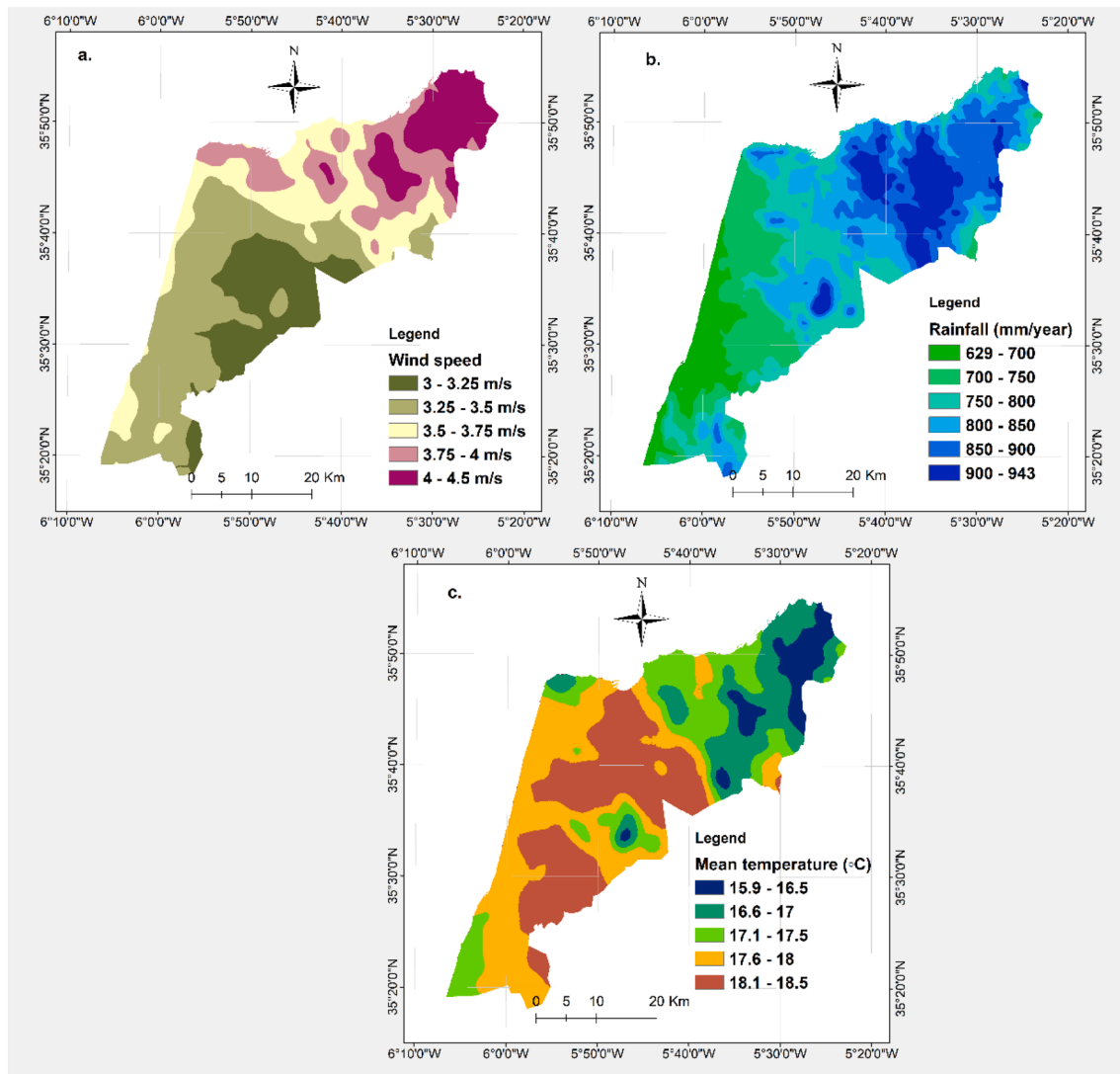


Fig. 4. Fire conditioning factors (a. Wind speed; b. Rainfall; c. Mean temperature;).

3.3. Logistic regression (LR)

LR, a statistical model for solving binary problems, was developed by (Cox, 1958). In LR, the target variable is a categorical binary value and the output of LR is likelihood values, which specify the probability of the occurrence of a certain class based on the feature values. Mathematically, a LR is a special case of linear regression and is based on the central mathematical concept of logit, the natural logarithm of an odds ratio (Hosmer and Lemeshow, 2000). The plot of the simplest case of linear regression for one feature and one dichotomous outcome is two parallel lines, each corresponding to a value of the dichotomous outcome. It is a linear plot in the middle and curved at the ends, by computing the mean of these two outcomes. It is not easy to describe this S-shaped plot with a linear equation as the ends are not linear and the errors are not normally distributed or constant across the entire range of data (Peng et al., 2001). The key to solving this problem is LR as it applies the logit transformation to the target variable and predicts the logit of outcomes from features (Peng et al., 2002). The conditional mean of the dichotomous outcome in LR is based on the binomial distribution which is the only assumption of LR and denotes that there is the same probability across the range of feature values. LR describes the effect of features by determining the coefficients of them, handles both continuous and categorical variables but just binary categorical ones for the dependent variable (Menard, 2000).

3.4. Classification and regression tree (CART)

CART method, a binary univariate decision tree algorithm and non-parametric supervised classification and prediction method, was introduced by (Breiman et al., 1984). CART classifies the dataset by constructing a tree. The procedure is conducted by sorting the whole training dataset down from the root to some leaf nodes in a recursive manner. At each stage of the process, the root node is the ideal informative feature, and the division of the training dataset into subsets is conducted based on the splitting rules applied to a single feature. The splitting rules in a CART decision tree algorithm are the Gini Index and towing criteria (Karimi et al., 2019a). The splitting procedure maximizes the homogeneity of subsets by determining the proper features and their corresponding thresholds. CART is applied to the training dataset repetitively to complete the tree which happens when three parameters of the stopping rules are fulfilled, including 1-the minimum number of records in a leaf node, 2- the minimum number of records in a parent node, and 3-the maximum number of splits at each node (Quinlan, 1993). The stopping rules make the CART model robust and efficient with the least complexity. Several decision rules can be utilized for classification and prediction studies are determined by the constructed tree. CART is able to identify the most important features (Delen et al., 2013). Based on the importance of the features, CART divides a complex dataset into smaller subsets hence it creates simple solutions which can

be understood and interpreted easily (Quinlan, 1993). Moreover, CART handles non-linear relations between the target variable and explanatory variables, process large datasets of variables including continuous and categorical, without any prior data preparation (Friedl and Brodley, 1997), makes no statistical assumptions, analyzes data in different measurement scales without the need to normalization (Qin et al., 2009), handles missing data and outliers (Delen et al., 2013) and makes good visualizations of the relationships between the features (Karimi et al., 2019a).

3.5. Random forest (RF)

RF (Breiman, 2001), a non-parametric supervised method applied for both classification and prediction (Costache et al., 2020a; Nefeslioglu et al., 2010), was developed by (Breiman, 2001). RF is composed of a combination of decision trees (Breiman, 2001). Each decision tree individually votes for assigning the most frequent class to the input data and the majority vote of the trees determines the class prediction (Breiman, 2001). In RF models trees are growing from different training data subsets to increase the diversity (Breiman, 1996) as a result, greater classifier stability is achieved (Breiman, 2001). The data subsets which are not included for the training of a tree are called out-of-bag which can be classified by that tree to assess the accuracy and the performance and calculate an internal unbiased estimate of the generalization error based on the number of trees (Breiman, 2001). Increasing the number of trees in the model, lead to the generalization error converges, and avoiding overfitting the data by RF. For building RF model, a suitable attribute selection measure is required to maximize dissimilarity measures between classes. The most frequent methods for selecting the attributes at each node are gain-ratio (Quinlan, 1993), Gini Index (Breiman et al., 1984), and Chi-square. When a tree is growing, each node is divided using the best split of a random subset of input features. This makes each tree to be less strong, but at the same time decreases the correlation between trees, as the result the generalization error is reduced and the model's accuracy is increased (Breiman, 2001). In addition, RF has the ability to run on large datasets efficiently, it can process massive amounts of input features, it gives estimates of the importance of the features used for modeling, it replaces missing data, and detects outliers through computing proximities between pairs of cases, and it is approximately robust to outliers and noise (Breiman, 2001; Rodriguez-Galiano et al., 2012).

3.6. Support vector Machine (SVM)

SVM developed by (Vapnik, 1963) at AT&T Bell Laboratories (Vapnik, 1995), is used for both classification and regression tasks and then, it was enhanced by (Boser et al., 1992) using an idea based on statistical learning theory. SVM is a binary classifier developed to find a linear hyperplane that separates two classes optimally, but it can be promoted to an n-class classifier (Belousov, 2002). SVM finds an optimal hyperplane for classification by projecting the input data into the higher-dimensional Hilbert space (Yang et al., 2008). SVM minimizes an upper bound of generalization error by widening the distance of the hyperplanes separating the two classes (Huang et al., 2002; Karimi et al., 2019b). This action guarantees a low generalization error, independent of data distribution (Huang et al., 2002; Karimi et al., 2019b). In SVM models a penalty parameter beyond the margin of the hyperplanes is incorporated to consider misclassification errors (Vapnik, 1998). The penalty parameter makes a trade-off between the margin size and the number of error instances in which a larger value for the penalty parameter leads to smaller misclassification error and smaller margin size (Vapnik, 1998). SVM addresses non-linearity uses a mathematical function called kernel trick through to explore the data in a higher-dimensional space (Statnikov, 2011). The performance of the SVM algorithm is dependent on suitable kernel functions which are, the polynomial kernel, sigmoid kernel, radial basis function, and linear kernel

(Chapelle et al., 1999; Choubin et al., 2019). Furthermore, SVM prevents overfitting in the model and assures good generalization and classification performance (Huang et al., 2002). Continuous and categorical variables can be both processed by SVM effectively, and it can also handle non-linear data, complex and noisy data with outliers (Huang et al., 2002; Karimi et al., 2019b).

4. Methodology for forest fire susceptibility mapping

The computation of forest fire susceptibility mapping was performed through the following steps. Also, the graphical representation of the order of methodological steps is highlighted in Fig. 5.

4.1. Establishment of fire database

The database containing the 10 fire conditioning factors and 510 fire locations was processed in ArcGIS 10.4 software. In order to be included in the methodological procedure, all the conditioning variables were converted into raster format with the pixel size equal to 30 m. Given the fact that forest fire susceptibility represents a binary classification, the preparation of another dataset containing the non-fire locations was required (Tien Bui et al., 2017). The use of a dataset representing the absence of the phenomenon in the modelling process can also increase the performance of the applied methods (Costache et al., 2020b). These non-fire points were randomly placed within the areas characterized by the negative values of NDVI which highlights the presence of water bodies and poor vegetation where the fire forest is almost impossible (Tien Bui et al., 2017). Also, it should be mentioned that the number of non-fire locations is equal to those of fire locations. The fire locations were attributed value "1" and the non-fire locations were attributed value "0".

4.2. Setup of training and validating dataset

The validation is a mandatory step in the forecasting studies as it can be affected by a specific phenomenon. Thus, the fire and non-fire samples were divided into training dataset (70%) and validating dataset (30%) (Tien Bui et al., 2017; Pourtaghi et al., 2015). Therefore, 714 fire and non-fire locations will be used to train the models, while another 306 fire and non-fire locations will be involved in the validation of fire susceptibility maps. The *Subset Features* tool available in ArcGIS 10.4 was employed in order to establish both the training and validation dataset.

4.3. Computing FR coefficients

Since FR coefficients are used as input to the machine learning models, these values were calculated only for the training dataset. The use of FR coefficients as input in natural hazards modelling is considered an effective solution to assign the same type of values to the classes/categories of all (Costache and Tien Bui, 2019; Tehrani et al., 2015a). Thus, by involving the FR values within the models the following ensembles will result MLP-FR, LR-FR, CART-FR, RF-FR and SVM-FR. The computation of FR coefficients will be done according to section 3.1. Moreover, in order to bring the value of the fire conditioning factor within the same range, the FR coefficients were normalized between 0.1 and 0.9 according to Equation (8) below (Costache, 2019a):

$$x = \frac{(v - \min(r)) * (\max(l) - \min(l))}{\max(r) - \min(r)} + \min(l)$$

where x is the standardized value of v ; v is the current value of the variable; r is the limits of the range value, and l is the limits of the standardization range.

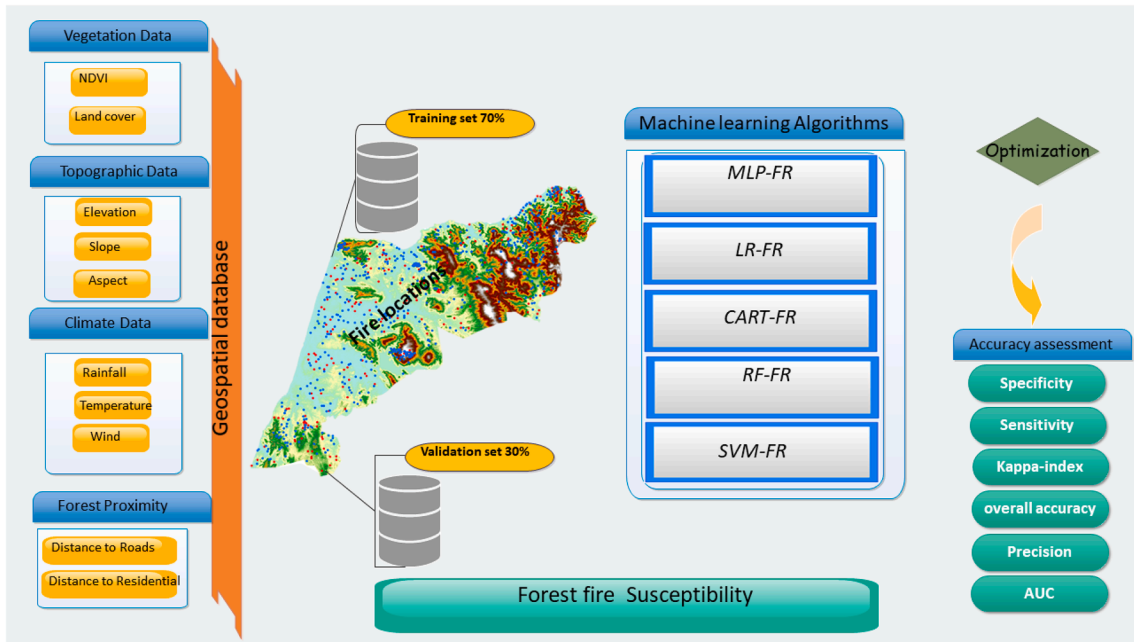


Fig. 5. Flowchart of the methodology adopted.

4.4. Configuration of fire susceptibility models

The next stage of the modelling process is represented by the configuration of the models used to calculate the fire forest susceptibility. In this regard, the FR values which characterize the factor classes/categories were assigned to fire and non-fire locations. In this study,

Python 3.8.2 (“Python Release Python 3.8.2,” n.d.) was used to apply machine learning models. Therefore, the fire and non-fire locations having attributed the FR coefficient were converted in tabular format in order to be read in Python. Then, different amounts of hyperparameters were set, and by running the models several times, the best configuration was chosen to reach the highest accuracy for each model. For RF

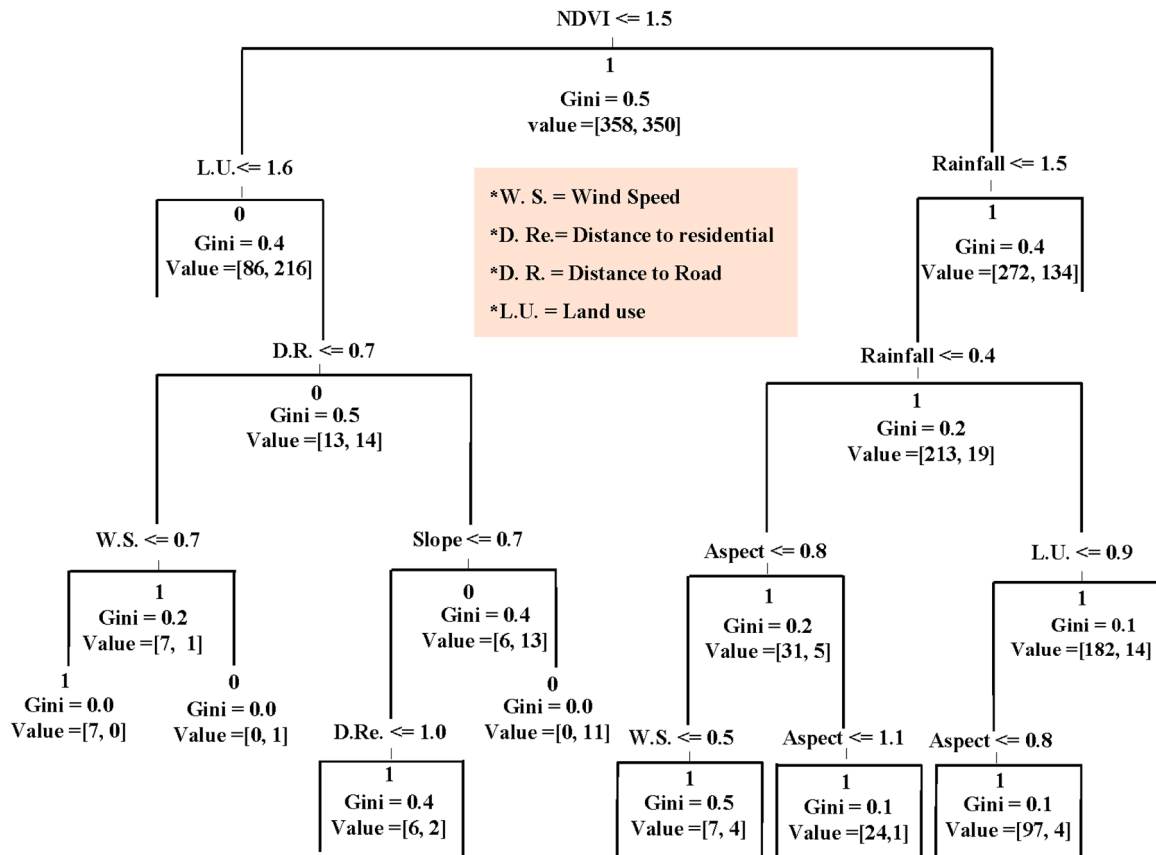


Fig. 6. Structure of the RF model used for forest fire susceptibility in this study.

model the number of trees equal to 100 and the random split variable equal to 1 leads to the highest accuracy with the minimum time to get the results. The advantage of RF comparing to other models is the smaller number of hyperparameters to be set. Based on these criteria, the structure of the RF model used in this research is shown in Fig. 6.

For SVM model, the penalty parameter equal to 5 and the kernel function of radial basis function and its parameter equal to 1 had the highest accuracy. For LR model, the maximum iteration equal to 1000, the inverse of regularization strength equal to 3, and the tolerance for stopping criteria equal to 0.001 were set. However, in this model changing the hyperparameters did not change the accuracy meaningful. For the NN model (Fig. 7) the optimizer was set to Adam, the activation function was set to Relu (Rectification Linear Unit), the number of hidden layers was set to 60, and the number of neurons in each layer was set to 2 to reach the highest accuracy. For the CART model, the maximum number of splits equal to 10, the minimum number of parent nodes equal to 10 and the minimum number of leaf nodes equal to 5 resulted in the highest accuracy. Fig. 8 illustrates the structure of the CART used for modelling forest fire in this research.

4.5. Evaluation of models performance

Once the models of the forest fire susceptibility are implemented, it is necessary to evaluate their performance. To achieve this, we used the receiver operating characteristic (ROC) curve (Chen et al., 2019; Jaafari et al., 2019b; Tien Bui et al., 2017), The ROC curve is plotted with the sensitivity as the y-axis and the 1-specificity as the x-axis (Costache, 2019b, 2019c; Pourghasemi et al., 2016; Tien Bui et al., 2017). The AUC value ranges from 0.5 to 1. The highest AUC value indicates a perfect measure of separability, while the lowest AUC value indicates the worst measure of separability.

Additionally, we used statistical measures such as overall accuracy, precision, specificity, sensitivity, and Kappa index (Tien Bui et al.,

2017). These statistical criteria were calculated based on the following equations:

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

$$Specificity = \frac{TN}{FP + TN} \tag{10}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{11}$$

In the above equations, TP (true positive) and TN (true negative) are the number of pixels which are considered correctly classified as forest fire and non-forest fire, respectively. While FP (false positive) and FN (false Negative) are the numbers of pixels that are incorrectly classified.

4.6. Fire susceptibility mapping

Based on the results obtained, the models for forest fire susceptibility mapping for the study area were adopted. To generate forest fire susceptibility maps, the layers obtained were then transformed to GIS environment, the five classes of very low, low, moderate, high, and very high susceptibility to forest fire were applied (Jaafari et al., 2019b; Tien Bui et al., 2017) using natural breaks method in ArcGIS 10.4 (Tehrany et al., 2019).

5. Results and discussion

5.1. Frequency Ratio weights

The results of the first of the workflow are represented by the values of FR coefficients. FR coefficients equal to 0 were achieved by the mean annual temperature between 15.9 °C and 16.5 °C and the water bodies land use category. On the surfaces characterized by these two

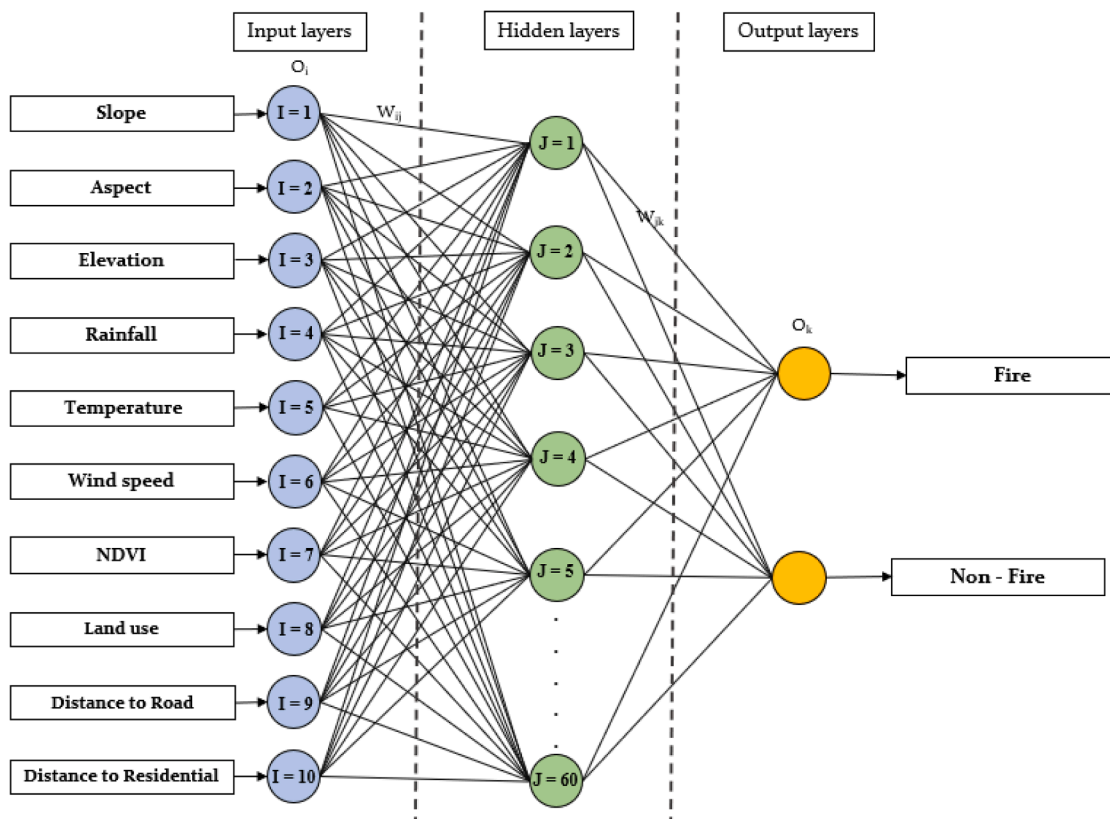


Fig. 7. Architecture of MLP used in this study.

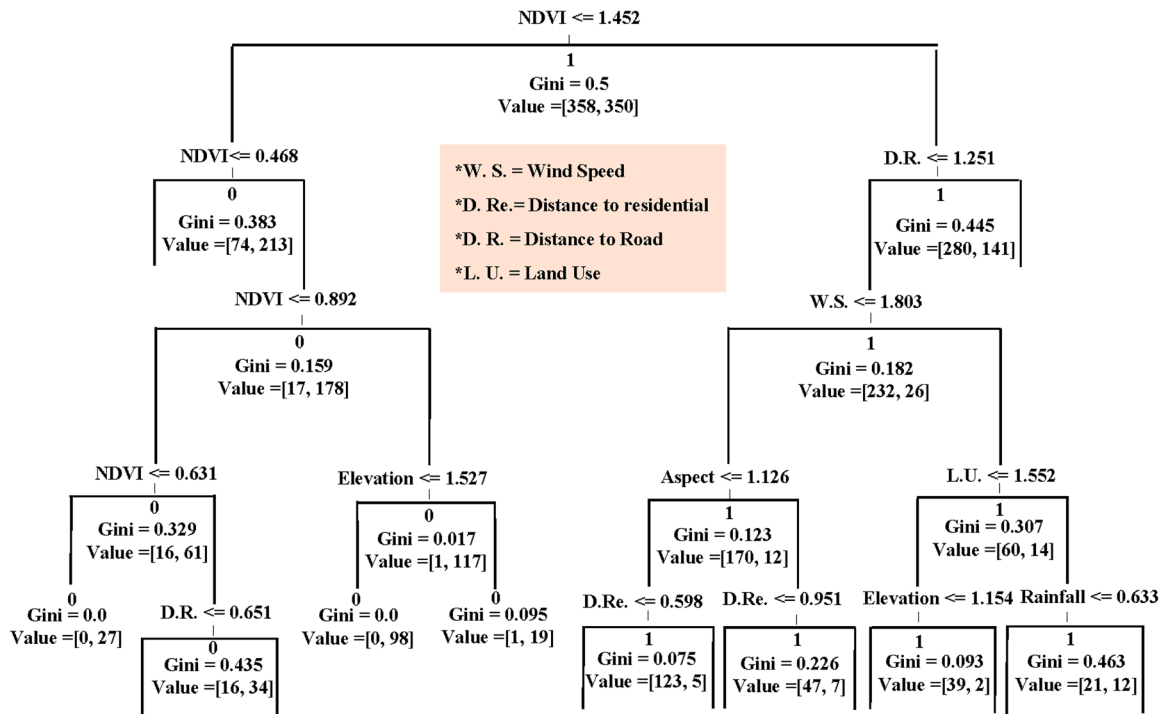


Fig. 8. Structure of the CART model used for forest fire susceptibility in this study.

parameters, no forest fire location was recorded. Low FR coefficients were also achieved by the flat zones (0.215), rainfall class between 629 and 700 mm/year (0.314), or the north-eastern slopes (0.445). The highest FR values were recorded by the areas with rainfall between 900 and 943 mm/year (3.58), distances to roads between 240.1 m and 480 m (3.302), and the surfaces where the average wind speed is between 3.75 m/s and 4 m/s (2.653) (Table 3). According to the described methodology, these values were used as input in the 5 machine learning models.

5.2. Fire susceptibility mapping and models performance

5.2.1. MLP-Fr

The statistical metrics were involved in the estimation of MLP-FR performance (Table 4). Thus, the accuracy of the training dataset was equal to 0.883, the Precision achieved a value of 0.856, the Sensitivity was 0.904, the Specificity is 0.863, while the K-index was situated at 0.766. In terms of validating dataset, the following values of statistical metrics were achieved: Accuracy = 0.844, Precision = 0.801, Sensitivity = 0.877, Specificity = 0.817 and K-index = 0.689.

The fire susceptibility map was derived through the MLP-FR model (FSM-MLP-FR) after the model performance assessment (Fig. 9). The values of FSM-MLP-FR were grouped into five classes using the *Natural Breaks* method. The first class, highlighting the areas with very low susceptibility, accounts around 20.2% of the study area, the low fire susceptibility is presented at approximately 10.44%, while the medium class of FSM-MLP-FR occupies a surface equal to 21.92% of the research zone. The high and very high fire susceptibility span on a total of 47.62% of the study area.

5.2.2. CART-Fr

The same statistical measures were used to assess the performance of CART-FR ensemble for both training and validating datasets. The use of training sample revealed the achievement of the following results (Table 4): Accuracy = 0.9, Precision = 0.907, Sensitivity = 0.894, Specificity = 0.905 and K-index = 0.799; while the involvement of validating data set helps to achieve the next values: Accuracy = 0.828, Precision = 0.795, Sensitivity = 0.851, Specificity = 0.807 and K-index

= 0.656. In both of the cases, the statistical metrics highlight very good performances of CART-FR ensemble.

Following the performance evaluation, the fire susceptibility map (FSM-CART-FR) was computed (Fig. 9). Similar to the previous case, the FSM-CART-FR values were grouped into 5 classes using the *Natural Breaks* method. The very low and low susceptibility is spread on around 22.65% of the research area, while the medium values are encountered at approximately 13.39%. The high and very high susceptibility can be found around 63.97% of the study area.

5.2.3. LR-Fr

The performances achieved by LR-FR ensemble were the lowest among all the applied models. Thus, in terms of training dataset the following values were achieved (Table 4): Accuracy = 0.766, Precision = 0.754, Sensitivity = 0.772, Specificity = 0.76 and K-index = 0.531. The validation dataset revealed the following values for statistical metrics: Accuracy = 0.768, Precision = 0.781, Sensitivity = 0.761, Specificity = 0.776 and K-index = 0.536.

The classification of Fires Susceptibility Map (FSM-LR-FR) into five classes (Fig. 9), according to the *Natural Breaks* method, highlights the following situation: very low susceptibility present on 17.26% of the study area, low susceptibility spread on 18.58% of the research area, medium values located on 23.4%, and high and very high susceptibility with a surface equal to 40.77% of the study area.

5.2.4. SVM-Fr

The application of SVM-FR ensemble is characterized by the following performances in terms of training dataset (Table 4): Accuracy = 0.91, Precision = 0.898, Sensitivity = 0.919, Specificity = 0.901 and K-index = 0.819. The use of validating dataset shows that the Accuracy = 0.858, the Precision = 0.841, the Sensitivity = 0.870, the Specificity = 0.846 while the K-index = 0.715. It should be mentioned that overall the SVM-FR ensemble achieved the second-best performance after RF-FR model.

The Forest Fire Susceptibility Map was represented by dividing the FSM-SVM-FR values into five classes using the *Natural Breaks* method (Fig. 9). The very low fire susceptibility appears on approximately

Table 3
Frequency Ratio values for factors classes/categories.

Factors	Classes	Class pixels	Fire pixels	FR
Slope	0 – 3°	350,477	42	0.458
	3.1 – 7°	536,399	125	0.890
	7.1 – 15°	707,757	223	1.203
	15.1 – 25°	285,455	101	1.351
	25.1–62.2°	48,361	14	1.105
Aspect	Flat zones	35,581	2	0.215
	North	225,210	44	0.746
	North-East	240,527	28	0.445
	East	234,241	47	0.766
	South-East	205,031	85	1.583
	South	230,302	92	1.525
	South-West	249,841	88	1.345
	West	267,906	62	0.884
	North-West	239,810	57	0.908
Distance to residential areas	0–1000 m	122,140	23	0.719
	1000–2000 m	282,685	36	0.486
	2000–3000 m	344,728	64	0.709
	3000–4000 m	343,593	123	1.367
	4000–10835 m	835,737	259	1.184
Distance to road	0–120 m	234,465	40	0.652
	120.1–240 m	120,365	57	1.809
	240.1–480 m	197,782	171	3.302
	480.1–840 m	248,156	45	0.693
	840.1–8988 m	1,128,115	192	0.650
Main temperature	15.9–16.5 °C	773,213	116	0.000
	16.6–17 °C	542,856	236	0.587
	17.1–17.5 °C	314,498	71	1.353
	17.6–18 °C	212,919	51	0.640
	18.1–18.5 °C	84,963	31	1.104
Land use	Water bodies	6598	0	0.000
	Forest	536,096	165	1.176
	Croplands	1,023,509	157	0.586
	build-up/Bare lands	362,639	183	1.927
Rainfall (mm/year)	629–700	194,672	16	0.314
	700–750	359,763	44	0.467
	750–800	553,382	81	0.559
	800–850	513,471	295	2.194
	850–900	291,610	54	0.707
	900–943	16,003	15	3.580
Wind speed (m/s)	3–3.25	86,078	11	0.488
	3.25–3.5	669,975	85	0.484
	3.5–3.75	625,851	110	0.671
	3.75–4	365,492	254	2.653
	4–4.5	180,327	45	0.953
Elevation	0–62 m	773,213	116	0.573
	62.1–142 m	542,856	236	1.660
	142.1–243 m	314,498	71	0.862
	243.1–370 m	212,919	51	0.915
	370.1–761 m	84,963	31	1.393
NDVI	–1 – –0.05	447,168	206	1.760
	–0.05–0	412,061	50	0.463
	0–0.05	281,062	47	0.639
	0.05–0.1	189,685	31	0.624
	0.1–0.5	550,391	165	1.145
	0.5–0.73	48,467	6	0.473

Table 4
Model performances estimated with training and validating samples.

Models	Sample	TN	FP	FN	Accuracy	Precision	Sensitivity	Specificity	K-index
RF-FR	Training	353	353	1	1	0.997	0.997	0.997	0.994
	Validating	138	135	13	16	0.904	0.914	0.896	0.808
SVM-FR	Training	318	326	36	28	0.910	0.898	0.919	0.819
	Validating	127	132	24	19	0.858	0.841	0.870	0.715
LR-FR	Training	267	275	87	79	0.766	0.754	0.772	0.531
	Validating	118	114	33	37	0.768	0.781	0.761	0.536
MLP-FR	Training	303	322	51	32	0.883	0.856	0.904	0.766
	Validating	121	134	30	17	0.844	0.801	0.877	0.689
CART-FR	Training	321	316	33	38	0.900	0.907	0.894	0.799
	Validating	120	130	31	21	0.828	0.795	0.851	0.656

26.74% of the study zone, the low susceptibility is spread on 8.09% of the total analyzed territory, the medium FSM-SVM-FR values are present on 11.09%, while the high and very high susceptibility accounts around 54.07% of the Fahs -Anjra province and Tanger-Asilah prefecture.

5.2.5. RF-Fr

After the training of the RF-FR ensemble, its performance was measured with the help of several statistical metrics (Table 4). Thus, in terms of training sample, the accuracy of 0.997 was the highest between all the applied models. In this case, the precision, sensitivity and specificity have the same values as the Accuracy while K-index is equal to 0.994. The involvement of validating sample in the assessment of models performance revealed that RF-FR achieved also the best results highlighted by the following values: Accuracy = 0.904, Precision = 0.914, Sensitivity = 0.896, Specificity = 0.912 and K-index = 0.808.

Once the model performance was assessed, the mapping of the Fire Susceptibility Index was performed (Fig. 9). Thus, the very low susceptibility has 18.68% of the total study area, the low susceptibility is present on 12.47% of the analyzed zone, the medium values are encountered on approximately 16.79%, while the areas exposed in a high and very high degree to fire occurrence can be found on around 52.05% of the study area.

5.3. Results validation

The results validation was done by using the ROC Curve. In this regard, the Success Rate, constructed with the training data, and Prediction Rate, constructed with the validating data were employed and their plots are shown in Fig. 10a and 10b respectively. Thus, in terms of success rate, the highest AUC value (0.998) was achieved by RF-FR ensemble followed by SVM-FR (0.973), CART-FR (0.967), MLP-FR (0.947), and LR-FR (0.822). The RF-FR ensemble was again the most performant in terms of Prediction Rate with an AUC of 0.952, followed by SVM-FR (0.928), MLP-FR (0.918), CART (0.905), and LR-FR (0.905). Given the fact that all the models, for both success and prediction rate, achieved AUC values higher than 0.8, we can assume that the applied algorithms were performant concerning the identification of areas susceptible to forest fire occurrence.

5.4. The importance of conditioning factors

Machine learning models have recently become the focus of intense interest of researchers in several environmental hazards studies. Modeling of Forest fire is a complex issue (Tien Bui et al., 2017), its propagation is commonly linked to several factors, such as climate, topography, human, and vegetation factors. However, at present, there is no agreement for the selection of explicative variables for forest fire susceptibility.

In this research, we showed that 10 influencing parameters (i.e. slope, aspect, elevation, distance to road, distance to residential, normalized difference vegetation index, land use, rainfall, temperature, and wind speed) could be used to implement FR-MLP, FR-LR, FR-CART,

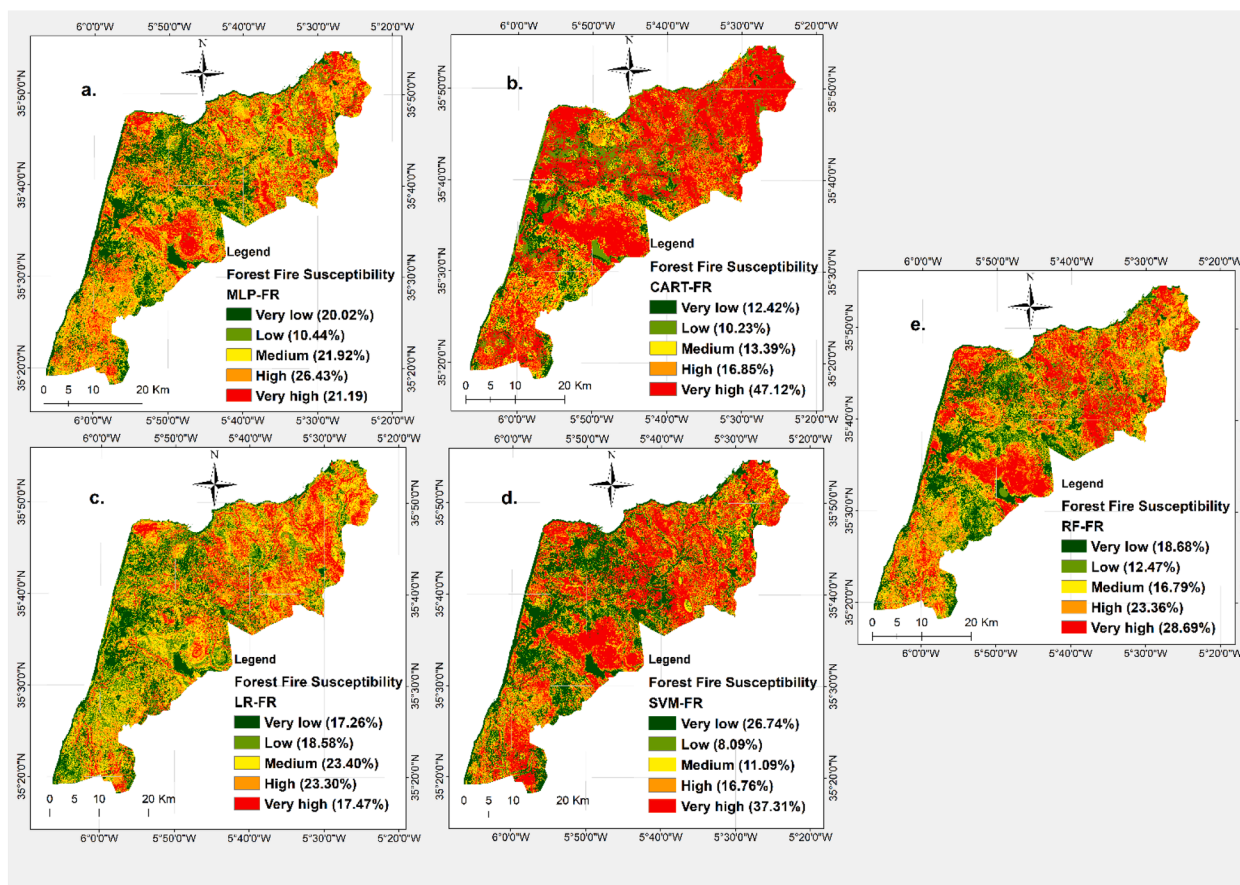


Fig. 9. Forest Fire Susceptibility using ensemble models.

FR-SVM and FR-RF for mapping forest fire susceptibility in the north of Morocco. These models classify the study area into 5 levels of risk. The outcomes from this study could be served as a reference to identify areas which require emergency intervention.

As mentioned before, one of the advantages of RF method is its ability to estimate the importance of the features used for modeling. Fig. 11 shows the result of evaluating the conditioning factors for fire forest modeling in the study area using RF method. It is obvious that NDVI is the most important conditioning factor, following by distance to roads, rainfall and wind speed. Slope has the least importance among the conditioning factors. This finding support other previous studies (e.g. Holsinger et al., 2016; Tien Bui et al., 2019, 2017, 2016). However, (Pourghasemi et al., 2020) emphasized that land use, annual mean rainfall, slope, and elevation are the best factors for the forest fire susceptibility. Similar to this finding, in a study developed by (Pourtaghi et al., 2016) the most important factors were soil type, average annual temperature, and land-cover for forest fire susceptibility.

In this paper, the RF -FR model achieved the highest accuracy among other models, these findings are consistent with prior research (Pourtaghi et al., 2016; Tien Bui et al., 2017). These authors stated that this outperformance of RF is due to its advantages to handle and to deal with nonlinearities between variables, another essential advantage for RF is that it is easy to use and processing large dataset. Indeed, Random Forest algorithm has been employed with promising results in other prior research related to natural hazards, for example for flood susceptibility mapping, (Costache et al., 2020c), in their paper proved that RF achieved high accuracy among other algorithms applied in their case study.

As discussed above, FR is Followed by SVM-FR model in term of accuracy, the SVM, in a way that it can be applied for both classification and regression task, has numerous advantages such as its capacity to fix complexity of overfitting and its applicability to handle smaller dataset

with high dimensionality (Chen et al., 2017). Due to these advantages, SVM has been widely used in natural hazards assessment proving successful results. Also, (Tien Bui et al., 2017) reported the predictive performance for both RF and SVM for forest fire prediction. MLP showed effective results due to the ability to be involved in processing non-linear datasets (Huang et al., 2020).

CART has interesting advantages, Nevertheless, it is worthy to note that the CART algorithm is rarely present in the literature review for forest fire susceptibility and in general for moderate prediction purposes (Jaafari et al., 2018).

LR achieved also good accuracy. similarly, it has shown successful results in other fields, for example, flood mapping (Costache et al., 2020a) and landslide susceptibility (Nhu et al., 2020). Overall, the hybrid models based on ML algorithms and bivariate statistical methods provided promising and more performance results in comparison to individual models (Costache and Tien Bui, 2019; Pham et al., 2018b).

To sum up, selecting a suitable ML algorithm for forest fire models like for any other environmental hazards assessment is a challenging task because each model presents its drawbacks and advantages (Tien Bui et al., 2017; Jaafari et al., 2019a). Therefore, all the aforementioned developed models in this study are recommended for forest fire susceptibility studies scope.

Forest ecosystems in the north of Morocco are dramatically influenced by a number of natural and anthropogenic disturbance, for instance, in a recent work, (Salhi et al., 2020) reported the fragility of soil quality, hence, a situation which requires a need of anti-erosion activities. Also, the major part of landscape areas in the north of Morocco is exploited for cannabis plantations (El Motaki et al., 2019) due to its economic value. All these factors increase the widespread forest fire disaster. This paper highlights the urgent need for government agencies to act for maintaining forest ecosystem prevention and

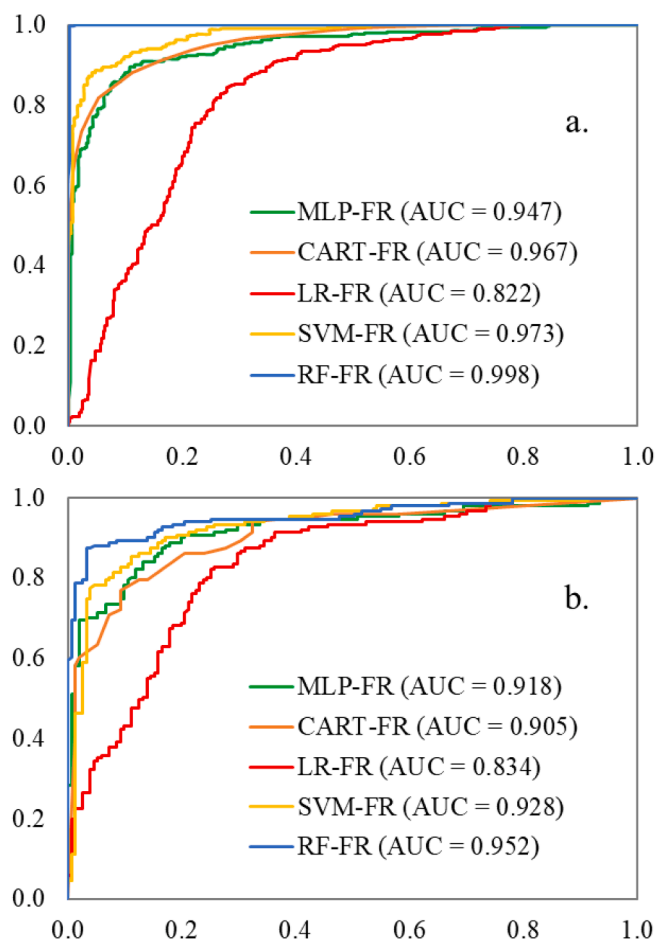


Fig. 10. ROC Curve (a. Success Rate; b. Prediction Rate).

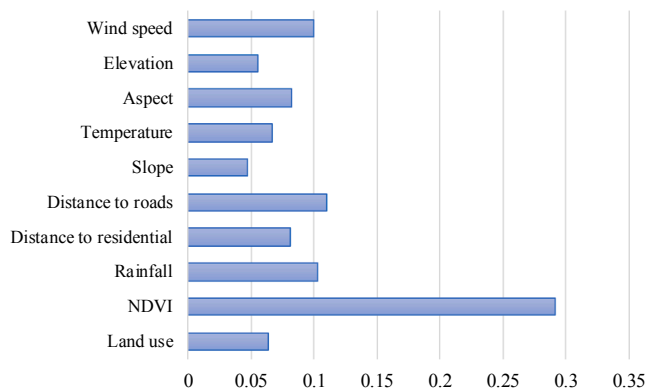


Fig. 11. The importance of conditioning factors.

planning mitigation strategies. Despite a huge effort provided by a public agency to cope with this situation even though dealing with local conflicts.

Finally, looking to an ecosystem forest sustainable for our future, such efforts could be a powerful tool and they must be accomplished: i- a public alarm system could be implemented to secure the urgent situation for the forest ecosystem. ii- delivering and communicating the results of the forest fire to the public community could be a pathway to reduce the influence of population behaviour.

6. Conclusion

Forest fire is one of the main causes of ecosystem service losses and deterioration of human life. Thus, it is considered one of the pressing issues. In many parts of the world governments' policy on forest protection and management is challenging to meet their goals. With these issues in mind, this paper is motivated by the need for preparation predictive models of the forest fire risk and identification of areas requiring immediate management actions with a goal of "Act Now before Tomorrow". In this paper, five hybrid models have been developed namely, RF-FR, SVM-FR, MLP-FR, CART-FR, and LR-FR for forest fire modelling based on 510 forest fire locations and a total of 10 forest fire conditioning factors (elevation, slope, aspect, distance to roads, distance to residential, land use, Normalized Difference Vegetation Index (NDVI), rainfall, temperature, and wind speed). Results of models proposed indicate that RF-FR achieved the highest performance (AUC = 0.989), followed by SVM-FR (AUC = 0.959), MLP-FR (AUC = 0.858), CART-FR (AUC = 0.847), LR-FR (AUC = 0.809) in the forecasting of forest fire, respectively. In light of these results, the maps generated here could be a very effective management tool for analyzing and developing forest fire management and strategies. Also, the methodology can further be extended to other areas with a similar issue.

CRediT authorship contribution statement

Meriam Mohajane: Conceptualization, Investigation, Project administration. **Romulus-Dumitru Costache:** Conceptualization, Methodology, Investigation, Data curation. **Firoozeh Karimi:** Conceptualization, Methodology, Resources, Data curation. **Quoc Bao Pham:** Methodology, Resources, Writing - original draft, Supervision, Project administration. **Ali Essahlaoui:** Validation, Visualization. **Hoang Nguyen:** Validation, Formal analysis, Writing - review & editing, Visualization. **Giovanni Laneve:** Formal analysis, Writing - review & editing. **Fatiha Oudija:** Formal analysis.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Ajbilou, R., Marañón, T., Arroyo, J., 2006. Ecological and biogeographical analyses of Mediterranean forests of northern Morocco. *Acta Oecologica* 29, 104–113. <https://doi.org/10.1016/j.actao.2005.08.006>.

Akay, H., 2021. Flood hazards susceptibility mapping using statistical, fuzzy logic, and MCDM methods. *Soft Comput.* 1–22. <https://doi.org/10.1007/s00500-021-05903-1>.

Amit Parashar, Sas Biswas, n.d. The Impact of Forest Fire on Forest Biodiversity in the Indian Himalayas, 2003. (Uttaranchal) [WWW Document]. URL <http://www.fao.org/3/XII/0358-B1.htm#fnB1> (accessed 4.27.20).

Arpaç, A., Malowerschnig, B., Sass, O., Vacik, H., 2014. Using multi variate data mining techniques for estimating fire susceptibility of Tyrolean forests. *Appl. Geogr.* 53, 258–270. <https://doi.org/10.1016/j.apgeog.2014.05.015>.

Basheer, I.A., Hajmeer, M., 2000. Artificial neural networks: fundamentals, computing, design, and application. *J. Microbiol. Methods, Neural Comput. Micrbiol.* 43 (1), 3–31. [https://doi.org/10.1016/S0167-7012\(00\)00201-3](https://doi.org/10.1016/S0167-7012(00)00201-3).

Bax, V., Francesconi, W., 2018. Environmental predictors of forest change: an analysis of natural predisposition to deforestation in the tropical Andes region, Peru. *Appl. Geogr.* 91, 99–110. <https://doi.org/10.1016/j.apgeog.2018.01.002>.

Belinchón, Rocío, Martínez, Isabel, Ojalora, Mónica A.G., Aragón, Gregorio, Dimas, Jesús, Escudero, Adrián, 2009. Fragment quality and matrix affect epiphytic performance in a Mediterranean forest landscape. *Am. J. Bot.* 96 (11), 1974–1982. <https://doi.org/10.3732/ajb.0900040>.

Belousov, Application aspects of support vector machines [WWW Document] <https://onlinelibrary.wiley.com/doi/abs/10.1002/cem.744> 2002 accessed 4.28.20.

Bonham-Carter, G.F., 2014. *Geographic Information Systems for Geoscientists: Modelling with GIS.* Elsevier.

Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers, in: *Proceedings of the Fifth Annual Workshop on Computational Learning Theory, COLT '92.* Association for Computing Machinery, Pittsburgh, Pennsylvania, USA, pp. 144–152. <https://doi.org/10.1145/130385.130401>.

- L. Breiman Random forests., Kluwer Academic Publishers. Manufactured in The Netherlands. ed. 2001 The Netherlands.
- Breiman, Leo, 1996. Bagging predictors. *Mach. Learn.* 24 (2), 123–140. <https://doi.org/10.1007/BF00058655>.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. *Classification and Regression Trees*. CRC Press.
- Tien Bui, Dieu, Bui, Quang-Thanh, Nguyen, Quoc-Phi, Pradhan, Biswajeet, Nampak, Haleh, Trinh, Phan Trong, 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.
- Chapelle, O., Haffner, P., Vapnik, V.N., 1999. Support vector machines for histogram-based image classification. *IEEE Trans. Neural Networks* 10, 1055–1064. <https://doi.org/10.1109/72.788646>.
- Chavez, P.S., 1996. Image-Based Atmospheric Corrections - Revisited and Improved 12. *Photogramm. Eng. Remote Sens.* 62(9), 1025–1035.
- Chebli, Y., Chentouf, M., Ozer, P., Hornick, J.-L., Cabarau, J.-F., 2018. Forest and silvopastoral cover changes and its drivers in northern Morocco. *Appl. Geogr.* 101, 23–35. <https://doi.org/10.1016/j.apgeog.2018.10.006>.
- Chebli, Y., Otmami, S.E., Chentouf, M., Hornick, J.-L., Bindelle, J., Cabarau, J.-F., 2020. Foraging behavior of goats browsing in Southern Mediterranean Forest Rangeland. *Animals* 10, 196. <https://doi.org/10.3390/ani10020196>.
- Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D.T., Duan, Z., Ma, J., 2017. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *CATENA* 151, 147–160. <https://doi.org/10.1016/j.catena.2016.11.032>.
- Chen, Wei, Yan, Xusheng, Zhao, Zhou, Hong, Haoyuan, Bui, Dieu Tien, Pradhan, Biswajeet, 2019. Spatial prediction of landslide susceptibility using data mining-based kernel logistic regression, naive Bayes and RBFNetwork models for the Long County area (China). *Bull. Eng. Geol. Environ.* 78 (1), 247–266. <https://doi.org/10.1007/s10064-018-1256-z>.
- Choubin, B., Moradi, E., Golshan, M., Adamowski, J., Sajedi-Hosseini, F., Mosavi, A., 2019. An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Sci. Total Environ.* 651, 2087–2096. <https://doi.org/10.1016/j.scitotenv.2018.10.064>.
- Chuvieco, Emilio, Congalton, Russell G., 1989. Application of remote sensing and geographic information systems to forest fire hazard mapping. *Remote Sens. Environ.* 29 (2), 147–159. [https://doi.org/10.1016/0034-4257\(89\)90023-0](https://doi.org/10.1016/0034-4257(89)90023-0).
- Conedera, Marco, Torriani, Damiano, Neff, Christophe, Ricotta, Carlo, Bajocco, Sofia, Pezzatti, Gianni Boris, 2011. Using Monte Carlo simulations to estimate relative fire ignition danger in a low-to-medium fire-prone region. *For. Ecol. Manag.* 261 (12), 2179–2187. <https://doi.org/10.1016/j.foreco.2010.08.013>.
- Costache, R., 2019a. Flash-Flood Potential assessment in the upper and middle sector of Prahova river catchment (Romania). A comparative approach between four hybrid models. *Sci. Total Environ.* 659, 1115–1134. <https://doi.org/10.1016/j.scitotenv.2018.12.397>.
- Costache, Romulus, 2019b. Flash-flood Potential Index mapping using weights of evidence, decision Trees models and their novel hybrid integration. *Stoch. Environ. Res. Risk Assess.* 33 (7), 1375–1402.
- Costache, Romulus, 2019c. Flood susceptibility assessment by using bivariate statistics and machine learning models-a useful tool for flood risk management. *Water Resour. Manag.* 33 (9), 3239–3256.
- Costache, Romulus, Tien Bui, Dieu, 2019. Spatial prediction of flood potential using new ensembles of bivariate statistics and artificial intelligence: a case study at the Putna river catchment of Romania. *Sci. Total Environ.* 691, 1098–1118.
- Costache, Romulus, Hong, Haoyuan, Pham, Quoc Bao, 2020a. Comparative assessment of the flash-flood potential within small mountain catchments using bivariate statistics and their novel hybrid integration with machine learning models. *Sci. Total Environ.* 711, 134514. <https://doi.org/10.1016/j.scitotenv.2019.134514>.
- Costache, Romulus, Pham, Quoc Bao, Sharifi, Ehsan, Linh, Nguyen Thi Thuy, Abba, S.I., Vojtek, Matej, Vojteková, Jana, Nhi, Pham Thi Thao, Khoi, Dao Nguyen, 2020b. Flash-flood susceptibility assessment using multi-criteria decision making and machine learning supported by remote sensing and GIS techniques. *Remote Sens.* 12 (1), 106. <https://doi.org/10.3390/rs12010106>.
- Costache, Romulus, Popa, Mihnea Cristian, Tien Bui, Dieu, Diaconu, Daniel Constantin, Ciobotaru, Nicu, Minea, Gabriel, Pham, Quoc Bao, 2020c. Spatial predicting of flood potential areas using novel hybridizations of fuzzy decision-making, bivariate statistics, and machine learning. *J. Hydrol.* 585, 124808. <https://doi.org/10.1016/j.jhydrol.2020.124808>.
- Cox, D.R., 1958. The regression analysis of binary sequences. *J. R. Stat. Soc. Ser. B Methodol.* 20 (2), 215–232. <https://doi.org/10.1111/rssb.1958.20.issue-2.10.1111/j.2517-6161.1958.tb00292.x>.
- Delen, Dursun, Kuzey, Cemil, Uyar, Ali, 2013. Measuring firm performance using financial ratios: a decision tree approach. *Expert Syst. Appl.* 40 (10), 3970–3983. <https://doi.org/10.1016/j.eswa.2013.01.012>.
- Díaz-Avalos, Carlos, Peterson, David L., Alvarado, Ernesto, Ferguson, Sue A., Besag, Julian E., 2001. Space–time modelling of lightning-caused ignitions in the Blue Mountains. *Oregon. Can. J. For. Res.* 31 (9), 1579–1593. <https://doi.org/10.1139/cjfr-31-9-1579>.
- DRATT (Direction Régionale de l'Agriculture de Tanger-Tétouan), 2015. *Monographie de la région du nord du Maroc. Rapports et Données statistiques sur l'agriculture et l'élevage*, n.d.
- El Hafyani, M., Essahlaoui, A., Van Rompaey, A., Mohajane, M., El Hmaid, A., El Ouali, A., Moudeden, F., Serhini, N.-E., 2020. Assessing regional scale water balances through remote sensing techniques: a case study of Boufakrane River Watershed, Meknes Region, Morocco. *Water* 12, 320. <https://doi.org/10.3390/w12020320>.
- El Motaki, H., El-Fengour, A., Aissa, E., Madureira, H., Monteiro, A., 2019. The global change impacts on forest natural resources in Central Rif Mountains in northern Morocco: extensive exploration and planning perspective. *GOT - J. Geogr. Spat. Plan.* 75–92. <https://doi.org/10.17127/got/2019.17.004>.
- Escudero, Adrián, Núñez, Yolanda, Pérez-García, Félix, 2000. Is fire a selective force of seed size in pine species? *Acta Oecologica* 21 (4–5), 245–256. [https://doi.org/10.1016/S1146-609X\(00\)01083-3](https://doi.org/10.1016/S1146-609X(00)01083-3).
- Fernández-García, V., Fulé, P.Z., Marcos, E., Calvo, L., 2019. The role of fire frequency and severity on the regeneration of Mediterranean serotinous pines under different environmental conditions. *For. Ecol. Manag.* 444, 59–68. <https://doi.org/10.1016/j.foreco.2019.04.040>.
- Freden, S.C., Mercanti, E.P., Becker, M.A., 1974. Third Earth Resources Technology Satellite-1 Symposium: The Proceedings of a Symposium Held by Goddard Space Flight Center at Washington, D.C. on December 10–14, 1973; Prepared at Goddard Space Flight Center. Scientific and Technical Information Office, National Aeronautics and Space Administration.
- Friedl, M.A., Brodley, C.E., 1997. Decision tree classification of land cover from remotely sensed data. *Remote Sens. Environ.* 61 (3), 399–409. [https://doi.org/10.1016/S0034-4257\(97\)00049-7](https://doi.org/10.1016/S0034-4257(97)00049-7).
- Ganteaume, Anne, Camia, Andrea, Jappiot, Marielle, San-Miguel-Ayanz, Jesus, Long-Fournel, Marlène, Lampin, Corinne, 2013. A review of the main driving factors of forest fire ignition over Europe. *Environ. Manage.* 51 (3), 651–662. <https://doi.org/10.1007/s00267-012-9961-z>.
- Giglio, Louis, Desclotres, Jacques, Justice, Christopher O., Kaufman, Yoram J., 2003. An enhanced contextual fire detection algorithm for MODIS. *Remote Sens. Environ.* 87 (2–3), 273–282. [https://doi.org/10.1016/S0034-4257\(03\)00184-6](https://doi.org/10.1016/S0034-4257(03)00184-6).
- Hantson, Stijn, Pueyo, Salvador, Chuvieco, Emilio, 2016. Global fire size distribution: from power law to log-normal. *Int. J. Wildland Fire* 25 (4), 403. <https://doi.org/10.1071/WF15108>.
- Harris, I., Osborn, T.J., Jones, P., Lister, D., 2020. CRU TS4.01: Climatic Research Unit (CRU) Time-Series (TS) version 4.01 of high-resolution gridded data of month-by-month variation in climate (Jan. 1901–Dec. 2016), Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci. Data* 7, 109. <https://doi.org/10.1038/s41597-020-0453-3>.
- HCEFLCD, n.d. HCEFLCD, 2011. Haut-Commissariat aux Eaux et Forêts et à la Lutte Contre la Désertification Les incendies de Forêts Maroc. Département des Eaux et Forêts [WWW Document]. URL <http://www.eauxetforets.gov.ma/ProtectionForet/Incendies/Pages/Incendies.aspx> (accessed 4.29.20).
- HCP, (Haut-Commissariat au Plan), 2014. *Monographie régionale de Tanger-Tétouan*.
- Holsinger, L., Parks, S.A., Miller, C., 2016. Weather, fuels, and topography impede wildland fire spread in western US landscapes. *For. Ecol. Manag.* 380, 59–69. <https://doi.org/10.1016/j.foreco.2016.08.035>.
- Hong, H., Naghibi, S.A., Moradi Dashtpajardi, M., Pourghasemi, H.R., Chen, W., 2017. A comparative assessment between linear and quadratic discriminant analyses (LDA-QDA) with frequency ratio and weights-of-evidence models for forest fire susceptibility mapping in China. *Arab. J. Geosci.* 10, 167. <https://doi.org/10.1007/s12517-017-2905-4>.
- Hong, H., Pradhan, B., Jebur, M.N., Bui, D.T., Xu, C., Akgun, A., 2016. Spatial prediction of landslide hazard at the Luxi area (China) using support vector machines. *Environ. Earth Sci.* 75, 40. <https://doi.org/10.1007/s12665-015-4866-9>.
- Hong, H., Tsangaratos, P., Ili, I., Liu, J., Zhu, A.-X., Xu, C., 2018. Applying genetic algorithms to set the optimal combination of forest fire related variables and model forest fire susceptibility based on data mining models. The case of Dayu County. *China. Sci. Total Environ.* 630, 1044–1056. <https://doi.org/10.1016/j.scitotenv.2018.02.278>.
- Hosmer Jr, D.W., Lemeshow, S., 2000. *Applied Logistic Regression*. Wiley, New York.
- Huang, C., Davis, S., Townshend, J.R.G., 2002. An assessment of support vector machines for land cover classification: *International Journal of Remote Sensing: Vol 23, No 4* [WWW Document]. URL <https://www.tandfonline.com/doi/abs/10.1080/01431160110040323> (accessed 4.28.20).
- Huang, F., Cao, Z., Guo, J., Jiang, S.-H., Li, S., Guo, Z., 2020. Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. *CATENA* 191, 104580. <https://doi.org/10.1016/j.catena.2020.104580>.
- Huebner, K., Lindo, Z., Lechowicz, M.J., 2012. Post-fire succession of collembolan communities in a northern hardwood forest. *Eur. J. Soil Biol.* 48, 59–65. <https://doi.org/10.1016/j.ejsobi.2011.10.004>.
- Jaafari, A., Razavi Termeh, S.V., Bui, D.T., 2019a. Genetic and firefly metaheuristic algorithms for an optimized neuro-fuzzy prediction modeling of wildfire probability. *J. Environ. Manage.* 243, 358–369. <https://doi.org/10.1016/j.jenvman.2019.04.117>.
- Jaafari, A., Panahi, M., Pham, B.T., Shahabi, H., Bui, D.T., Rezaie, F., Lee, S., 2019b. Meta optimization of an adaptive neuro-fuzzy inference system with grey wolf optimizer and biogeography-based optimization algorithms for spatial prediction of landslide susceptibility. *CATENA* 175, 430–445. <https://doi.org/10.1016/j.catena.2018.12.033>.
- Jaafari, A., Zenner, E.K., Pham, B.T., 2018. Wildfire spatial pattern analysis in the Zagros Mountains, Iran: a comparative study of decision tree based classifiers. *Ecol. Inform.* 43, 200–211. <https://doi.org/10.1016/j.ecoinf.2017.12.006>.
- Jiang, Z., Huete, A.R., Chen, J., Chen, Y., Li, J., Yan, G., Zhang, X., 2006. Analysis of NDVI and scaled difference vegetation index retrievals of vegetation fraction. *Remote Sens. Environ.* 101, 366–378. <https://doi.org/10.1016/j.rse.2006.01.003>.
- Karimi, F., Sultana, S., Babakan, A.S., Suthaharan, S., 2019a. Urban expansion modeling using an enhanced decision tree algorithm. *GeoInformatica*. <https://doi.org/10.1007/s10707-019-00377-8>.

- Karimi, F., Sultana, S., Shirzadi Babakan, A., Suthaharan, S., 2019b. An enhanced support vector machine model for urban expansion prediction. *Comput. Environ. Urban Syst.* 75, 61–75. <https://doi.org/10.1016/j.compenvurbys.2019.01.001>.
- Keeley, J.E., Pausas, J.G., Rundel, P.W., Bond, W.J., Bradstock, R.A., 2011. Fire as an evolutionary pressure shaping plant traits. *Trends Plant Sci.* 16, 406–411. <https://doi.org/10.1016/j.tplants.2011.04.002>.
- Kingma, D.P., Ba, J., 2017. Adam: A Method for Stochastic Optimization. ArXiv1412.6980 Cs.
- Knudby, A., LeDrew, E., Brenning, A., 2010. Predictive mapping of reef fish species richness, diversity and biomass in Zanzibar using IKONOS imagery and machine-learning techniques. *Remote Sens. Environ.* 114, 1230–1241. <https://doi.org/10.1016/j.rse.2010.01.007>.
- Krawchuk, M.A., Cumming, S.G., Flannigan, M.D., Wein, R.W., 2006. Biotic and abiotic regulation of lightning fire initiation in the mixedwood boreal forest. *Ecology* 87, 458–468. <https://doi.org/10.1890/05-1021>.
- Lee, Saro, Talib, Jasmi Abdul, 2005. Probabilistic landslide susceptibility and factor effect analysis. *Environ. Geol.* 47 (7), 982–990.
- Lentile, L.B., Smith, F.W., Shepperd, W.D., 2006. Influence of topography and forest structure on patterns of mixed severity fire in ponderosa pine forests of the South Dakota Black Hills, USA. *Int. J. Wildland Fire* 15, 557. <https://doi.org/10.1071/WF05096>.
- Li, J., Shan, Y., Yin, S., Wang, M., Sun, L., Wang, D., 2019. Nonparametric multivariate analysis of variance for affecting factors on the extent of forest fire damage in Jilin Province, China. *J. For. Res.* 30, 2185–2197. <https://doi.org/10.1007/s11676-019-00958-1>.
- Lippmann, R., 1987. An introduction to computing with neural nets. *IEEE ASSP Mag.* 4, 4–22. <https://doi.org/10.1109/MASSP.1987.1165576>.
- Menard, S., 2000. Coefficients of determination for multiple logistic regression analysis. *Am. Stat.* 54, 17–24. <https://doi.org/10.1080/00031305.2000.10474502>.
- Moayedi, H., Mehri, M., Bui, D.T., Pradhan, B., Foong, L.K., 2020. Fuzzy-metaheuristic ensembles for spatial assessment of forest fire susceptibility. *J. Environ. Manage.* 260, 109867. <https://doi.org/10.1016/j.jenvman.2019.109867>.
- Mohajane, M., Essahlaoui, A., Oudija, F., El Hafyani, M., Hmaid, A.E., El Ouali, A., Randazzo, G., Teodoro, A.C., 2018. Land use/land cover (LULC) using landsat data series (MSS, TM, ETM+ and OLI) in Azrou Forest, in the Central Middle Atlas of Morocco. *Environments* 5, 131. <https://doi.org/10.3390/environments5120131>.
- Nami, M.H., Jaafari, A., Fallah, M., Nabituni, S., 2018. Spatial prediction of wildfire probability in the Hyrcanian ecoregion using evidential belief function model and GIS. *Int. J. Environ. Sci. Technol.* 15, 373–384. <https://doi.org/10.1007/s13762-017-1371-6>.
- Nefeslioglu, H.A., Sezer, E., Gokceoglu, C., Bozkir, A.S., Duman, T.Y., 2010. Assessment of landslide susceptibility by decision trees in the Metropolitan area of Istanbul. *Turkey. Math. Probl. Eng.* 2010, 1–15. <https://doi.org/10.1155/2010/901095>.
- Nhu, V.-H., Mohammadi, A., Shahabi, H., Ahmad, B.B., Al-Ansari, N., Shirzadi, A., Geertsema, M., R. Kress, V., Karimzadeh, S., Valizadeh Kamran, K., Chen, W., Nguyen, H., 2020. Landslide Detection and Susceptibility Modeling on Cameron Highlands (Malaysia): A Comparison between Random Forest, Logistic Regression and Logistic Model Tree Algorithms. *Forests* 11, 830. <https://doi.org/10.3390/f11080830>.
- Ni, X., 2008. Research of data mining based on neural networks. *World Acad. Sci. Eng. Technol.* 39 (1), 381–384.
- Nsengiyumva, J.B., Luo, G., Amanambu, A.C., Mind'je, R., Habiyaemye, G., Karamage, F., Ochege, F.U., Mupenzi, C., 2019. Comparing probabilistic and statistical methods in landslide susceptibility modeling in Rwanda/Centre-Eastern Africa. *Sci. Total Environ.* 659, 1457–1472. <https://doi.org/10.1016/j.scitotenv.2018.12.248>.
- Oliveira, S., Félix, F., Nunes, A., Lourenço, L., Laneve, G., Sebastián-López, A., 2018. Mapping wildfire vulnerability in Mediterranean Europe. Testing a stepwise approach for operational purposes. *J. Environ. Manage.* 206, 158–169. <https://doi.org/10.1016/j.jenvman.2017.10.003>.
- 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. <https://doi.org/10.1016/j.foreco.2012.03.003>.
- Olivella, M., Ribalta, T., Defebris, A., Mollet, J., Delasheris, F., 2006. Distribution of polycyclic aromatic hydrocarbons in riverine waters after Mediterranean forest fires. *Sci. Total Environ.* 355, 156–166. <https://doi.org/10.1016/j.scitotenv.2005.02.033>.
- Pausas, J.G., Vallejo, V.R., 1999. The role of fire in European Mediterranean ecosystems, in: Chuvieco, E. (Ed.), *Remote Sensing of Large Wildfires*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 3–16. https://doi.org/10.1007/978-3-642-60164-4_2.
- Peng, C.-Y.J., Lee, K.L., Ingersoll, G.M., 2002. An introduction to logistic regression analysis and reporting. *J. Educ. Res.* 96, 3–14. <https://doi.org/10.1080/00220670209598786>.
- Peng, C.-Y.J., Manz, B.D., Keck, J., 2001. Modeling Categorical Variables by Logistic Regression [WWW Document]. <https://doi.org/info:doi/10.5993/AJHB.25.3.15>.
- Pham, Q.B., Yang, T.C., Kuo, C.M., Tseng, H.W., Yu, P.S., 2019. Combining random forest and least square support vector regression for improving extreme rainfall downscaling. *Water* 11 (3), 451. <https://doi.org/10.3390/w11030451>.
- Pham, Q.B., Afan, H.A., Mohammadi, B., Ahmed, A.N., Linh, N.T.T., Vo, N.D., et al., 2020. Hybrid model to improve the river streamflow forecasting utilizing multi-layer perceptron-based intelligent water drop optimization algorithm. *Soft. Comput.* 24, 18039–18056. <https://doi.org/10.1007/s00500-020-05058-5>.
- Pham, Q.B., Mohammadpour, R., Linh, N.T.T., Mohajane, M., Pourjassam, A., Sammen, S., Anh, D.T., 2021. Application of soft computing to predict water quality in wetland. *Environ. Sci. Pollut. Res.* 28 (1), 185–200. <https://doi.org/10.1007/s11356-020-10344-8>.
- Pham, B.T., Prakash, I., Tien Bui, D., 2018a. Spatial prediction of landslides using a hybrid machine learning approach based on Random Subspace and Classification and Regression Trees. *Geomorphology* 303, 256–270. <https://doi.org/10.1016/j.geomorph.2017.12.008>.
- Pham, B.T., Shirzadi, A., Tien Bui, D., Prakash, I., Dholakia, M.B., 2018b. A hybrid machine learning ensemble approach based on a Radial Basis Function neural network and Rotation Forest for landslide susceptibility modeling: A case study in the Himalayan area, India. *Int. J. Sediment Res.* 33, 157–170. <https://doi.org/10.1016/j.ijsrc.2017.09.008>.
- Pour, S.H., Wahab, A.K.A., Shahid, S., 2020. Physical-empirical models for prediction of seasonal rainfall extremes of Peninsular Malaysia. *Atmospheric Res.* 233, 104720. <https://doi.org/10.1016/j.atmosres.2019.104720>.
- 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, 80–98. <https://doi.org/10.1080/02827581.2015.1052750>.
- Pourghasemi, H.R., Gayen, A., Lasaponara, R., Tiefenbacher, J.P., 2020. Application of learning vector quantization and different machine learning techniques to assessing forest fire influence factors and spatial modelling. *Environ. Res.* 184, 109321. <https://doi.org/10.1016/j.envres.2020.109321>.
- Pourghasemi, H.R., Yousefi, S., Kornejady, A., Cerda, A., 2017. Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling. *Sci. Total Environ.* 609, 764–775. <https://doi.org/10.1016/j.scitotenv.2017.07.198>.
- Pourghasemi, H.R., Beheshtirad, M., Pradhan, B., 2016. A comparative assessment of prediction capabilities of modified analytical hierarchy process (M-AHP) and Mamdani fuzzy logic models using Netcad-GIS for forest fire susceptibility mapping. *Geomat. Nat. Hazards Risk* 7, 861–885. <https://doi.org/10.1080/19475705.2014.984247>.
- 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. <https://doi.org/10.1016/j.ecolind.2015.12.030>.
- Pourtaghi, Z.S., Pourghasemi, H.R., Rossi, M., 2015. Forest fire susceptibility mapping in the Minudasht forests, Golestan province, Iran. *Environ. Earth Sci.* 73, 1515–1533. <https://doi.org/10.1007/s12665-014-3502-4>.
- Pradhan, B., Lee, S., 2009. Landslide risk analysis using artificial neural network model focussing on different training sites. *Int. J. Phys. Sci.* 4, 1–15. <https://doi.org/10.5897/IJPS.9000343>.
- Python Release Python 3.8.2 [WWW Document], n.d. Python.org. URL <https://www.python.org/downloads/release/python-382/> (accessed 1.17.21).
- Qin, B., Xia, Y., Li, F., 2009. DTU: A Decision Tree for Uncertain Data, in: Theeramunkong, T., Kijisirikul, B., Cercone, N., Ho, T.-B. (Eds.), *Advances in Knowledge Discovery and Data Mining, Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg, pp. 4–15. https://doi.org/10.1007/978-3-642-01307-2_4.
- Quinlan, J.R., 1993. *C4.5: programs for machine learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, p. 302.
- Recknagel, F., 2001. Applications of machine learning to ecological modelling. *Ecol. Model.* 146, 303–310. [https://doi.org/10.1016/S0304-3800\(01\)00316-7](https://doi.org/10.1016/S0304-3800(01)00316-7).
- Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm. Remote Sens.* 67, 93–104. <https://doi.org/10.1016/j.isprsjprs.2011.11.002>.
- Rosenblatt, F., 1958. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.* 65, 386–408. <https://doi.org/10.1037/h0042519>.
- Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1985. Learning Internal Representations by Error Propagation (No. ICS-8506). California Univ San Diego La Jolla Inst For Cognitive Science.
- Sachdeva, S., Bhatia, T., Verma, A.K., 2018. GIS-based evolutionary optimized Gradient Boosted Decision Trees for forest fire susceptibility mapping. *Nat. Hazards* 92, 1399–1418. <https://doi.org/10.1007/s11069-018-3256-5>.
- Salhi, A., Benabdelouahab, S., Bouayad, E.O., Benabdelouahab, T., Larifi, I., El Mousaoui, M., Acharrat, N., Himi, M., Casas Ponsati, A., 2020. Impacts and social implications of landuse-environment conflicts in a typical Mediterranean watershed. *Sci. Total Environ.* 142853. <https://doi.org/10.1016/j.scitotenv.2020.142853>.
- Santana, V.M., Alday, J.G., Baeza, M.J., 2014. Mulch application as post-fire rehabilitation treatment does not affect vegetation recovery in ecosystems dominated by obligate seeders. *Ecol. Eng.* 71, 80–86. <https://doi.org/10.1016/j.ecoleng.2014.07.037>.
- Scarascia-Mugnozza, G., Oswald, H., Piuksi, P., Radoglou, K., 2000. Forests of the Mediterranean region: gaps in knowledge and research needs. *For. Ecol. Manage.* 132, 97–109. [https://doi.org/10.1016/S0378-1127\(00\)0383-2](https://doi.org/10.1016/S0378-1127(00)0383-2).
- Schmidt, D.A., Taylor, A.H., Skinner, C.N., 2008. The influence of fuels treatment and landscape arrangement on simulated fire behavior, Southern Cascade range, California. *For. Ecol. Manage.* 255, 3170–3184. <https://doi.org/10.1016/j.foreco.2008.01.023>.
- Shafizadeh-Moghadam, H., Valavi, R., Shahabi, H., Chapi, K., Shirzadi, A., 2018. Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping. *J. Environ. Manage.* 217, 1–11. <https://doi.org/10.1016/j.jenvman.2018.03.089>.
- Simioni, G., Marie, G., Davi, H., Martin-St Paul, N., Huc, R., 2020. Natural forest dynamics have more influence than climate change on the net ecosystem production of a mixed Mediterranean forest. *Ecol. Model.* 416, 108921. <https://doi.org/10.1016/j.ecolmodel.2019.108921>.
- Skapura, D.M., 1996. *Building Neural Networks*. Addison-Wesley Professional.

- Stambouli, H., El Bouri, A., Bellimam, M.A., Bouayoun, T., El Karni, N., 2005. Cultivation of *Cannabis sativa* L. in northern Morocco. *Bull. Narc.* 57, 79–118.
- Statnikov, A., 2011. A Gentle Introduction to Support Vector Machines in Biomedicine: Theory and methods. World Scientific.
- Tehrany, M.S., Jones, S., Shabani, F., Martínez-Álvarez, F., Tien Bui, D., 2019. A novel ensemble modeling approach for the spatial prediction of tropical forest fire susceptibility using LogitBoost machine learning classifier and multi-source geospatial data. *Theor. Appl. Climatol.* 137, 637–653. <https://doi.org/10.1007/s00704-018-2628-9>.
- Tehrany, M.S., Pradhan, B., Jebur, M.N., 2015a. Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method. *Stoch. Environ. Res. Risk Assess.* 29, 1149–1165.
- Tehrany, M.S., Pradhan, B., Mansor, S., Ahmad, N., 2015b. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *CATENA* 125, 91–101. <https://doi.org/10.1016/j.catena.2014.10.017>.
- 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, 699–720. <https://doi.org/10.1080/13658816.2012.721554>.
- Tien Bui, D., Ho, T.C., Revhaug, I., Pradhan, B., Nguyen, D.B., 2014. Landslide Susceptibility Mapping Along the National Road 32 of Vietnam Using GIS-Based J48 Decision Tree Classifier and Its Ensembles, in: Buchroithner, M., Prechtel, N., Burghardt, D. (Eds.), *Cartography from Pole to Pole, Lecture Notes in Geoinformation and Cartography*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 303–317. https://doi.org/10.1007/978-3-642-32618-9_22.
- Tien Bui, D., Hoang, N.-D., Samui, P., 2019. Spatial pattern analysis and prediction of forest fire using new machine learning approach of Multivariate Adaptive Regression Splines and Differential Flower Pollination optimization: a case study at Lao Cai province (Viet Nam). *J. Environ. Manage.* 237, 476–487. <https://doi.org/10.1016/j.jenvman.2019.01.108>.
- Tien Bui, D., Le, K.-T., Nguyen, V., Le, H., Revhaug, I., 2016. 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, 347. <https://doi.org/10.3390/rs8040347>.
- Truong, X., Mitamura, M., Kono, Y., Raghavan, V., Yonezawa, G., Truong, X., Do, T., Tien Bui, D., Lee, S., 2018. Enhancing prediction performance of landslide susceptibility model using hybrid machine learning approach of bagging ensemble and logistic model tree. *Appl. Sci.* 8, 1046. <https://doi.org/10.3390/app8071046>.
- Tsangaratos, P., Ilija, I., Hong, H., Chen, W., Xu, C., 2017. Applying Information Theory and GIS-based quantitative methods to produce landslide susceptibility maps in Nancheng County, China. *Landslides* 14, 1091–1111. <https://doi.org/10.1007/s10346-016-0769-4>.
- Vapnik, V., 1998. The support vector method of function estimation. In: Suykens, J.A.K., Vandewalle, J. (Eds.), *Nonlinear Modeling: Advanced Black-Box Techniques*. Springer, US, Boston, MA, pp. 55–85. https://doi.org/10.1007/978-1-4615-5703-6_3.
- Vapnik, V., 1963. Pattern recognition using generalized portrait method. *Autom. Remote Control* 24, 774–780.
- Vapnik, V.N., 1995. *The Nature of Statistical Learning Theory*. Springer, New York, New York, NY. <https://doi.org/10.1007/978-1-4757-2440-0>.
- Vecín-Arias, D., Castedo-Dorado, F., Ordóñez, C., Rodríguez-Pérez, J.R., 2016. Biophysical and lightning characteristics drive lightning-induced fire occurrence in the central plateau of the Iberian Peninsula. *Agric. For. Meteorol.* 225, 36–47. <https://doi.org/10.1016/j.agrformet.2016.05.003>.
- Venkatesh, K., Preethi, K., Ramesh, H., 2020. Evaluating the effects of forest fire on water balance using fire susceptibility maps. *Ecol. Indic.* 110, 105856. <https://doi.org/10.1016/j.ecolind.2019.105856>.
- Verdú, F., Salas, J., Vega-García, C., 2012. A multivariate analysis of biophysical factors and forest fires in Spain, 1991–2005. *Int. J. Wildland Fire* 21, 498. <https://doi.org/10.1071/WF11100>.
- Wotton, B.M., Martell, D.L., Logan, K.A., 2003. Climate Change and People-Caused Forest Fire Occurrence in Ontario 21.
- Yang, Q., Li, X., Shi, X., 2008. Cellular automata for simulating land use changes based on support vector machines. *Comput. Geosci.* 34, 592–602. <https://doi.org/10.1016/j.cageo.2007.08.003>.
- Zema, D.A., Nunes, J.P., Lucas-Borja, M.E., 2020. Improvement of seasonal runoff and soil loss predictions by the MMF (Morgan-Morgan-Finney) model after wildfire and soil treatment in Mediterranean forest ecosystems. *CATENA* 188, 104415. <https://doi.org/10.1016/j.catena.2019.104415>.
- Zhou, C., Yin, K., Cao, Y., Ahmed, B., 2016. Application of time series analysis and PSO-SVM model in predicting the Bazimen landslide in the Three Gorges Reservoir, China. *Eng. Geol.* 204, 108–120. <https://doi.org/10.1016/j.enggeo.2016.02.009>.
- Zidane, I., Lhissou, R., Bouli, A., Mabrouki, M., 2019. An improved algorithm for mapping burnt areas in the Mediterranean forest landscape of Morocco. *J. For. Res.* 30, 981–992. <https://doi.org/10.1007/s11676-018-0669-7>.