The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-4/W6-2022
Geoinformation Week 2022 “Broadening Geospatial Science and Technology”, 14–17 November 2022, Johor Bahru, Malaysia (online)

USING MACHINE LEARNING COUPLED WITH REMOTE SENSING FOR FOREST FIRE SUSCEPTIBILITY MAPPING. CASE STUDY TETOUAN PROVINCE, NORTHERN MOROCCO.

Mariem SEDDOUKI*[1], Mohamed BENAYAD*[1], Zakaria AAMIR[1], Mohamed TAHIRI[2], Mehdi MAANAN[1], Hassan RHINANE[1]

[1]Geosciences Laboratory, Faculty of Sciences-Ain Chock, Hassan II University, Casablanca, Morocco.
[2]Organic Synthesis Laboratory, Faculty of Sciences-Ain Chock, Hassan II University, Casablanca, Morocco.

Commission IV, WG 7

KEY WORDS: FOREST FIRE, MACHINE LEARNING, SVM, RF, XGBOOST, NDVI, GIS, RS.

ABSTRACT:

Currently there is a public awareness to protect the environment especially forest ecosystems and the forest fire dilemma has become a topic of intense research around the world. In this setting, this study evaluates forest fire susceptibility (FFS) in northern Morocco using three geographic information system (GIS) based on machine learning algorithms: XGBoost (XGB), Random Forest (RF), and Support Vector Machine (SVM). To this effect, a Geographic Information System (GIS) database was developed involving ten independent causal factors (elevation, aspect, slope, distance to roads, distance to residential areas, land cover, normalized difference vegetation index (NDVI), rainfall, temperature and wind speed) and 345 fire pixels. The 345 pixels were split into two sets for training (70%) and validation (30%) and the spatial relationships between factors affecting FFs and fire pixels was analyzed using XGB, RF, and SVM models to generate the FFS maps. The effectiveness of the models was evaluated using the receiver operating characteristic curve, the area under the curve (AUC), and several statistical measures. The results of the three models hinted that XGBoost had the highest performance (AUC = 0.856), followed by RF (AUC) = 0.827), and SVM (AUC = 0.803) in the forecasting of the forest fire. The resulting FFS maps areas can provide crucial support for the management of Mediterranean forest ecosystems and can enhance the effectiveness of planning and management of forest resources and ecological balances in these areas.

1. INTRODUCTION

In the natural world, forest fires seem to be inevitable and they play an important role in vegetation succession and landscape transformation (Tien Bui et al., 2019). Nevertheless, uncontrolled forest fires may bring about negative impacts on the environment and the local communities. These fires do not only damage human life and properties but also threaten the stability of ecosystems (Rajan1 et al., 2018). In the past decade, there has been an increasing trend in both number and severity of forest fires occurred across the globe (Catry et al., 2009; Jaafari et al., 2017; Robinne et al., 2016). This notable trend is spurring public concerns about the ecological and socioeconomic impacts of forest fires (Molina et al., 2018; Valdez et al., 2017).

Due to the impact of forest fires on socio-economic conditions and ecosystems, the prevention and suppression of forest fires have become a common interest of governments and researchers around the world (Nami et al., 2018) . To establish an effective fire prevention, it is necessary to construct the fire susceptibility maps at the regional scale. Such maps do not only facilitate the reasonable allocation of resources needed for fire prevention and suppression but also tremendously support the tasks of land use planning (Bax and Francesconi, 2018). Various methods have been proposed to model forest fire behaviors and they can be classified into three groups including physics-based method, statistical method, and machine learning methods (Tien Bui et al., 2017). Effective management requires the use of new technologies using remote sensing and geographic information systems (GIS) (Sunar and Özkan, 2001). In fire damage assessment, satellite sensor images combined with ground-based observations are the main source of information from which to determine the condition of vegetation cover (Hua and Shao, 2017). These technologies have proven their effectiveness and achieved successful results as they enable the monitoring and analysis of fires in large areas in a timely and cost-effective manner using satellite sensor imagery along with spatial analysis as provided in Geographical Information Systems (GIS) as reported by (Adab et al., 2013; Kanga and Singh, 2017; Sunar and Özkan, 2001). Recently, machine-learning approaches have achieved priority as an alternative to traditional field survey methods for predicting forest fire susceptibility by elucidating the relationship between historical fire events and various explanatory variables in order to predict future fires (Pham et al., 2020). Based on the performance comparison, it has been found that computational intelligence methods are able to provide more accurate prediction results than those of traditional statistical approaches (Tien Bui et al., 2019). There are many researches on the proposed machine learning methods used to predict forest fires that have proven their effectiveness and success, including decision tree based classification and regression (Mohajane et al., 2021), Artificial Neural Network (ANN) (Tien Bui et al., 2018), Multivariate Adaptive Regression Splines(MARS) (Tien

Bui et al., 2019), Multivariate Logistic Regression(MLP) (Pham et al., 2020), Mixture Discriminant Analysis(MDA) (Pourghasemi et al., 2020), Multilayer Perceptron Neural Network (MLP-Net) (Ngoc Thach et al., 2018), Support Vector Machine (Syifa et al., 2020), and Random Forest (Mohajane et al., 2021).

In Morocco, forests extend over an area of approximately 9 million hectares, or 12.7% of Moroccan territory at altitudes ranging from 0 to 2,700 meters, according to the latest 2021 report from the Department of Water and Forests under the Ministry, of Agriculture, Rural Development and Water and Forests (Flanagan,n.d.). As a result of the expansion of economic activities and climate change in recent years, forest fires in this country have become a serious natural hazard that destroyed vast amounts of natural resources, degraded the soil, and caused air pollution. Besides the activities of human in landuse altering, prolonged dry weather with exceptionally high temperature increases the number of fires in a large number of provinces of Morocco. Moreover, these phenomena are also observed in many other counties (Assali et al.). For example, the exceptional heat wave from July 9 to 11, 2021, favored the outbreak of 20 concomitant fires across the Kingdom and which burned 1,200 ha of forests in 10 provinces (M. Diao, n.d.).

In this study, research was conducted for northern Morocco as the most fire-affected case study nationwide. The study will assess and predict forest fire susceptibility in our area based on machine learning algorithms, GIS tools and remote sensing. The main objectives of this study are: (a) to explore the effectiveness of three forest fire forecasting models in the study area. (b) produce spatial sensitivity maps using the proposed models to identify critical areas requiring emergency response. Thus, the conclusions of this study should be a useful and effective tool to provide crucial guidance for the management of the forest ecosystem in the Tetouan region.

2. MATERIALS

2.1 Study Area

The study area is located in the region of Tanger-Tetoun-Al Hociema, in the north of Morocco (Fig. 1), it’s precisely in the province of Tetouan which is located in the extreme north of the Kingdom. It is bordered to the north by the Prefecture of M’diq- Fnideq and the Province of Fahs-Anjra, to the west by the Prefecture of Tanger-Asilah, to the south by the provinces of Larache and Chefchaouen and to the east by the Mediterranean Sea (“DRATT”, 2015). Its territory covers an area of 2,541 km², representing 0.36% of the total area of the national territory (710,850 km²).

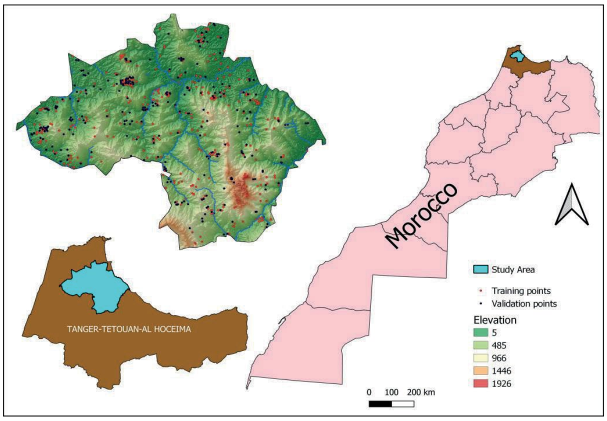

Figure. 1. Location of the study area and forest fire inventory points

Tetouan Province belongs to a region with a Mediterranean climate, marked by the existence of two different seasons. The area generally records an annual cumulative rainfall exceeding 700 mm on average. As for the temperatures, they are influenced on one side, by the action of the Mediterranean Sea and the Atlantic Ocean, and on the other side, by the altitude and the winds, in particular Chergui. In general, temperatures vary between 5.3° in cold periods and 32.9° in hot periods (HCP, 2014). The region is mainly crossed by Oued Martil and Oued Laou whose flows vary between 15 and 70 l/s. It includes two dams, one of which is large, namely the Nakhla dam (HCP, 2014). Precipitation is the highest there and varies between 800 and 1,400 mm/year sometimes reaching 2,000 mm/year with snowfall. In the Tanger- Tetoun-Al Hociema region, forests cover an area of 506,442 ha, which represents 31.63% of the forest area of the region. Therefore, the study area covers a forest area of 86,360 ha, which represents 33.98% of the area of the Tetouan province (“Regional Directorate, 2018,”n.d.) (Table 1).

Provinces	Coverage area (Ha)	Forest Area (Ha)	% of total forest area
Al Hoceima	355 000	105 290	29.65 %
Chefchaouen	333 900	210 540	63.05 %
Fahs Anjra	73 300	5 710	7.78 %
Larache	268 300	66 070	24.62 %
Ouezzane	21 300	20 192	94.798 %
M’diq-Fnideq	199 900	*	*
Tanger-Assilah	95 200	12 280	12.98 %
Tetouan	254 100	86 360	33.98 %
Tanger-Tetoun-Al Hociema region	1 601 000	506 442	31.63 %
Morocco	71 085 000	9 000 000	12.7 %

Table 1. Forest cover surfaces in northern Morocco (2017):

2.2 Data

2.2.1 Historical forest fires: The forest fire sensitivity model analyzes the correlation between relevant conditioning factors and past events. (Tehrany et al., 2019). The first step in forest fire modeling is to establish the forest fire inventory map that corresponds to past events exhibiting the target variable. In this research, information on 345 fire pixels that occurred from

January 2008 to May 2021 is used as a database of historical forest fires. These fire pixels were extracted using a fire information system for resource management (FIRMS), https://firms.modaps.cosdis.na sa.gov/, Most of the fires in the studied area occurred in months from March to August, September, and October. To improve the quality of historical forest fire data (Table 2).

Data pre-processing was performed to exclude anomalous points based on the extent of the presence of severe collateral damage and the occurrence of the fire. Of the 345 pixels, 241 pixels (70%) were randomly selected and used for model training, and the remaining 104 pixels (30%) were used for model validation. Since the forest fire risk modeling in this study is done as a binary classification, i.e. Fire class and non-fire class. The probability that a pixel belongs to a fire class was used as the fire hazard index. AS a result, the same number of fire-free spots were randomly detected. Therefore, the training data was created with 482 samples and the validation data was 207 samples. Finally, a sampling procedure was performed to derive the values of the 10 influential factors.

Month	Total number of forest fire
March	7
April	11
May	11
June	42
July	77
August	84
September	76
October	37

Table 2. Total number of forest fires that occurred during the period 2008–2021.

2.2.2 Explanatory Factors: Another key step in forest fire mapping and modeling is collecting a set of independent explanatory variables known as fire causative factors in accordance with their potential relationship with the local characteristics of the area being investigated, historical fires, and data availability. Based on previous research, there are four categories of variables that can explain forest fire modeling: topography, climate, vegetation and anthropogenic activities (Jaafari et al., 2018; Mohajane et al., 2021; Pham et al., 2020; Tehrany et al., 2019). These include; slope angle, aspect, elevation, distance to roads and residential, rainfall, temperature, wind speed, Land use, and NDVI (Table 3).

Topographic data rate is one of the most important factors to include in the fire exposure model. The effects of slope degree, aspect, and elevation on fire behavior have been covered extensively (Nami et al., 2018; Tehrany et al., 2019). 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 (Mohajane et al., 2021; Nami et al., 2018; Tehrany et al., 2019). In this study, the Digital Elevation Model (DEM) obtained from ASTER Global Digital Elevation Model (ASTGTM) with a pixel size of 30 m, sourced from the United States Geological Survey (USGS) archive (http://earthexplorer.usgs.gov), was used to derive topography data including slope, aspect, and elevation.

Conditioning factor	Unit	Source
Elevation	Meters(m)	DEM 30 m from,
Slope	Degrees(◦)	http://earthexplorer.usgs.gov
Aspect	-	
Land cover	-	Sentinel-2A at a resolution of 10m
NDVI	Ratio	https://scihub.copernicus.eu/dhus
Rainfall	(mm)	
Temperature	Degree Celsius	https://crudata.uea.ac.uk/cru/data/hrg/
Wind	m/s	https://globalwindatlas.info/
Road distance	km	Roads map
Residential distance	km	Land use map of the study

Table 3. Explanatory factors used in this study

All topographic characteristics were calculated using Surface tool in spatial analyst tools available in ArcGIS 10.3 software. Slope (Fig. 2c) was calculated and divided into five groups including (1) 0–1.5◦, (2) 1.5◦–3◦, (3) 3◦–4.5◦, (4) 4.5◦–6◦, and (5) 6◦–8.5◦, (6)8.5◦–10◦, (7) >10◦. Up-slope areas are more affected by intensive fires whereas down slopes areas are less affected. Aspect (Fig. 2d) of FFS is also an important factor. (Tien Bui et al., 2017). It indicates the orientation of a slope, which influences soil moisture, sunlight, wind and precipitation (Tien Bui et al., 2017), the main conditions that influence the behaviors of FFs in the study area. The aspect was calculated and classified into five classes such as flat zones, north, east, south and west. The elevation (Fig. 2a) was composed of five classes as 0–5 m, 5–485 m, 485–966 m, 966–1446 m, and 1446–1926 m, respectively.

Many potential sources of ignition created by human activity can affect the vulnerability of forest fires (Tehrany et al., 2019). Also, land use was chosen as a conditioning factor. Land cover information is widely recognized as a fuel proxy (Tien Bui et al., 2016). The land cover was extracted with the Sentinel 2A Multispectral Instrument (MSI) and consists of 13 bands with a resolution of 10 m extracted from the archives of the European Space Agency (ESA). (https://scihub.copernicus.eu/dhus/). Random Forest Supervised Classification Method Using Sentinel Application Platform (SNAP, version 8.0.x) applied to classify images into five class categories: water bodies, forests, croplands, building and vacant lands. The produced land cover achieved an overall accuracy of 92% (Fig. 3b).\

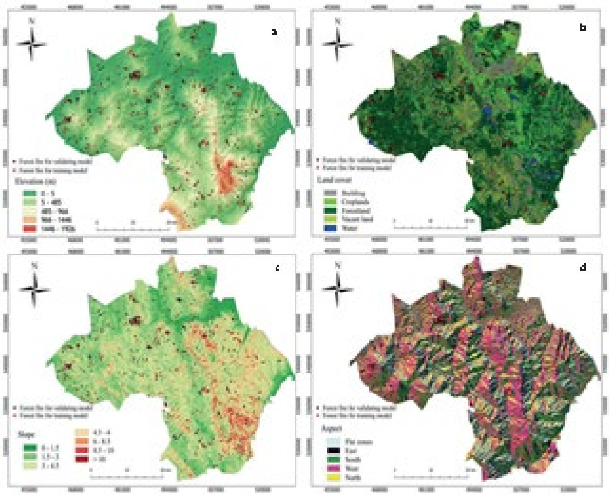

Figure 2. Fire conditioning factors (a. Elevation; b. Land Cover; c. Slope; d. Aspect).

The normalized difference vegetation index (NDVI) constituted the assessment of the vegetation cover. NDVI is the most popular index for assessment of live fuel moisture levels. The NDVI factor was calculated using the near-infrared (NIR) and visible red bands through the following equation (1).

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

The NDVI index varies between -0.996 to 0.999, the higher values representing the vegetation cover, and the lower values indicate non-vegetative cover. In our case, the NIR and RED bands are band 8 (842 nm) and band 4 (665 nm), respectively. We applied an NDVI thresholding to classify NDVI image (Fig.3e) and we generated five groups: (-0.996 – (-0.267)), (-0.267 - 0.295), (0.295 - 0.498), (0.498 - 0.710), and (0.710 - 0.999). Fires are more noticeable in forest ecosystems near roads and in populated areas due to traffic accidents and the impact of living beings on natural ecosystems which have become uncontrollable (Tien Bui et al., 2019, 2016). Therefore, in this study, we considered the distance from roads and residential areas. The distance to roads and residential areas was assessed by measuring the location FF against the vector of the street and residential areas using the distance function tool in ArcGIS. The distance to the roads (Fig. 4j) 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 residential areas were extracted from the land use map and used to generate the distance to the residential areas (Fig. 4i) into five groups including (0–1000 m), (1000–2000 m), (2000–3000 m), (3000–4000 m) and (4000–20000 m).

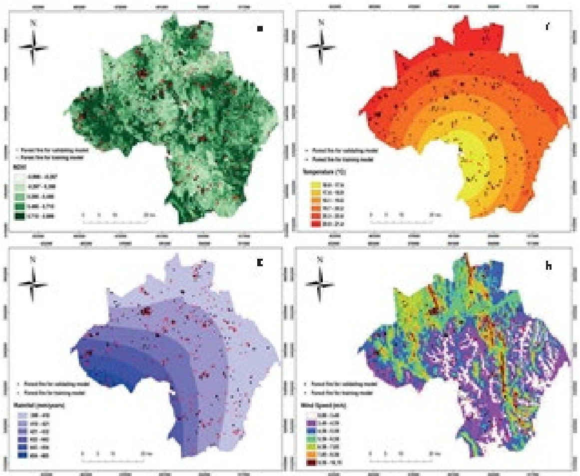

Figure 3. Fire conditioning factors (e. NDVI; f. Temperature; g. Rainfall; h.Wind Speed)

Weather conditions such as rainfall, temperature, and wind are considered the main factors that strongly affect the behavior of forest fire, in which, forest fire is more possible to occur under hot, windy and dry weather conditions (Ngoc Thach et al., 2018). In this study, thematic maps were created using meteorological data from 2011 to 2021. The wind speed map (Fig. 3h) was classified into five classes including: (0.89–3.49 m/s), (3.49–4.59 m/s), (4.59–5.59 m/s), (5.59– 6.58 m/s), (6.58 –7.85 m/s), (7.85–9.58 m/s), and (9.58 –16.15 m/s). The rainfall map (Fig. 3g) was created with six classes as (399–410 mm), (410–421 mm), (421–432 mm), (432–443 mm), (443–454 mm), and (454–465 mm). The mean temperature map (Fig. 3f) was created and classified as (16.9– 17.4 ◦C), (17.4– 18.9 ◦C), (19.1– 19.6 ◦C), (19.7– 20.2 ◦C), (20.3– 20.8 ◦C), and (20.9– 21.4 ◦C).

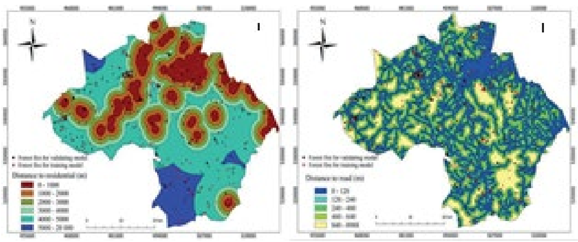

Figure. 4. Fire conditioning factors (i. Distance to Residential; j. Distance to Road).

3. THEORETICAL BACKGROUND OF THE MODELS

Based on the data collected and generated, the forest fire occurrence probability model was generated for each cell of the polygonal grid for the territory of the Tetouan province.

3.1 Random Forest (RF)

This method uses a large number of decision trees, which produce their predictions and combine them into a single, more prediction, accurate (Breiman, 2001). The RF algorithm applies a bootstrap to a subset of the observed values to build a random binary tree. The original data is randomly sampled to generate training data to build the model. In the RF model, the tree expands from different subsets of training data for greater diversity (Breiman, 1996). A subset of the data not included in the training is called the out-of-bag, which classifies from that tree to evaluate accuracy and performance and provides an internal unbiased estimate of the generalization error based on the number of calculated trees (Breiman, 2001).

3.2 XGBoost (XG)

XGBoost is a flexible, optimized and efficient distributedgradient amplification system, which allows to generate a machine learning algorithm in the context of gradient boosting. XGBoost provides a parallel tree boosting (also known as GBDT, GBM) which solve huge data science problems accurately and quickly. It was built and developed by Tianqi Chen, Ph.D. student at the University of Washington. More details about XGBoost can be found here (http://dmlc.cs.washington.edu/xgboost.html). XGB algorithm has become a dominating algorithm in the field of applied machine learning. It is used over other gradient boosting machines (GBMs) due to its fast execution speed and model performance.

3.3 Support Vector Machine (SVM)

One of the most popular machine learning algorithms is SVM, a binary classifier for supervised learning based on the principle of structural risk minimization. (Yao et al., 2008). Developed by (Vapnik, 1995) at AT& T Bell Laboratories to find a linear hyperplane that separates two classes optimally, but it can be promoted to an n-class classifier (Belousov, 2002). Since the maximum separation margin between classes is recognized by the SVM, we build a classification hyperplane in the center of the maximum margin. (Chapelle et al., 1999). SVM minimizes an upper bound of generalization error by widening the distance of the hyperplanes separating the two classes (Karimi et al., 2019). This action guarantees a low generalization error, independent of data distribution (Karimi et al., 2019). The performance of the SVM algorithm depends on the appropriate kernel functions which are polynomial kernel, sigmoid kernel, radial basis function and linear kernel (Chapelle et al., 1999). Moreover, SVM prevents overfitting in the model and ensures good generalization and classification performance (Huang et al., 2020). Both continuous and categorical variables can be handled efficiently by SVM, and it can also handle nonlinear data, complex and noisy data with outliers (Karimi et al., 2019).

4. METHODOLOGY FOR MAPPING FOREST FIRE SUSCEPTIBILITY

In calculating the forest fire susceptibility mapping, we followed these steps: A graph of the order of the systematic steps is also highlighted in (Fig. 5). Which is divided into three steps:

- Geospatial database and feature selection.
- Model Training, Layout, and Model Performance.
- Forest fire Sensitivity Maps.

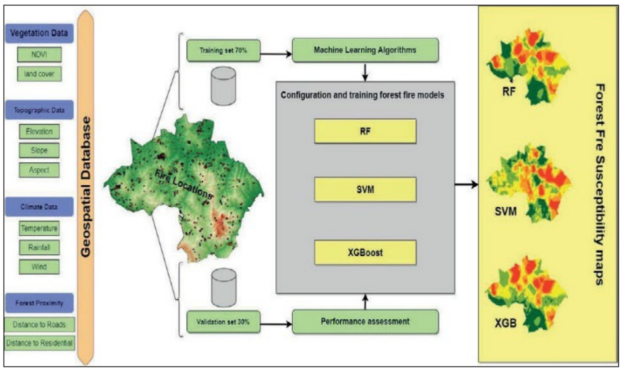

Figure. 5. Workflow of the methodology employed in this study

4.1 Establishment of fire database

As mentioned earlier, the historical forest fires and the ten variables have been collected from various sources. And 345 fire pixels was processed in ArcGIS 10.3 software. Therefore, a geographic processing has been employed to construct a GIS database for this research. The 10 adjustment variables have been converted to a raster format with a pixel size of 30. In addition, all categories of these maps have been converted to numeric format by the method developed by Tien Bui et al. (Tien Bui et al., 2017). Finally, a data normalization process was performed against 10 factors, and forest fires hedged the potential bias caused by the imbalanced magnitude of the factors. Given the fact that forest fire susceptibility represents a binary classification, the preparation of another dataset containing the non-fire pixels 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.,2019). These non-fire points were placed outside on the forest zone these areas are characterized by negative NDVI values, which indicates the presence of water bodies and poor vegetation where forest fires are almost impossible. (Tien Bui et al., 2017). It should also be mentioned that the number of non-burnt spots is equal to the number of burnt spots. A value of "1" was assigned to the fire locations and a value of "0" was assigned to the non-fire locations.

4.2 Training and Validating dataset

The validation is a mandatory step in the forecasting studies as it can be affected by a specific phenomenon. Accordingly, the fire and non-fire samples were split into one at a time of training data (70%) and one at a time of validation data (30%). (Pourtaghi et al., 2015; Tien Bui et al., 2017). Therefore, 482 fire and non-fire locations will be used to train the models, while another 207 fire and non-fire locations will be involved in the validation of fire susceptibility maps.

4.3 Configuration of the susceptibility models

The next step in the modeling process is to configure the model used to calculate the forest fire vulnerability. In this study, Python 3.8.8 was used to apply machine learning models. Hence, a sampling procedure was performed to derive the values from the 10 influential factors. and the values of each fire and non-fire pixel were converted in tabular format in order to be read in Python. Then, different amounts of hyperparameters were defined, and by running the models multiple times, the best configuration was chosen to achieve the highest accuracy for each model. For RF model the number of trees equal to 500 and the random split variable equal to 1 leads to the highest accuracy with the minimum time to obtain the results. The advantage of RF over other models is the smaller number of hyperparameters to be set. For the XGBoost model the binary: logistic loss function and the number of trees was set to 100 to reach the highest accuracy. For SVM model, the radial basic function (RBF) kernel with the kernel width equal to 1 and the regularization value of 1000 had the highest accuracy.

4.4 Validation of Metrics

One of the important steps after developing a model is to evaluate its training and predictive performance (Pham et al., 2020). In this study, to evaluate and compare the models developed for forest fire susceptibility mapping we used the receiver operating characteristic (ROC) curve. Receiver operating characteristic (ROC) curves are used to evaluate the performance of forest fire sensitivity modeling. (Tehrany et al., 2019). The ROC curve used to visualize the performance of binary classification is plotted with sensitivity as the y-axis and 1-specificity as the x-axis. (Tien Bui et al., 2017). The AUC score is the area below the ROC curve and its value ranges from 0 to 1. The highest AUC value indicates the ideal measure of separability, and the lowest AUC value indicates the lowest measure of separability. Additionally, we used several statistical measures like specificity, overall accuracy, precision, sensitivity, true positive (TP), true negative (TN), false positive (FP), false negative (FN), negative predictive value (NPV), and positive predictive value (PPV). The following equations provide a brief description of each metric:

$$\text{OverallAccuracy} = TP + TN / TP + TN + FP + FN \quad (2)$$

$$\text{Specificity} = TN / FP + TN \quad (3)$$

$$\text{Sensitivity} = TP / TP + FN \quad (4)$$

$$\text{Precision} = TP / TP + FP \quad (5)$$

$$PPV = TP / FP + TP \quad (6)$$

$$NPV = TN / FN + TN \quad (7)$$

4.5 Fire susceptibility mapping and models performance

Based on the results obtained, a model was adopted in the study area to describe the vulnerability to forest fires. To generate forest fire susceptibility maps, the resulting layers were then changed to a GIS environment. The five affect classes were very low, low, moderate, high and very susceptible to forest fire (Fig. 7) were applied (Jaafari et al., 2019; Tien Bui et al., 2017) using Gaussian process in ArcGIS 10.3 (Tehrany et al., 2019).\

(NPV) = 0.804 meaning the probability of predicting pixels at non-forest fire is 80%. 80.4%.

The forest fire susceptibility map was plotted by dividing the FSM-SVM values into five classes using the Gaussian process (Fig. 7). Very low fire sensitivity zones occur in approximately 26.74% of the study area, the low susceptibility is spread on 8.09% of the total analyzed territory, the medium FSM-SVM values are present on 11.09%, while about 54% of the study area represents high and very high susceptibility.

5. RESULTS AND DISCUSSION

5.1 Fire susceptibility mapping and models performance

5.1.1 Random Forest (RF): The prediction power of the RF model was assessed using the validation dataset that was not used during the training phase (Table 4). The following values for the statistical metric were obtained: Accuracy = 0.80, Precision = 0.891, Sensitivity = 0.713, Specificity = 0.891, Positive predictive value (PPV) = 0.891 meaning that the probability of predicting pixels to forest fires is 89.1%, and Negative predictive value (NPV) = 71.3% meaning that the probability of predicting pixels to non-forest fires is 71.3%. It should be mentioned that overall the RF model achieved the second-best performance after XGBoost model. The fire sensitivity map was derived from the RF model (FSM-RF) after evaluating the model performance (Fig. 7). The values of FSM-RF were grouped into five classes using the Gaussian process. The first class, highlighting the areas with very low susceptibility, accounts around 22.48% of the study area, the low fire susceptibility is presented at approximately 22.70%, while the medium class of FSM-RF occupies a surface equal to 23.90% of the research zone. High fire susceptibility and very high fire susceptibility represent a total of 30.92% of the study area

5.1.2 XGboost (XG): After the training of the XGBoost model, its performance was measured with the help of several statistical metrics (Table 4). Thus, in terms of training sample, the accuracy of 0.859 was the highest between all the applied models. The involvement of validating sample in the assessment of models performance revealed that XGBoost model achieved also the best results highlighted by the following values: Accuracy = 0.827, Precision = 0.816, Sensitivity = 0.832, Specificity = 0.821, Positive predictive value (PPV) = 0.815 meaning that the probability of predicting pixels to forest fires is 81.5%, and Negative predictive value (NPV) = 0.837 meaning that the probability of predicting pixels to non-forest fires is 83.7%. Once the model performance was assessed, the mapping of the Fire Susceptibility Index was performed (Fig. 7). Thus, the very low susceptibility has 16.11% of the total study area, the low susceptibility is present on 26.31% of the analyzed zone, the medium values are encountered on approximately 25.57%, while the areas exposed in a high and very high degree to fire occurrence can be found on around 32.01% of the study area.

Evaluation Metrics	XGBosst	RF	SVM
True Positive	84	83	80
True Negative	87	80	86
False Positive	19	9	20
False Negative	17	35	21
Accuracy	0.827	0.80	0.802
Precision	0.816	0.891	0.80
Sensitivity	0.832	0.713	0.792
Specificity	0.821	0.891	0.812
Positive predictive value %	81.5	89.1	80
Negative predictive value %	83.7	71.3	80.4
AUC	0.856	0.827	0.803

Table 4. Model performances estimated with validating samples.

5.1.3 Support Vector Machine (SVM: The application of SVM model is characterized by the following performances in terms of validating dataset (Table 4):

Accuracy = 0.802, Precision = 0.80, Sensitivity = 0.792, Specificity = 0.812, Positive Predictive Value (PPV) = 0.80 meaning the probability of predicting pixels at forest fire is 80%, and Negative Predictive Value.

5.2 Results validation

The results validation was done by using the ROC Curve. In this regard, the validating data were employed and their plots are shown in Fig. 6. Thus, the highest AUC value (0.856) was achieved by XGBoost model followed by RF (0.827), and SVM (0.803). Given the fact that all the models, achieved AUC values higher than 0.7, we can assume that the applied algorithms were performant concerning the identification of areas susceptible to forest fire occurrence.

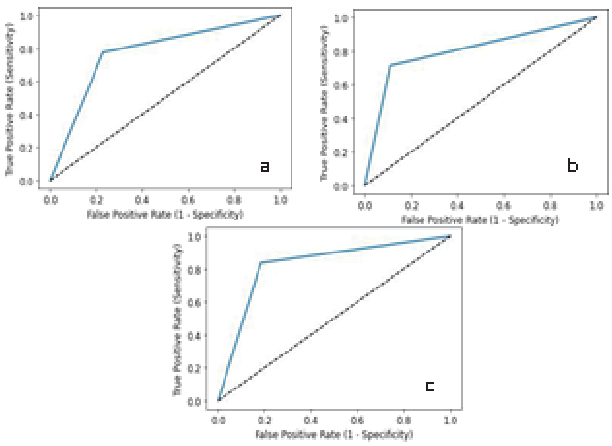

Figure. 6. ROC Curve (a. SVM; b. RF; c. XGBoost)

5.3 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). These models classify the study area into 5 levels of risk. The outcomes from this study could be served as a benchmark to identify areas, which require emergency intervention.

One of the advantages of RF method is its ability to estimate the importance of the features used for modeling (Table 5). Fig.8 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 elevation, rainfall, distance to residential areas, wind speed, temperature and aspect. Distance to roads 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, (Pham et al., 2020) suggested that proximity to roads and residence areas, land use, elevation, and annual temperature intensifies the likelihood and frequency of fire ignitions. Similar to this finding, in a study developed by (Pourghasemi et al., 2020) the most important factors were land cover, slope, annual mean rainfall, and elevation for forest fire susceptibility.

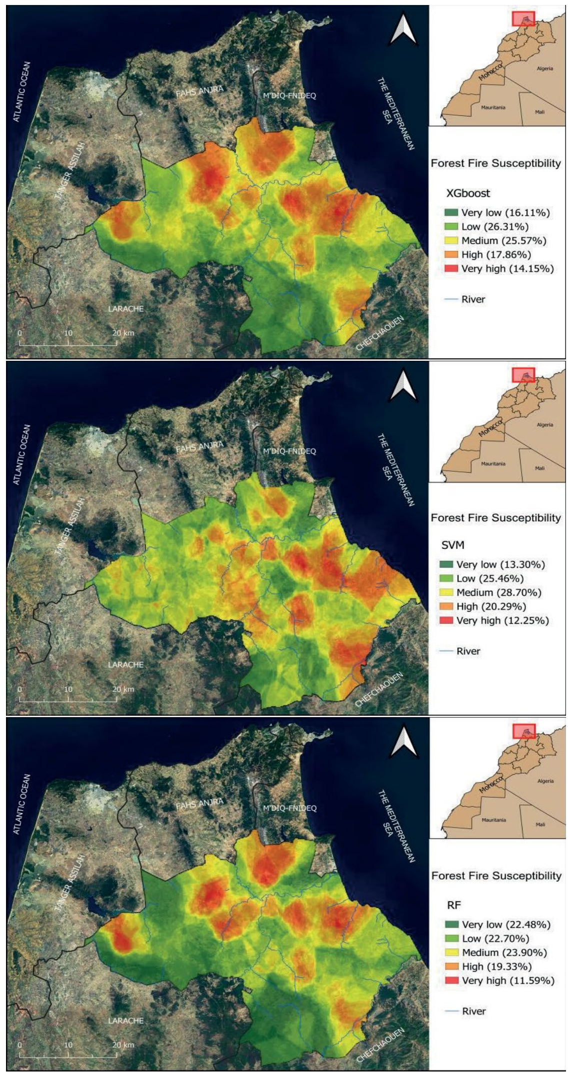

Figure. 7. Forest Fire Susceptibility using the 3 ensemble models.

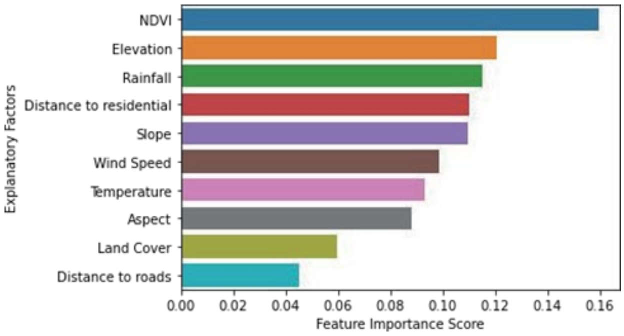

Figure. 8. The feature importance factors.

Factors	Scores
NDVI	0.159624
Elevation	0.120555
Rainfall	0.115202
Distance to residential	0.110238
Slope	0.109800
Wind Speed	0.098787
Temperature	0.093133
Aspect	0.088099
Land Cover	0.059420
Distance to roads	0.045141

Table 5. Scores of importance conditioning factors.

In this research, the XGBoost model achieved the highest accuracy among other models, and these results are consistent with previous research (Ghatkar et al., 2019; Zamani Joharestani et al., 2019). These authors stated that the XGBoost has better control against overfitting by using more regularized model formalization, in comparison to prior algorithms. The outperformance of XGBoost is due to its advantages in parallelization in tree building through the use of the CPU cores during training. Indeed, eXtreme Gradient Boosting algorithm has been employed with promising results in other prior research related to natural hazards, for example for Prediction of gully erosion susceptibility mapping (Arabameri et al., 2021), in their paper proved that XGBoost achieved high accuracy among other algorithms applied in their case study.

As discussed above, XGBoost is followed by RF model in term of accuracy. The RF can be applied for both classification and regression task, and it is fast, easy to use and processing large dataset (Zamani Joharestani et al., 2019). Another essential advantage for RF is that it can handle and deal with nonlinearities between variables (Mohajane et al., 2021). RF has been widely used in natural hazards assessment proving successful results. For example, flood mapping (Thanh Son et al., 2021) and landslide susceptibility (Taalab et al., 2018).

SVM 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 shown successful results in several fields like flood mapping (Tehrany et al., 2014). Also, (Mohajane et al.,2021; Tien Bui et al., 2017) reported the predictive performance for both RF and SVM for forest fire prediction. Selecting an appropriate ML algorithm for forest fire models as for any other environmental risk assessment is a complex task as each model has its advantages and disadvantages (Tien Bui et al., 2017). Therefore, all the models mentioned above in this study are recommended for forest fire sensitivity studies.

A number of natural and anthropogenic disturbances significantly affects forest ecosystems in northern Morocco, for example, in a recent research, (Salhi et al., 2021) reported the fragility of soil quality, thus, the situation requires anti-erosion activities. In addition, the most landscape areas in northern Morocco are exploited for cannabis plantations (El Motaki et al., 2019) due to their economic value. All these factors increase the spread of forest fire disaster. This research highlights the urgent need for government agencies to work towards preserving forest ecosystems and planning mitigation strategies, as well as the need to implement a public warning system to secure the urgent status of the forest ecosystem.

6. CONCLUSION

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 research 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 to assist forest managers and local authorities in forest management and fire suppression. In this study , three models have been developed namely, RF, SVM, and XGBoost for forest fire modelling based on 345 forest fire locations and a total of 10 forest fire conditioning factors (elevation, aspect, slope, distance to roads, distance to residences, normalized difference vegetation index (NDVI), precipitation, temperature, wind speed and land cover). The results of the proposed model show that XGBoost has high performance (AUC = 0.856), followed by RF (AUC = 0.827), and SVM (AUC = 0.803), When predicting forest fires, or given the results obtained, the maps created here can be very useful management tools for developing and analyzing forest fire strategies and management. In addition, this methodology may be compatible with other fields that present similar problems.

REFERENCES

Friedman, J.H., 2002. Stochastic gradient boosting. Computational Statistics & Data Analysis 38, 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2
Karimi, F., Sultana, S., Shirzadi Babakan, A., Suthaharan, S., 2019. An enhanced support vector machine model for urban expansion prediction. Computers, Environment and Urban Systems 75, 61–75. https://doi.org/10.1016/j.compenvurbsys.2019.01.001

Vapnik, V., n.d. 3 THE SUPPORT VECTOR METHOD 31.
Taalab, K., Cheng, T., Zhang, Y., 2018. Mapping landslide susceptibility and types using Random Forest. Big Earth Data 2, 159–178. https://doi.org/10.1080/20964471.2018.1472392

Tehrany, M.S., Pradhan, B., Jebur, M.N., 2014. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. Journal of Hydrology 512, 332–343. https://doi.org/10.1016/j.jhydrol.2014.03.008
Tehrany, M.S., Pradhan, B., Jebur, M.N., 2013. Spatial prediction of flood

susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. Journal of Hydrology 504, 69–79. https://doi.org/10.1016/j.jhydrol.2013.09.034

Belousov, Applicational aspects of support vector machines https://onlinelibrary.wiley.com/doi/abs/10.1002/cem.744
2002 accesVapnik, 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
Yao, X., Tham, L.G., Dai, F.C., 2008. Landslide susceptibility mapping based on Support Vector Machine: A case study on natural slopes of Hong Kong, China. Geomorphology 101, 572–582.
https://doi.org/10.1016/j.geomorph.2008.02.011sed 4.28.20.

Nami, M.H., Jaafari, A., Fallah, M., Nabiuni, 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

Chapelle, O., Haffner, P., Vapnik, V.N., 1999. Support vector machines for histogram-based image classification. IEEE Trans. Neural Netw. 10, 1055–1064. https://doi.org/10.1109/72.788646

Mohajane, M., Costache, R., Karimi, F., Bao Pham, Q., Essahlaoui, A., Nguyen, H., Laneve, G., Oudija, F., 2021. Application of remote sensing and machine learning algorithms for forest fire mapping in a Mediterranean area. Ecological Indicators 129, 107869.
https://doi.org/10.1016/j.ecolind.2021.107869

Kanga, S., Singh, S.K., 2017. Forest Fire Simulation Modeling using Remote Sensing & GIS. International Journal of Advanced Research in Computer Science 8.
Syifa, M., Panahi, M., Lee, C.-W., 2020. Mapping of Post-Wildfire Burned Area Using a Hybrid Algorithm and Satellite Data: The Case of the Camp Fire Wildfire in California, USA. Remote Sensing 12, 623. https://doi.org/10.3390/rs12040623 Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L., 1996. Bagging predictors. Mach Learn 24, 123–140. https://doi.org/10.1007/BF00058655

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. Environmental Research 184, 109321. https://doi.org/10.1016/j.envres.2020.109321

Salhi, A., Benabdelouahab, S., Bouayad, E.O., Benabdelouahab, T., Larifi, I., El Mousaoui, M., Acharrat, N., Himi, M., Casas Ponsati, A., 2021. Impacts and social implications of landuse-environment conflicts in a typical Mediterranean watershed. Science of The Total Environment 764, 142853. https://doi.org/10.1016/j.scitotenv.2020.142853

Pham, B.T., Jaafari, A., Avand, M., Al-Ansari, N., Dinh Du, T., Yen,
H.P.H., Phong, T.V., Nguyen, D.H., Le, H.V., Mafi-Gholami, D., Prakash, I., Thi Thuy, H., Tuyen, T.T., 2020. Performance Evaluation of Machine Learning Methods for Forest Fire Modeling and Prediction. Symmetry 12, 1022. https://doi.org/10.3390/sym12061022

Zamani Joharestani, M., Cao, C., Ni, X., Bashir, B., Talebiesfandarani, S., 2019. PM2.5 Prediction Based on Random Forest, XGBoost, and Deep Learning Using Multisource Remote Sensing Data. Atmosphere 10, 373. https://doi.org/10.3390/atmos10070373

Taalab, K., Cheng, T., Zhang, Y., 2018. Mapping landslide susceptibility and types using Random Forest. Big Earth Data 2, 159–178. https://doi.org/10.1080/20964471.2018.1472392

Arabameri, A., Chandra Pal, S., Costache, R., Saha, A., Rezaie, F., Seyed Danesh, A., Pradhan, B., Lee, S., Hoang, N.-D., 2021. Prediction of gully erosion susceptibility mapping using novel ensemble machine learning algorithms. Geomatics, Natural Hazards and Risk 12, 469–498. https://doi.org/10.1080/19475705.2021.1880977.

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

Deka, B., Maji, P., Mitra, S., Bhattacharyya, D.K., Bora, P.K., Pal, S.K. (Eds.), 2019. Pattern Recognition and Machine Intelligence: 8th International Conference, PReMI 2019, Tezpur, India, December 17-20, 2019, Proceedings, Part I, Lecture Notes in Computer Science. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-34869-4

Holsinger, L., Parks, S.A., Miller, C., 2016. Weather, fuels, and topography impede wildland fire spread in western US landscapes. Forest Ecology and Management 380, 59–69. https://doi.org/10.1016/j.foreco.2016.08.035

Syifa, M., Panahi, M., Lee, C.-W., 2020. Mapping of Post-Wildfire Burned Area Using a Hybrid Algorithm and Satellite Data: The Case of the Camp Fire Wildfire in California, USA. Remote Sensing 12, 623. https://doi.org/10.3390/rs12040623 Sunar, F., Özkan, C., 2001. Forest fire analysis with remote sensing data.International Journal of Remote Sensing 22, 2265–2277.
https://doi.org/10.1080/01431160118510

Jaafari, A., Panahi, M., Pham, B.T., Shahabi, H., Bui, D.T., Rezaie, F., Lee, S., 2019. 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

DRATT, 2015 DRATT (Direction Régionale de l'Agriculture de Tanger-Tétouan) Monographie de la région du nord du Maroc Rapports et Données statistiques sur l'agriculture et l'élevage , nd (2015)

HCP, (Haut-Commissariat au Plan), 2014 HCP, (Haut-Commissariat au Plan), 2014. Monographie régionale de Tanger-Tétouan.
Costache, R., Pham, Q.B., Sharifi, E., Linh, N.T.T., Abba, S.I., Vojtek,
M., Vojteková, J., Nhi, P.T.T., Khoi, D.N., 2019. Flash-Flood Susceptibility Assessment Using Multi-Criteria Decision Making and Machine Learning Supported by Remote Sensing and GIS Techniques. Remote Sensing 12, 106. https://doi.org/10.3390/rs12010106

Tien Bui, D., Bui, Q.-T., Nguyen, Q.-P., Pradhan, B., Nampak, H., Trinh, P.T., 2017. A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. Agricultural and Forest Meteorology 233, 32–44. https://doi.org/10.1016/j.agrformet.2016.11.002

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

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

Catry, F.X., Rego, F.C., Bação, F.L., Moreira, F., Catry, F.X., Rego, F.C., Bação, F.L., Moreira, F., 2009. Modeling and mapping wildfire ignition risk in Portugal. Int. J. Wildland Fire 18, 921–931. https://doi.org/10.1071/WF07123

Flanagan, R., n.d. National Tree Day. Green-Schools. https://greenschoolsireland.org/national-tree-day/ (accessed 1.24.22).

Sunar, F., Özkan, C., 2001. Forest fire analysis with remote sensing data.International Journal of Remote Sensing 22, 2265–2277. https://doi.org/10.1080/01431160118510

Adab, H., Kanniah, K.D., Solaimani, K., 2013. Modeling forest fire risk in the northeast of Iran using remote sensing and GIS techniques. Nat Hazards 65, 1723–1743. https://doi.org/10.1007/s11069-012-0450-8

Robinne, F.-N., Parisien, M.-A., Flannigan, M., 2016. Anthropogenic influence on wildfire activity in Alberta, Canada. Int. J. Wildland Fire 25, 1131–1143. https://doi.org/10.1071/WF16058

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

Ngoc Thach, N., Bao-Toan Ngo, D., Xuan-Canh, P., Hong-Thi, N., Hang Thi, B., Nhat-Duc, H., Dieu, T.B., 2018. Spatial pattern assessment of tropical forest fire danger at Thuan Chau area (Vietnam) using GIS-based advanced machine learning algorithms: A comparative study. Ecological Informatics 46, 74–85. https://doi.org/10.1016/j.ecoinf.2018.05.009

M. Diao, n.d. Feux de forêts au Maroc: les leçons à tirer pour l’avenir [WWW Document]. http://fnh.ma. URL https://fnh.ma/article/developpement-durable/feux-de-forets-au-maroc-les-lecons-a-tirer-pour-l-avenir (accessed 1.22.22).

Molina, J.R., Moreno, R., Castillo, M., Rodríguez y Silva, F., 2018. Economic susceptibility of fire-prone landscapes in natural protected areas of the southern Andean Range. Science of The Total Environment 619–620, 1557–1565. https://doi.org/10.1016/j.scitotenv.2017.11.233