Application of Machine Learning Algorithms for Forest Fire Mapping

Forest fire disaster is considered as one of the main causes of dramatic depletion of the forest ecosystems worldwide due to both anthropogenic or natural processes. 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. The impacts include deterioration of soil ecology and water hydrology, and land degradation and soil erosion. 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, and it acts as a selective factor of the traits of plants. 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.

Creating forest fire susceptibility maps requires the establishment of forest fire inventory as the target variable and some explanatory variables. Therefore, five machine learning algorithms NN, LR, DT, SVM, and RF are used. The research was conducted for the north of Morocco as the case study which is most affected by the fire at the national scale.

Data

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. Historical data provided by (HCEFLCD) and previous reports, were used for preparing the forest fire inventory map of the study area. The independent variables for forest fire modeling including slope angle, aspect, elevation, distance to roads and residential, rainfall, temperature, wind speed, Land use, and NDVI.

independent variables

Conditioning factor	Unit
Slope	Degrees (°)
Elevation	Meters (m)
Aspect	–
Land cover	–
NDVI	Ratio
Rainfall	(mm)
Temperature	Degree Celsius
Wind	m/s
Road distance	km
Residential distance	km

Logistic Regression (LR)

LR, a statistical model for solving binary problems, was developed by. 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. 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. 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. 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 .

Decision Tree (DT)

DT 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 DT decision tree algorithm are the Gini Index and towing criteria. The splitting procedure maximizes the homogeneity of subsets by determining the proper features and their corresponding thresholds. DT 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. The stopping rules make the DT 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. DT is able to identify the most important features. Based on the importance of the features, DT divides a complex dataset into smaller subsets hence it creates simple solutions which can be understood and interpreted easily. Moreover, DT 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, makes no statistical assumptions, analyzes data in different measurement scales without the need to normalization, handles missing data and outliers and makes good visualizations of the relationships between the features.

Random Forest (RF)

RF, a non-parametric supervised method applied for both classification and prediction , was developed by . RF is composed of a combination of decision trees . 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 . In RF models trees are growing from different training data subsets to increase the diversity as a result, greater classifier stability is achieved . 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 . 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, Gini Index , 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 . 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 .

Support Vector Machine (SVM)

SVM developed at AT&T Bell Laboratories , is used for both classification and regression tasks and then, it was enhanced 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. SVM finds an optimal hyperplane for classification by projecting the input data into the higher-dimensional Hilbert space. SVM minimizes an upper bound of generalization error by widening the distance of the hyperplanes separating the two classes. This action guarantees a low generalization error, independent of data distribution. In SVM models a penalty parameter beyond the margin of the hyperplanes is incorporated to consider misclassification errors. 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. SVM addresses non-linearity uses a mathematical function called kernel trick through to explore the data in a higher-dimensional space. 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. Furthermore, SVM prevents overfitting in the model and assures good generalization and classification performance. 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.

Training and testing dataset

The testing 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 testing dataset (30%). 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.

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 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 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 Figure.

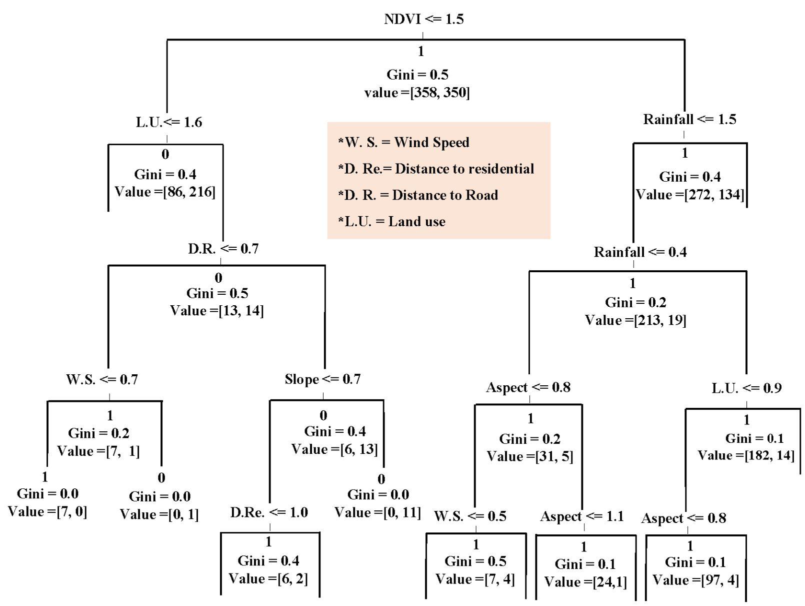

Figure. Structure of the RF model used for forest fire susceptibility in this study

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 0.001were set. However, in this model changing the hyperparameters did not change the accuracy meaningful.

For the NN model 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.

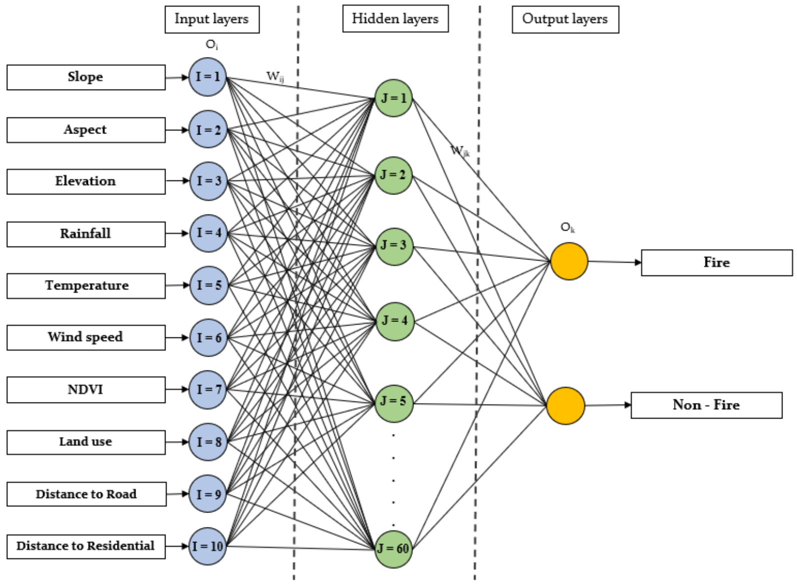

Figure. Architecture of MLP used in this study

For the DT 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. Figure illustrates the structure of the DT used for modelling forest fire in this research.

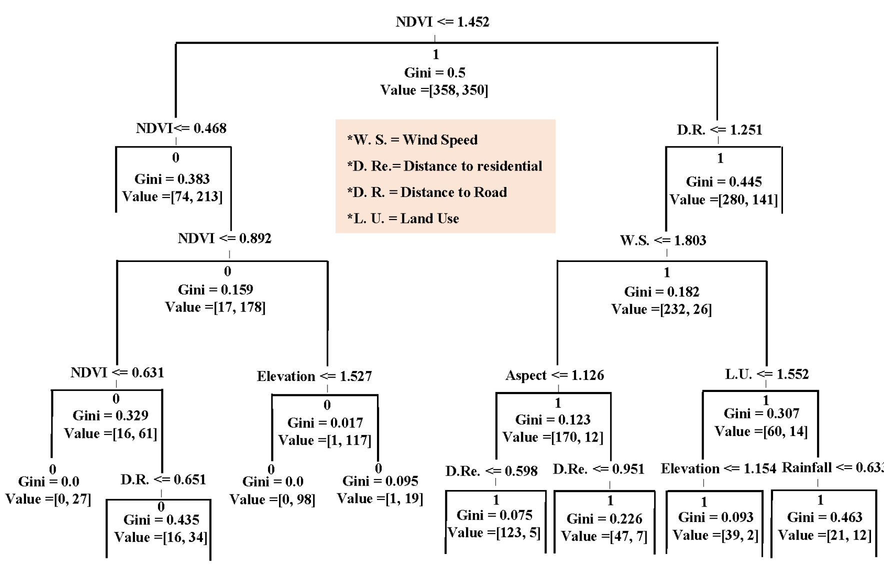

Figure. Structure of the DT model used for forest fire susceptibility in this study

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, The ROC curve is plotted with the sensitivity as the y-axis and the 1-specificity as the x-axis. 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. These statistical criteria were calculated based on the following equations:

$$Overall\ Accuarcy = \frac{TP+TN}{TP+TN+FP+FN}$$

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

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

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.

Results

NN

The statistical metrics were involved in the estimation of NN performance. 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 NN model (FSM-NN) after the model performance assessment. The values of FSM-NN 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-NN 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.

DT

The same statistical measures were used to assess the performance of DTensemble 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 DTensemble.

Following the performance evaluation, the fire susceptibility map (FSM-DT-FR) was computed. Similar to the previous case, the FSM-DTvalues 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.

LR

The performances achieved by LRensemble 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 (Figure 8), 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.

SVM

The application of SVMensemble is characterized by the following performances in terms of training dataset: 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 SVMensemble achieved the second-best performance after RFmodel.

The Forest Fire Susceptibility Map was represented by dividing the FSM-SVMvalues into five classes using the *Natural Breaks* method (Figure 8). The very low fire susceptibility appears on approximately 26.74% of the study zone, the low susceptibility is spread on 8.09% of the total analyzed territory, the medium FSM-SVMvalues 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.

RF

After the training of the RFensemble, its performance was measured with the help of several statistical metrics. 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 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. 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.

Table Model performances estimated with training and validating samples

Models	Sample	TN	TN	FP	FN	Accuracy	Precision	Sensitivity	Specificity	K-index
RF	Training	353	353	1	1	0.997	0.997	0.997	0.997	0.994
	Testing	138	135	13	16	0.904	0.914	0.896	0.912	0.808
SVM	Training	318	326	36	28	0.910	0.898	0.919	0.901	0.819
	Testing	127	132	24	19	0.858	0.841	0.870	0.846	0.715
LR	Training	267	275	87	79	0.766	0.754	0.772	0.760	0.531
	Testing	118	114	33	37	0.768	0.781	0.761	0.776	0.536
NN	Training	303	322	51	32	0.883	0.856	0.904	0.863	0.766
	Validating	121	134	30	17	0.844	0.801	0.877	0.817	0.689
DT	Training	321	316	33	38	0.900	0.907	0.894	0.905	0.799
	Validating	120	130	31	21	0.828	0.795	0.851	0.807	0.656

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 Figures 9a and 9b respectively. Thus, in terms of success rate, the highest AUC value (0.998) was achieved by RFensemble followed by SVM(0.973), DT(0.967), NN (0.947), and LR(0.822). The RFensemble was again the most performant in terms of Prediction Rate with an AUC of 0.952, followed by SVM(0.928), NN (0.918), DT (0.905), and LR(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.

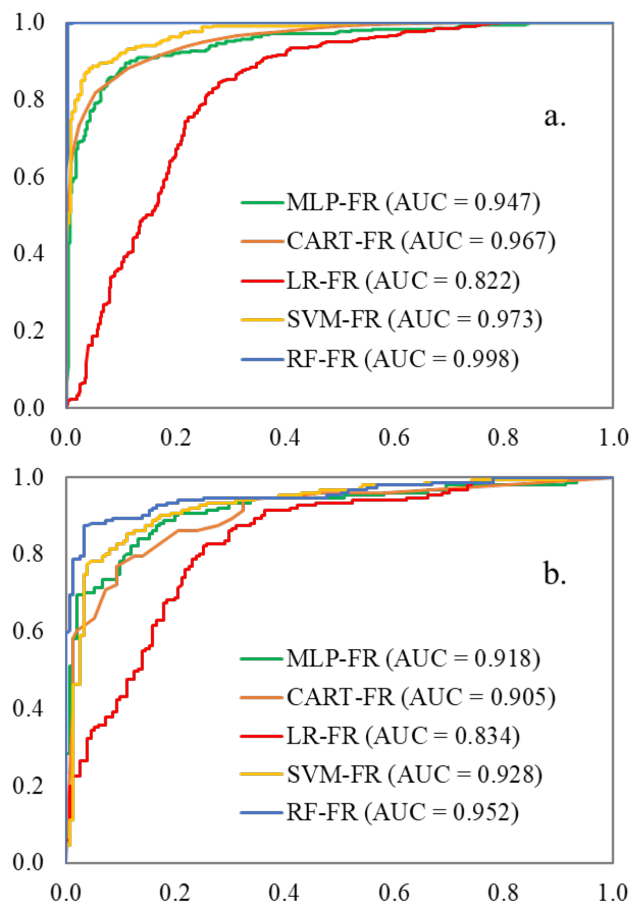

Figure ROC Curve (a. Success Rate; b. Prediction Rate)

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, 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 MLP,LR,DT,SVM andRF 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. Figure 14 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. However, emphasized that land use, annual mean rainfall, slope, and elevation are the best factors for the forst fire susceptibility. Similar to this finding, in a study developed by the most important factors were soil type, average annual temperature, and land-cover for forest fire susceptibility.

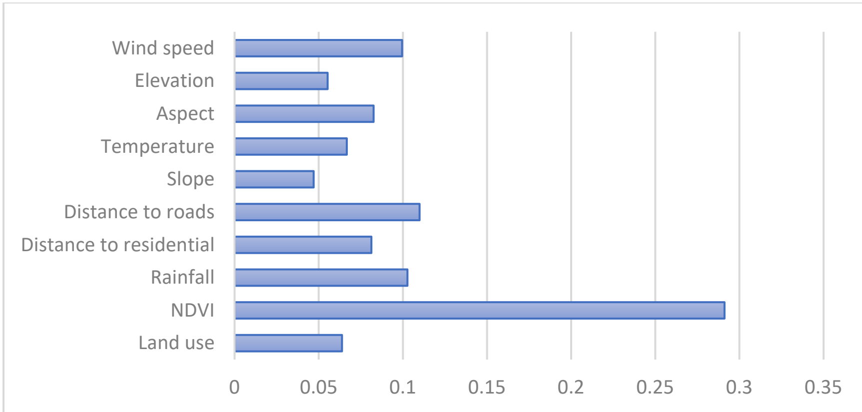

Figure The importance of conditioning factors

In this paper, the RF model achieved the highest accuracy among other models, these findings are consistent with prior research. 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, in their paper proved that RF achieved high accuracy among other algorithms applied in their case study.

As discussed above, FR is Followed by SVMmodel 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. Due to these advantages, SVM has been widely used in natural hazards assessment proving successful results. Also, 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).

DT has interesting advantages, Nevertheless, it is worthy to note that the DT algorithm is rarely present in the literature review for forest fire susceptibility and in general for moderate prediction purposes.

LR achieved also good accuracy. similarly, it has shown successful results in other fields, for example, flood mapping and landslide susceptibility. Overall, the hybrid models based on ML algorithms and bivariate statistical methods provided promising and more performance results in comparison to individual models.

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. 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, 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 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 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.