Prediction performance of Frequency Ratio, Weighting Factor, Weight of Evidence and Logistic Regression models in landslide susceptibility mapping: a case study of Ourika Basin, Marrakech High Atlas, Morocco

Namous Mustapha[1*], Ouayah Mustapha[1], Krimissa Samira[1], Jadoud Mourad[2] and Ourribane Malika[1]

1. Laboratory of Biotechnology and Sustainable Development of Natural Resources, Polydisciplinary Faculty, Sultan Moulay Slimane University, Mghila B.P. 592, Beni Mellal, MOROCCO

2. Laboratory Geosciences and Environmental Techniques, Faculty of Sciences El Jadida, Chouaïb Doukkali University, MOROCCO

*mustaphanamous@yahoo.fr

Abstract

This research aims to assess and compare Frequency Ratio, Weighting Factor, Weight of Evidence and Logistic Regression models for landslide susceptibility mapping using Geographic Information Systems and Remote Sensing data in the Ourika watershed, Morocco. A set of 100 landslides were identified and mapped by evaluating observations from satellite images (Google Earth images) and fieldwork undertaken from 2010 to 2018. The landslide inventory data was arbitrarily divided into two groups for training (75%) and validation (25%). Thirteen landslide conditioning factors were selected for landslide susceptibility modelling, based on multicollinearity analyses and the information gain method. Validation of the results is based on statistical rules for the Spatial Effective Method, Statistical Measures and Receiver Operating Characteristics Curve (ROC).

The validation results show that all four models exhibit reasonably good performance and that the Logistic Regression model shows the most stable and judicious results for landslide susceptibility mapping in the study area. In regards to conditioning factors contribution to landslides, different methods show that topographic factors had the most impact on landslides occurrence in the study area.

Keywords: Landslide susceptibility, Statistical Models, Ourika watershed, Statistical Measures.

Introduction

Landslides represent mass movements of the land surface and represent dangerous phenomena affecting life, property and environmental degradation worldwide, particularly in mountain areas[4,8,13,33,29]. Landslides are considered as one of the most common natural hazards because they influence settlements and engineered structures around the world[25].

In terms of risk to property and persons, landslides rank alongside other natural disasters such as earthquakes, floods and storms. In order to reduce the damage caused by landslides, we should first determine the affected areas and then evaluate the probability of landslide occurrence.

Natural hazard mapping identifies the previous occurrence of natural events (such as landslides, floods, earthquakes and volcanic eruptions) and includes information about possible future occurrences[47]. During recent decades, progress in Geographical Information Systems (GIS) and remote sensing techniques has facilitated landslide susceptibility and hazard mapping[25,26].

In the literature, some researchers have cited several methods for mapping and assessing landslide hazards[26, 34]. The most used methods in landslide susceptibility modelling can be grouped into three approaches: qualitative factor overlay, statistical models and geotechnical process models[14].

In addition, many researchers have chosen statistical methods, the most used being: Frequency Ratio Model[4,8,29,49], Information Value Model and Statistical Index Method[4,36], Weight of Evidence Model[4,21,22,40,48], Multivariate Logistic Regression Model[4,34,48], Bivariate Statistical Analysis (BS)[48], Fuzzy Logic Method[30,39] and Artificial Neural Network Method[6,9,12,41].

Recently, numerous innovative approaches and advanced machine learning models are being applied to susceptibility assessment, such as logistic model tree (LMT), random forest (RF) and classification and regression tree[31]. Moreover, several hybrid methods have shown good results in the prediction of landslides in several regions of the world[15]. Hence, it is very important to examine and compare machine learning methods and statistical methods to reach judicious conclusions for landslide susceptibility assessment.

Accordingly, this study proposes to evaluate and compare the performance of several statistical methods including the Frequency Ratio Model, Weighting Factor Method, Weight of Evidence Model and a Machine Learning algorithm that is Logistic Regression model, for the spatial prediction of landslides in the Ourika watershed, Hight Atlas, Morocco. Thus, thirteen landslide conditioning factors were considered in this study and the calculation and modeling were performed in a GIS environment. The results were validated via the area under the receiver operating characteristic (ROC) statistical measures and statistical rules for the spatial effective method.

The contribution of this study can be summarized in three main points:

(1) The Frequency Ratio (FR), Weighting Factor (WF), Weight of Evidence (WOE) and Logistic Regression (LR) models were applied and their results were compared and validated by three methods, which are the area under the receiver operating characteristic (ROC) statistical measures and statistical rules for the spatial effective method.
(2) IG method was applied to assess and choose the most important conditioning factors for landslide modeling and to compare the importance of topographic, geological, hydrological, land cover, anthropogenic and climatic factors in the genesis of landslides.
(3) The performance of methods and models to predict landslides should be well assessed in a complex area characterized by several types of landslides: rotational and translational landslides, block slide, rockfall and rock avalanche. To achieve this objective, we took care that the different types of landslides will be represented in the inventory data.

In Morocco, the areas threatened by landslide hazards are mountainous regions, particularly the Rif mountains and the Atlas chain[21,22,35]. Furthermore, several studies into landslide assessment using statistical methods have been completed in recent years, the majority being located in the Rif mountains[1,21,22]. For our project, the study area is the Ourika watershed, which is a part of The Marrakech High Atlas (MHA). The MHA is the most prominent topographic feature in North Africa (Toubkal summit is at ca. 4167 m) and it is located in an active, compressional setting[24,38,46]. The MHA generally and the Ourika watershed especially, are the seats of several natural phenomena such as devastating floods and landslides[3,19,20,45].

In the same way, on the night of July 24 to 25, 2020, high-intensity rainfall triggered a rock collapse in a valley in the High Atlas of Marrakech located about a few dozen kilometers from the Ourika watershed in the same geological, geomorphological and climatic context. The damage was severe and 15 people was killed. Hence, it is essential to evaluate assess susceptible areas to landslides and to and reduce the risk to property and people and with a view toward minimizing environmental degradation and achieving more sustainable development.

Material and Methods

Study area: Our research area is located on the northern flank of the Marrakech High Atlas in an active compressional setting[16]. The Ourika watershed lies between latitudes 31°23′ N and 31°03′ N and longitudes 7°35′ W and 7°53′ W (Figure 1) and covers 575 km². Topographically, the altitude is between 800 and 2000 m in the downstream with a relief dominated by hills and plateaus. In upstream, the summits exceed 3500 m and reach up to 4000 m (Figure 1).

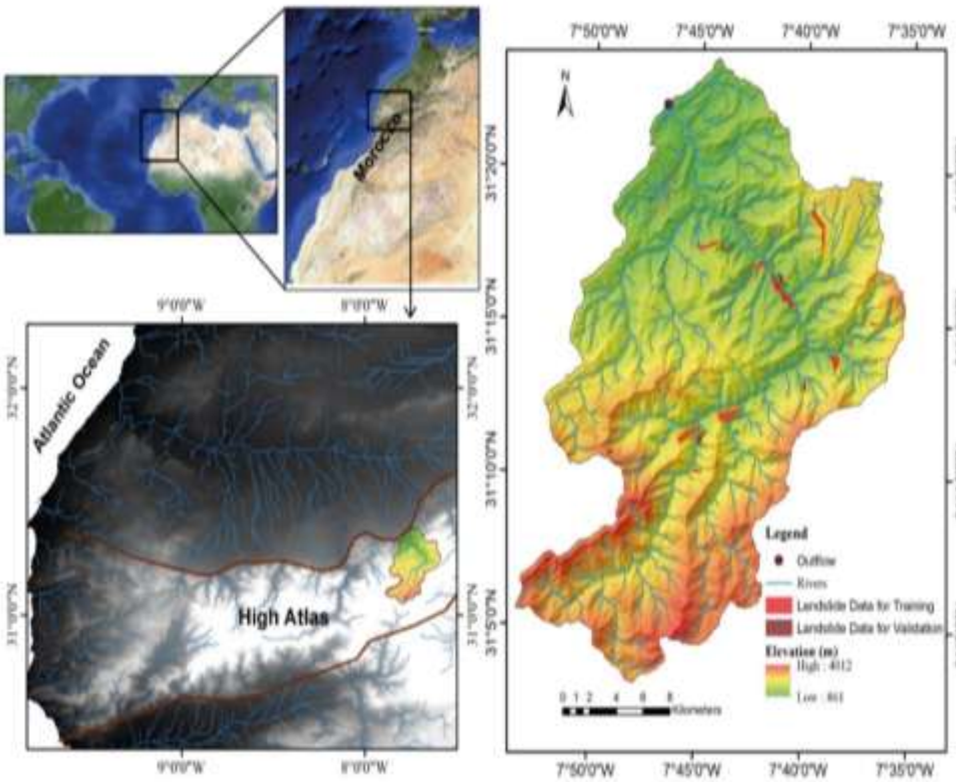


Figure 1: The study area and spatial distribution of landslides inventory

The lithological characteristics of the study area are shown in figure 2. Geologically, the northern part of the area is composed of Triassic sandstones and clays. This emergence rests unconformably on Palaeozoic schist. Further south, the Ourika River is located at an outcrop of older clastic and magmatic terrains comprising Cambrian trachy-andesites and conglomerates, Precambrian granites and granodiorites. In the southern part of the basin, the Precambrian rocks comprise granites and granodiorites (Precambrian I, II and III), mica schists and gneisses (Precambrian I) and volcanic massifs (Trachy-andesites of Precambrian III) (Figure 2). Structurally, the basin is crossed by numerous reverse faults, the most important are: The Fault of Sidi Ali Ou Fars, the Oukaidemen Fault and the Tizi N'Test Faults[17] (Figure 2). The trigger for collapse was established as being seismic activity related to the proximity of the major Tizin'Test fault[16,28].

The climate of the Ourika watershed is complex; it represents an amalgam of a semi-arid and mountain climate. Annual precipitation ranges from 400 to 650 mm on the Haouz Plain and the centre of the drainage basin up to 800 to 1000 mm on the high summits exposed to the humid Atlantic Ocean winds. However, the precipitation sometimes occurs as storms and exceeds 100 mm per day which can initiate landslides and floods[19,45].

Methodology: Our approach comprises a quantitative method and is based on the principle that anticipation of future landslides is based on determining the causative factors of the past[10]. For this, we chose to apply four statistical methods: Frequency Ratio Model (Fr), Weighting Factor Model (Wf), Weight of Evidence Model (WOE) and Logistic Regression Model (LR).

Frequency Ratio Model (FR): The FR method was tested by several researchers and is based on the principle of establishing the relationship between the spatial distribution of landslides and causative factors[8,29]. Each causal factor is subdivided into several classes, so the Fr index value is calculated for each class of factors using eq. 1:

$$Fr = \frac{PLi}{PDi} = \frac{\left(\frac{NLi}{NLt}\right)*100}{\left(\frac{NAi}{NAt}\right)*100} \tag{1}$$

where *PLi* denotes the percentage of landslide mapping for each class i of causative factors relative to the total number of landslides mapped in the study area, *PDi* is the percentage of each class i of causative factors, relative to the total area of the watershed, *NLi* is the number of landslide pixels in a thematic class i, *NLt* is the number of pixels of all landslides, *NAi* is the total number of pixels in a thematic class i and *NAt* is the total number of all pixels.

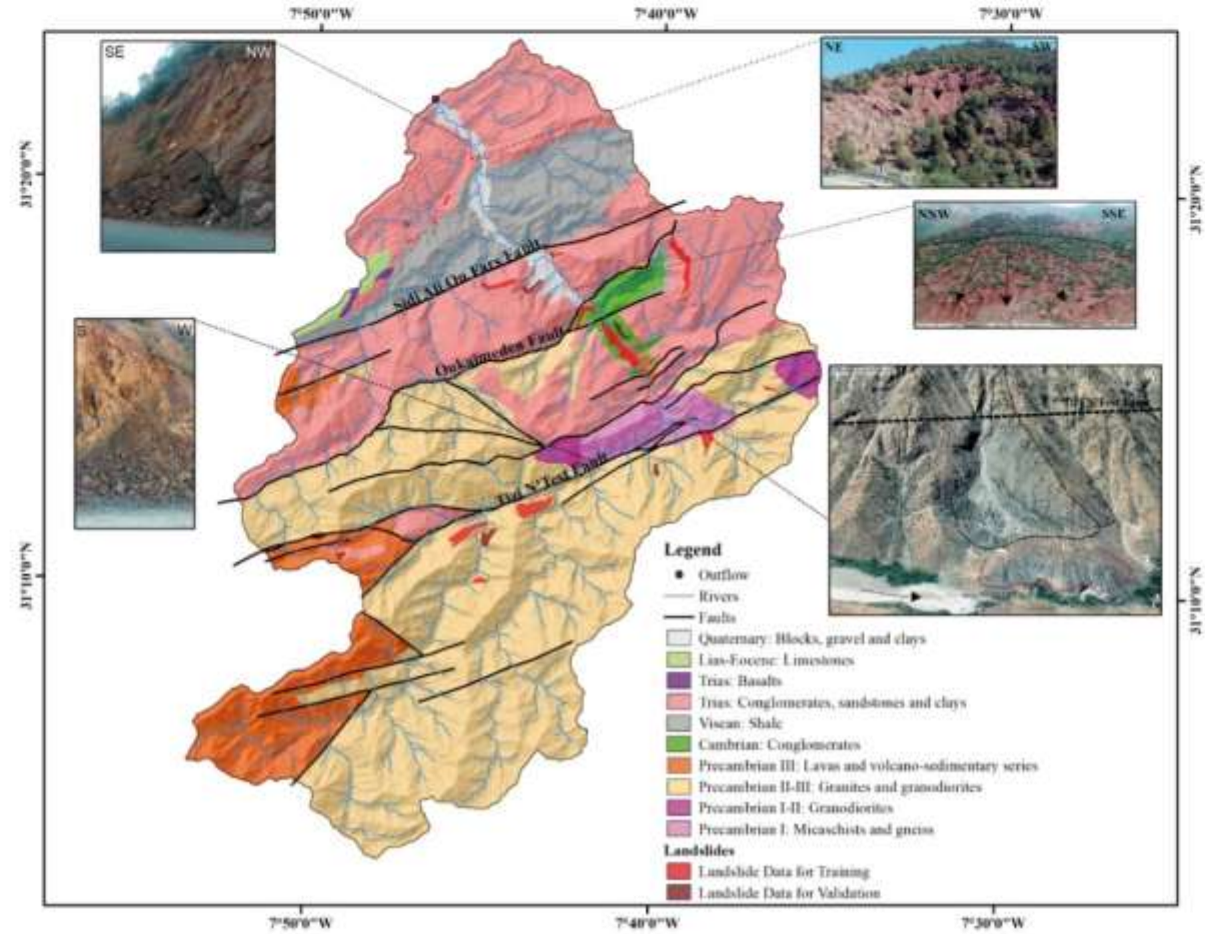


Figure 2: Geological map of the study area

The results obtained represent the correlation between each class of causal factor and landslide occurrence. Thus, values less than 1 indicate a lower probability of the occurrence of a landslide and the values greater than 1 indicate a higher probability of landslide occurrence. Calculated Fr values are used to determine the weight of each class of causative factor to achieve a final map that represents the index of susceptibility of landslides in a GIS environment (LSI). The landslide susceptibility index (LSI) was calculated by a summation of each factor ratio value (eq. 2)[32]:

$$LSI\ (Fr) = \sum Fr \tag{2}$$

where *Fr* is the frequency ratio of each factor type or range.

Weighting Factor Model (Wf): This method is described by several researchers as the Statistical Index Method. First, the statistical index (Si) is calculated for each class of causative factor and is equal to the ratio of the density of landslides in the class relative to that in the whole study area[49]. Si is calculated as per eq. 3:

$$Si = Ln\,\frac{NL_i/NA_i}{NL_t/NA_t} \tag{3}$$

where *Si* represents the weight of each class i of the causative factors, NL_i is the number of landslide pixels in a thematic class i, NA_i is the total number of pixels in a thematic class i, NL_t is the number of pixels of all landslides and NA_t is the total number of all pixels.

Secondly, to evaluate the weight of each parameter in the genesis of the landslides, a weighting factor (Wf) for each causal factor has been calculated using the max–min approach shown in eq. 4 and eq. 5[4,36]. Then, we calculate the final weighting index value of each class i of the causative factors (Wfi) using equation eq. 6:

$$Tsi = Si * NLi \tag{4}$$

$$Wf = \frac{Tsi - Min(Tsi)}{Max(Tsi) - Min(Tsi)} * 9 + 1 \tag{5}$$

$$Wfi = Si * Wf \tag{6}$$

where *Wf* is weighting index value of each causative factor, *Wfi* is final weighting index value of each class i of the causative factors, *Tsi* is the total weighting index value of each class i of the causative factors, *NLi* is the number of landslide pixels in a thematic class i, *Min(Tsi)* is the minimum total weighting index value within the selected layers and *Max(Tsi)* is the maximum total weighting index value within the selected layers.

Finally, to generate the final landslide hazard map, we use the following equation 7:

$$LSI\ (Wf) = \sum Wfi \tag{7}$$

Weight of Evidence Model (WOE): Weight of Evidence Model is a statistical method that uses the log linear from the Bayesian probability model to estimate the relative importance of evidence by statistical means[4,40].

The model is based on the calculation of positive and negative weights W^+ and W^- in relation to the presence or absence of the landslide in each class of the causative factors[8]. The positive and negative weights are calculated by equations 8 and 9 respectively[40]:

$$W^+ = Log_e \frac{P\{F/L\}}{P\{F/\bar{L}\}} = Log_e \frac{N_1/N_1+N_2}{N_3/N_3+N_4} \tag{8}$$

$$W^- = Log_e \frac{P\{\bar{F}/L\}}{P\{\bar{F}/\bar{L}\}} = Log_e \frac{N_2/N_1+N_2}{N_4/N_3+N_4} \tag{9}$$

where P is the probability, F indicates the presence of a causative factor, $\bar{F}$ indicates the absence of a causative factor, L indicates the occurrence of landslides, $\bar{L}$ indicates the absence of landslides, N_1 indicates the number of pixels where a causative factor and a landslide are both present, N_2 is the number of pixels in the study area where a landslide is present but a causative factor is absent, N_3 is the number of pixels where a causative factor is present but a landslide is absent and N_4 is the number of pixels where a landslide and a causative factor are both absent.

From the values of the negative and positive weights, we have to establish the correlation between the causative factors and the occurrence of a landslide (Table 1). We then need to calculate the weight contrast (index C) according to equation 10. Based on index C, a final LSZ map was prepared (eq. 11):

$$C = W^+ - W^- \tag{10}$$

$$LSI\ (WOE) = \sum C \tag{11}$$

Logistic Regression Model (LR): Logistic Regression is a Machine Learning algorithm which is used as a classification solution, it is a predictive analysis algorithm and based on the notion of probability. This method has been widely used for landslide assessment. LR represents the analysis of a multivariate regression relationship between landslide occurrence and the causative variables[49].

In other words, LR aims to determine the relationship between a dependent variable (Landslide) and independent variables (causal factors)[37]. The benefit of the LR method is that variables can be continuous or discrete and they do not necessarily have a normal distribution[4].

In LR modelling, the dependent factor is a binary variable where the presence of a landslide is coded as a 1 and absence is coded by 0. However, independent variables can be continuous or categorical.

Table 1
The Correlation between causative factor and the occurrence of a landslide according to the values of W^+ and W^- in WOE method (from 40)

	Positive weights W^+	Negative weights W^-	Correlation between causative factor and the occurrence of a landslide
Value	> 0	< 0	Correlation positive
	< 0	> 0	Correlation negative
	0	0	Uncorrelated

In this study we chose to convert the parameters from nominal to numerical by coding and ranking the classes based on the relative percentage of the area affected by landsliding[37]. This method is cited by Bourenane et al[4] and is based on calculation of the weight factor for each factor class by summing the ratios of the observed landslide area to the area of each class. Weight factors have been transferred to the quantitative values from 0 to 10. The class with the maximum has been given a weight of 10 and the other classes were given weights <10 based on their proportions[4].

Finally, the probability of the presence, or not, of a landslide is calculated according to equations 12 and 13:

$$P = {}^{1}/_{1 + e^{-z}} \tag{12}$$

$$Z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \tag{13}$$

where P is the probability, Z is the linear combination of independent variables, β_0 is the intercept of the model, $\beta_1, \beta_2 \ldots \beta_n$ are the coefficients of the logistic regression model, $x_1, x_2 \ldots x_n$ are the independent variables and n is the number of independent variables.

Conditioning factors database

Landslide inventory: An inventory of existing landslides was performed to determine the relationship between the distribution of landslides and causative factors. A set of 100 landslides were identified and mapped by evaluating observations from satellite images (Google Earth images), followed by fieldwork carried out from 2010 to 2018. From this landslide inventory, approximately 75% were randomly selected and reserved for landslide susceptibility assessment and 25% were selected for validation (Figure 2).

Based on lithological and geomorphological features, three main types of landslides were identified: (1) Major landslides identified in the Triassic clay formations approaching streams and roads (2) Rockfall along the main Ourika River and its principal tributaries especially at cliffs comprising Triassic sandstone and conglomerate (3) Rock avalanche occurs on the Triassic cliffs and resistant Precambrian rocks, particularly along the escarpment of the Tizi N'Test faults.

Causal factors: According to several researchers and depending on the characteristics of the study area and data availability, 13 landslide-conditioning factors were considered in the current study and are grouped into six classes : topographic factors (elevation, slope angle, slope aspect, stream power index (SPI), sediment transport index (STI), topographic wetness index (TWI) and curvature), geological factors (distance to faults and lithology), hydrological factors (distance to rivers), land cover factors (NDVI), anthropogenic factors (distance to roads) and climatic factors (rainfall)[8]. The maps of slope angle, elevation, slope aspect, curvature, distance to rivers, SPI, STI and TWI are generated from a digital elevation model (DEM) with a resolution of 30 m and reclassified into numerous categories (Table 2 and Figure 4). The lithology and distance to faults maps are produced from the geological map (Table 2 and Figure 4).

To determine the density of the vegetation, we calculate the normalised difference vegetation index (NDVI) from the LANDSAT satellite image. The distance to roads map is produced from the minimum distance to a road represented in a vector format calculated at 100 m intervals (Figure 4). The map of rainfall was constructed by spatial modelling of the average rainfall at 16 climate stations from 1996 to 2016 using the Kriging method and reclassifiying it into 4 categories (Figure 4). The rainfall data were obtained from the Hydraulic Basin Agency of Tensift k;

Landslide conditioning factors analysis

Multicollinearity analysis: Multicollinearity analysis is used in statistics to detect the linearity between the explanatory factors of a given phenomenon and it is usually used in a multiple regression model. Multicollinearity refers to the non-independence of landslide conditioning factors that may occur in datasets[8]. In our study, multicollinearity for the causative factors of landslides was identified using tolerances and VIF methods, according to equations 14 and 15:

$$Tolerance = 1 - R_j^2 \tag{14}$$

$$VIF = \left[\frac{1}{Tolerance}\right] \tag{15}$$

where R_j^2 is the coefficient of determination.

Selection of landslide conditioning factors: The ability to predict a landslide depends on the factors introduced into the model. In fact, some factors can reduce this capacity and so a preliminary selection of the factors is indispensable. For this purpose, we used Information Gain value to select landslide conditioning factors[8].

Table 2
Spatial database of the study area

Factors	Data layers	Data provider
Landslide inventory		Google Earth data Field investigation
Topographic factors	Elevation Aspect Slope Curvature SPI STI TWI	SRTM-DEM from (http://gdex.cr.usgs.gov/gdex/) pixel size of 30 m × 30 m.
Geologic factors	Lithology Distance to Faults	Geological map of Morocco at the scale 1:500000 Geological map of Proust, 1971 Field work
Hydrologic factors	Distance to Stream	DEM at 30m
Land cover factors	NDVI	LANDSAT satellite image at 30m from (https://earthexplorer.usgs.gov/)
Anthropogenic Factors	Distance to Road	Topographic map at the scale 1:50000 Google Earth data Field investigation
Climatic Factors	Rainfall	Climatic stations data from hydraulic basin agency of Tensift

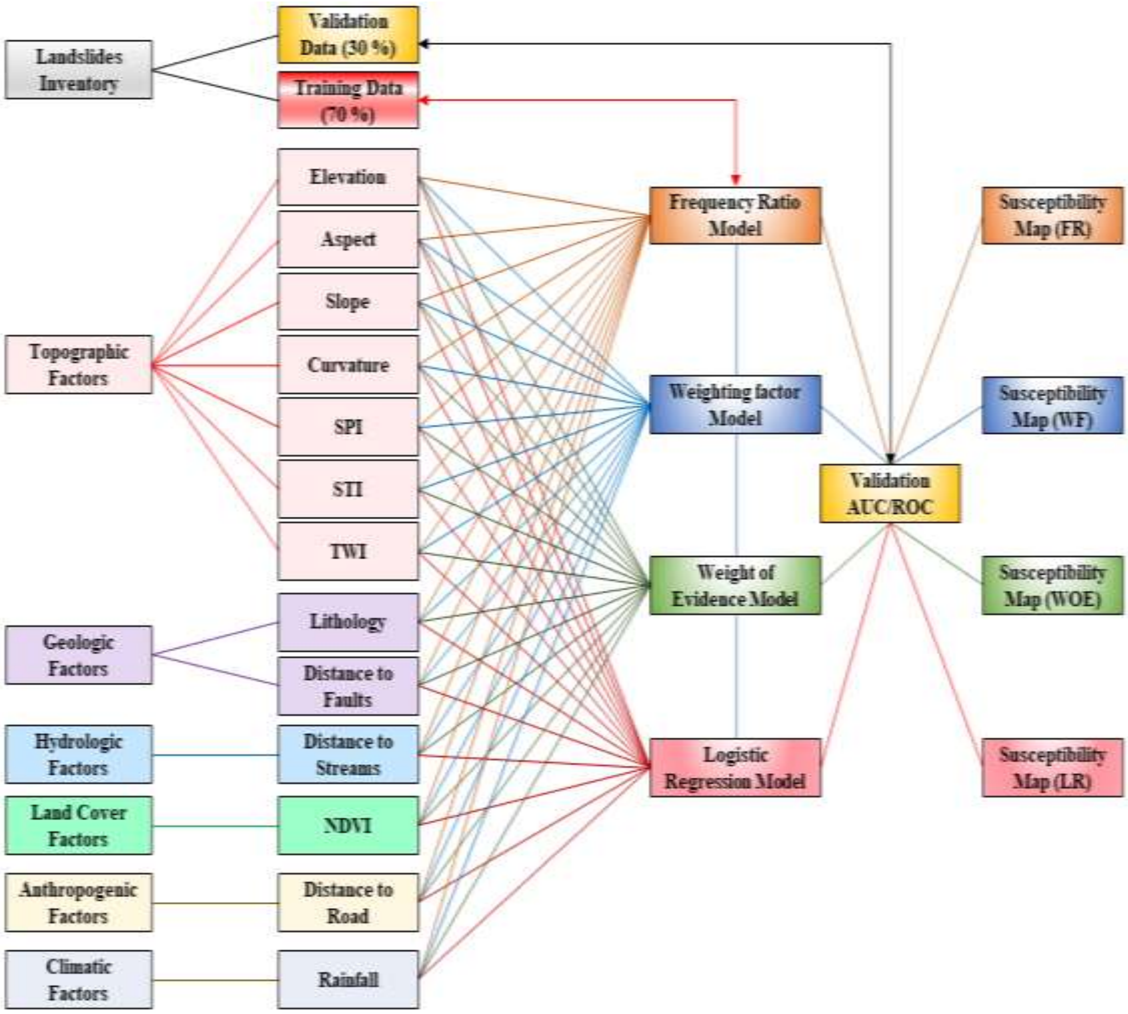
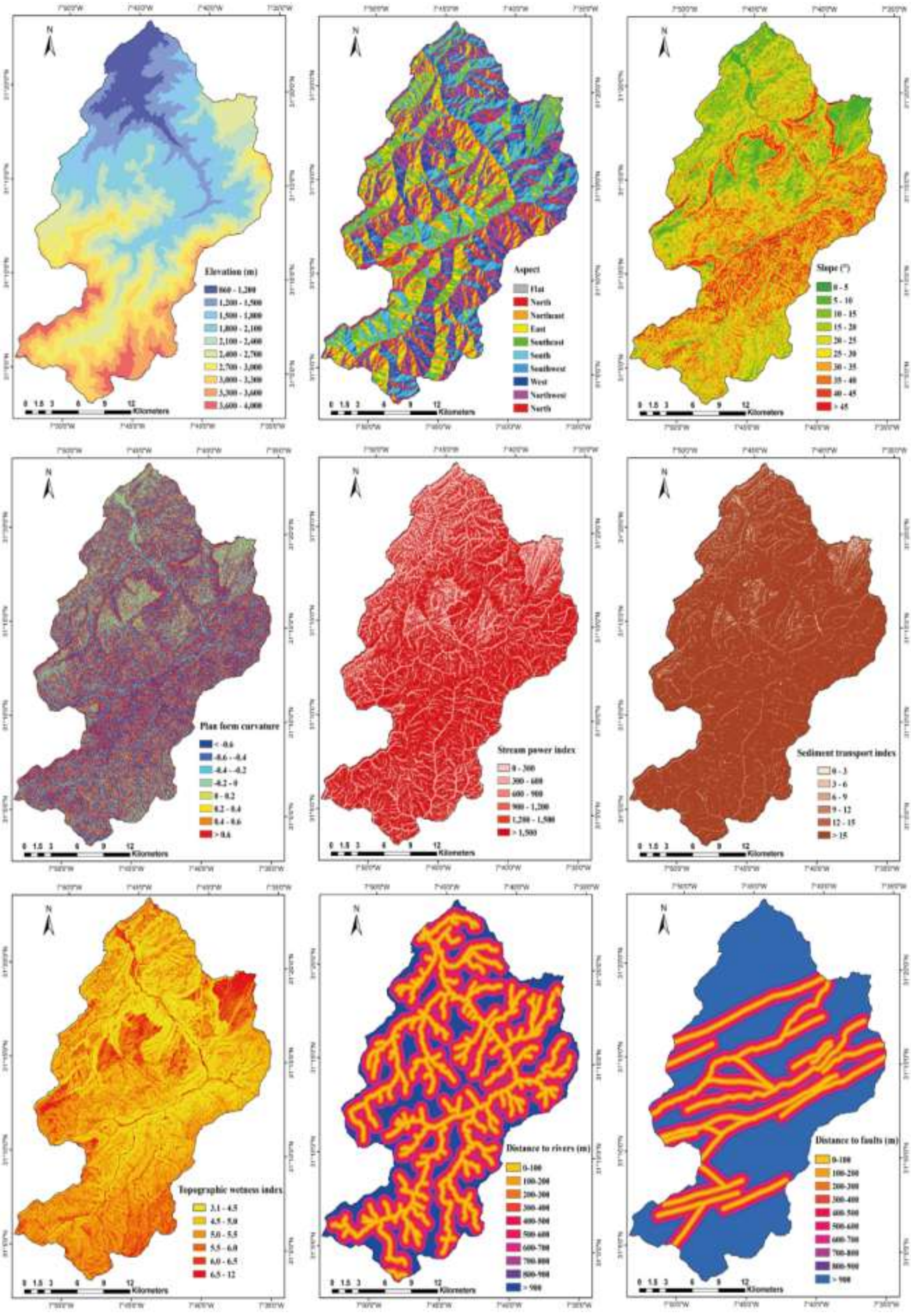


Figure 3: Flowchart of the developed methodology


Elevation (m)
860 - 1,200
1,200 - 1,500
1,500 - 1,800
1,800 - 2,100
2,100 - 2,400
2,400 - 2,700
2,700 - 3,000
3,000 - 3,300
3,300 - 3,600
3,600 - 4,000
Aspect
Flat
North
Northeast
East
Southeast
South
Southwest
West
Northwest
North
Slope (°)
0 - 5
5 - 10
10 - 15
15 - 20
20 - 25
25 - 30
30 - 35
35 - 40
40 - 45
> 45
Plan form curvature
Stream power index
Sediment transport index
Topographic wetness index
Distance to rivers (m)
Distance to faults (m)
Kilometers

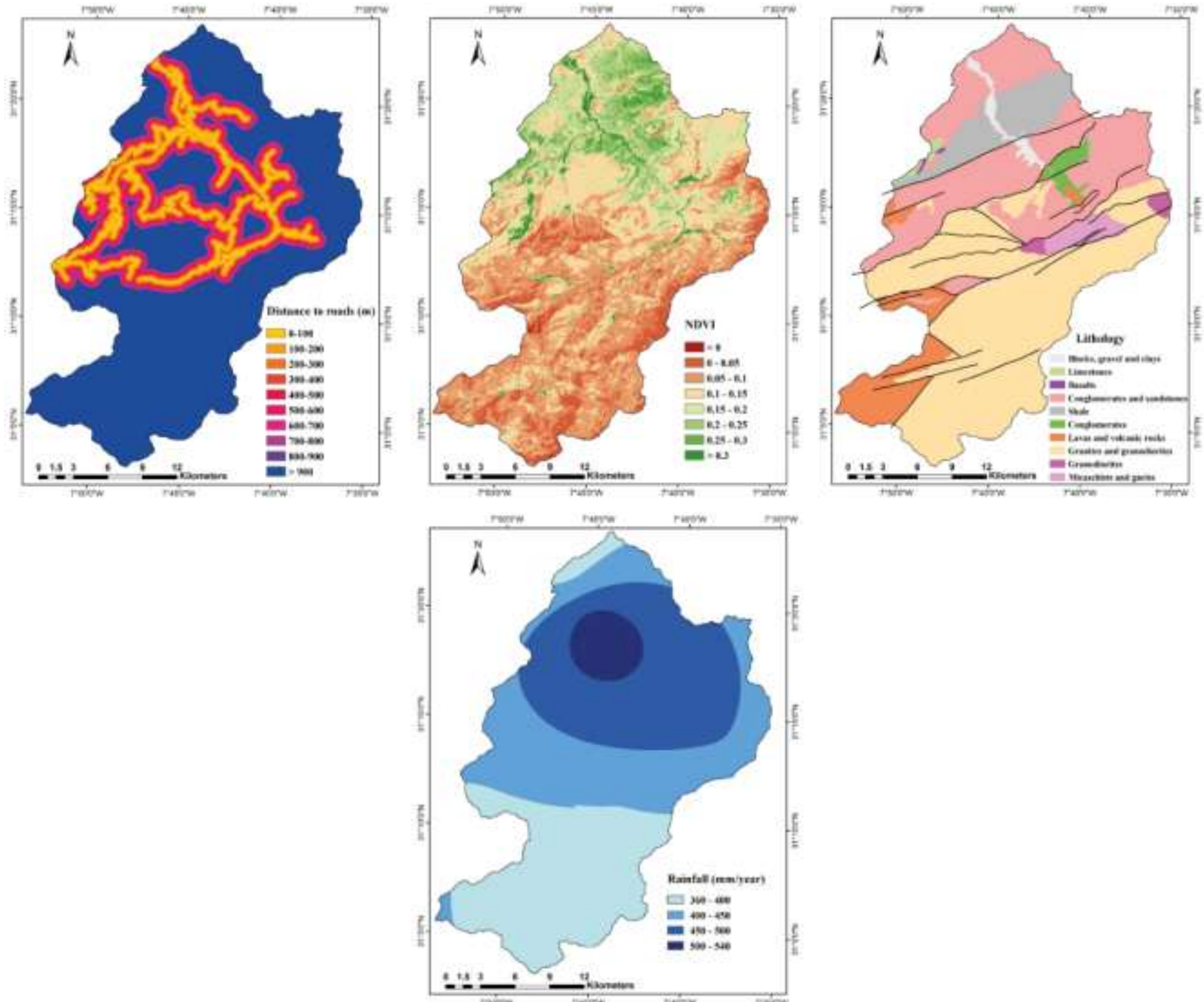


Figure 4: Landslide conditioning factors used in landslide susceptibility analysis

The information gain (IG) value for a landslide conditioning factor Xi and class Y is calculated using equations 16, 17 and 18[8]:

$$IG(Y, X_i) = H(Y) - H(Y_i|L_i) \tag{16}$$

where

$$H(Y) = -\sum_i P(Y_i)\, Log_2\,(P(Y_i)) \tag{17}$$

$$H(Y_i|L_i) = -\sum_i P(Y_i)\, \sum_j P(Y_i|L_i) Log_2\,((P(Y_i|L_i)) \tag{18}$$

where $H(Y)$ is the entropy value of Yi, $H(Y_i|L_i)$ is the entropy of Y after associating values of landslide conditioning factor L_i, $P(Y_i)$ is the prior probability of the out-class Y and $P(Y_i|L_i)$ is the posterior probabilities of Y given the values of conditioning factor L_i.

Calculation of the weights and susceptibility map: Initially, all of the data used in the present study were geo-referenced to the Nord Maroc coordinate system. The database construction and analysis required several steps:

Step (1): Classification of thirteen maps of the factors controlling landslides and their conversion to a raster format with the same spatial resolution (30 m x 30 m). The total number of pixels covering the Ourika watershed is 637256.

Step (2): After compiling the inventory map of historical landslides, it was converted to a raster format, then divided into two parts: approximately 75% for training and 25% for validation. The total size of the historical landslides area is 8104 pixels; 6040 pixels were used for training and 2064 pixels were used for validation.

Step (3): The results obtained were introduced into the Microsoft Excel calculator and the number of pixels in each class of each causal factor and the number of landslide pixels in each class were all calculated. This calculates the percentage of domain (PD) and percentage of landslide (PL) in each class of causal factors.

Step (4): Weight calculation was carried out for Frequency Ratio, Weighting Factors and Weight of Evidence Models using the equations 1, 6 and 10. Then, landslide susceptibility maps were constructed in GIS software using equations 2, 7 and 11 and reclassified into five susceptibility classes using the Natural Breaks (Jenks) method with rankings of very low, low, moderate, high and very high susceptibility.

Step (5): For the Logistic Regression Model, selected causal factors and inventory maps (landslides and non-landslides) were converted to POINT format, then exported as a dBASE table and imported into an SPSS statistical package; the correlations between the dependent factor (landslide) and independent factors were then calculated. The non-landslide data were mapped based on field missions and google earth images in equal proportion to landslide inventory pixels. Finally, the landslide susceptibility map was constructed using equations 12 and 13 in GIS software and reclassified into five susceptibility classes using the Natural Breaks (Jenks) method.

Model validation and comparison: Validation of the results in landslide susceptibility modelling is an essential step for the performance of landslide models[8,11]. In our study, we chose three validation methods: (1) Statistical rules for spatial effective LHM, (2) Statistical measures and (3) Receiver Operating Characteristics Curve (ROC).

Statistical rules for spatial effective LHM: The method is based on the spatial distribution of the inventoried landslides for a training area according to the different landslide susceptibility class (low, very low, moderate, high and very high). The goal is to test two characteristics: (1) if the percentage of landslides increases with a rise in landslide susceptibility and (2) if the highest percentage of observed landslides belong to the high-risk class, provided that the high-risk class only covers small areas[4,43].

Statistical measures: The principle is based on the calculation of two parameters: sensitivity and specificity. Sensitivity is the proportion of landslide pixels that are correctly classified as landslide occurrences and specificity is the proportion of the non-landslide pixels that are correctly classified as non-landslides[8,9]. The calculation of sensitivity and specificity is carried out according to eq. (19) and (20) and then the accuracy of each model is determined by Eq. (21) and as the accuracy gets closer to 1, the model becomes more efficient.

$$Sensitivity = \frac{TP}{TP+FN} \tag{19}$$

$$Specificity = \frac{TN}{FP+TN} \tag{20}$$

$$Accuracy = \frac{TN+TP}{TP+FP+TN+TP} \tag{21}$$

where TP (true positive) and TN (true negative) are the number of pixels that are correctly classified and FP (false positive) and FN (false negative) are the numbers of pixels incorrectly classified.

The Receiver Operating Characteristics curve (ROC): This method is the most useful way to validate the results and estimate the excellence and the performance of landslide hazard mapping[4,8,44]. For this method, a comparison is made between the landslide susceptibility map and the training and validation inventory maps of the historical landslides. The receiver operating characteristics curve is a graphical representation that plots the cumulative percentage of landslides falling into each class on the landslide susceptibility maps in the y-axis and the cumulative percentage of susceptibility classes in the x-axis[18]. To finish, the air under the curve is calculated (AUC) from the ROC curve and the precision of the model is evaluated. The area under the ROC curve varies between 0 and 1 which can be categorised as poor (0.5–0.6), average (0.6–0.7), good (0.7–0.8), very good (0.8–0.9) and excellent (0.9–1.0)[4,8,23].

Results and Discussion

Landslide conditioning factor analysis

Multicollinearity analyses of landslide conditioning factors: The multicollinearity analyses of the 13 conditioning factors of landslides show that tolerance values vary between 0.254 and 0.907. Accordingly, the VIF values vary between 1.102 (for Curvature Factor) and 3.936 (for TWI Factor) (Table 4). These results are tolerable since the tolerance values are greater than 0.1 and those of VIF are less than 10; it reveals that 13 landslide conditioning factors have no multicollinearity.

Table 3
Description of geological units of the study area

Class	Lithology	Formation	Geological age
1	Micaschists and Gneiss	-	Precambrian I
2	Granodiorites	-	Precambrian I-II
3	Granites and Granodiorites	-	Precambrian II-III
4	Lavas and Volcanic rocks	-	Precambrian III
5	Shale	-	Paleozoic (viseen)
6	Conglomerate, Sandstone and Siltstone	F5: Oukaimeden Sandstone F6: Upper Siltstone	Triassic and Cambrian
7	Basalts		Trias
8	Limestones		Lias-Eocene
9	Blocks, Gravels and Clays	-	Quaternary

Table 4
Multicollinearity diagnosis and Average Information Gain for the landslide conditioning factors

Landslide conditioning factors		Collinearity Statistics: Tolerance	Collinearity Statistics: VIF	Information Gain: Average merit
Topographic factors	Elevation	0.561	1.782	0.233
	Aspect	0.677	1.477	0.121
	Slope	0.261	3.833	0.528
	Curvature	0.907	1.102	0.061
	SPI	0.852	1.174	0.095
	STI	0.858	1.166	0.049
	TWI	0.254	3.936	0.498
Land cover factors	NDVI	0.803	1.245	0.094
Hydrological factors	Distance to Rivers	0.653	1.531	0.189
Geological factors	Distance to Faults	0.640	1.563	0.059
	Lithology	0.618	1.619	0.065
Anthropogenic Factors	Distance to Road	0.454	2.202	0.051
Climatic Factors	Rainfall	0.904	1.107	0.130

Selection of landslide conditioning factors: The results of the analysis using Information Gain method are shown in table 4, they show that the Slope and TWI factors have the highest Information Gain (0.528 and 0.498 respectively), followed by the Elevation (0.233), Distance to Rivers (0.189), Rainfall (0.130), Aspect (0.121), SPI (0.095), Lithology (0.065), Curvature (0.061), Distance to Road (0.059), Distance to Faults (0.051) and the STI (0.049). The 13 factors have positive Information Gain and all of them were included in this analysis.

Correlation between landslides and conditioning factors: Figure 6 and table 6 show the spatial relationship between each landslide conditioning factor and FR, WF, WOE and LR index.

Application of the FR, WF and WOE models: Generally, the highest value of FR index is observed in the class of slopes exceeding 45°; the value is 6.016. For WF and WOE index, the highest values are observed in the first class of TWI (10.01 for WF index model and 0.926 for WOE index model). This indicates the importance of slope and TWI factors in the manifestation of landslides.

Likewise, the results show that the FR, WF and WOE index values are higher for altitudes less than 2400 m. Thus, the values are either zero (FR and WF) or negative (WOE) for the other classes. The highest values are present in classes 1200-1500 m and 2100-2400 m.

With regards to aspect, the north class has a higher FR, WF and WOE values (2.514, 3.128 and 0.474 respectively), while flat areas had a null value. For curvature, the first class (<-0.6) has the highest value for FR index (1.207), WF index (0.228) and WOE index (0.126). The values of the three indexes decrease to class (0-0.02) and then increase again.

In the case of slope, SPI and STI factors, the values of FR, WF and WOE index increase in parallel from the first class to the last class. The highest values are FR = 6.016, WF = 10.010 and WOE = 0.926 observed in the >45° Slope class; FR = 1.320, WF = 0.415 and WOE = 0.238 observed in the >1500 SPI class; and FR = 1.059, WF = 0.062 and WOE = 0.085 for the >15 STI class. Conversely, the FR, WF and WOE values decrease for the TWI factors, the most important values being 2.820, 10.077 and 1.024 in the first class (0-3). In terms of NDVI factor, the class of 0.21-0.25 has the highest FR, WF and WOE values (1.641, 0.816 and 0.220 respectively) followed by <0 (bare ground) (FR = 1.641, WF = 0.816 and WOE = 0.220). The other classes have lower FR, WF and WOE values.

In the case of distance to rivers, the three first classes are most prone to landslides because they have the maximum FR, WF and WOE values. The maximum FR value is 1.920 in the 100-200 m class followed by 1.524 in the 0-100 m class and 1.485 in 200-300 m class; FR values progressively decrease away from the rivers. Likewise, WF and WOE values are at a maximum in the 100-200 m class (1.589 and 0.346 respectively), their values are negative 300 m away from the rivers. In general, less distances to rivers correspond to a higher probability of landslide-occurrence.

For the distance to roads, the 600-700 m class has the highest FR value (2.457), WF value (2.348) and WOE value (0.418). The minimum value is observed in the last class (>900 m) where FR = 0.599, WF = -1.485 and WOE = -0.499. For distance to faults, the 600-700 m class has the highest FR value (2.137), WF value (1.575) and WOE value (0.360) followed by the 700-800 m class (FR = 2.099, WF = 1.537 and WOE = 0.349). From a lithological point of view, class 6 (corresponding with Conglomerate, Sandstone and Siltstone from the Triassic and Cambrian) holds the highest values of Fr (1.458), WF (0.624) and WOE (0.273). The other classes have low or null values for FR and negative values for WF and WOE index. In the case of rainfall, the highest FR, WF and WOE values correspond to the class of >500 mm whereas the lowest NFR values (0) correspond to the class of <400 mm.

Application of the LR model: In the LR model, a comparison with the 25% and 75% quotas of overall sample is first realised and the confusion matrices are reported in table 5. The table reveals a considerable stability in the overall performance in both tests (25% and 75%) with values of 91.3% and 90.3% respectively which means there is no change in the regression coefficients. Next, using the logistic regression model, the spatial relationship between landslide occurrence and landslide conditioning factors is evaluated (Table 6 and Figure 6).

Table 5
Confusion matrix with validation sample constituted by the 25% and 75% of the overall sample for LR model (0: absence of phenomena; 1: presence of phenomena)

			Predicted		**Percentage Correct**
			0	1	
25 % of overall sample	Observed	0	1505	28	98.2
		1	39	470	92.3
	Overall Percentage				**96.7**
75 % of overall sample	Observed	0	4440	110	97.6
		1	122	1413	92.1
	Overall Percentage				**96.2**

Table 6
Spatial relation between thematic layers and landslides using FR, IV, WOE and LR methods

Factors	*Class*	*Nb of pixels in class*	*% of domain (PD)*	*Nb of landslide pixels*	*% of landslide (PL)*	*FR index*	*Wfi index (WFI)*	*W+*	*W-*	*WOE (C)*	*Factor Weight index (W)*	*LR coefficients (β)*
												β_0 = -9.012
Elevation	869 - 1200	49722	7.803	749	12.401	1.589	2.908	0.203	-0.022	0.226	5.495	0.219
	1200 - 1500	64213	10.076	1334	22.086	2.192	4.926	0.345	-0.063	0.408	9.787	
	1500 - 1800	74234	11.649	911	15.083	1.295	1.622	0.113	-0.017	0.130	6.684	
	1800 - 2100	87226	13.688	1348	22.318	1.631	3.069	0.214	-0.046	0.260	9.890	
	2100 - 2400	78571	12.330	1363	22.566	1.830	3.794	0.265	-0.054	0.319	10.000	
	2400 - 2700	87835	13.783	335	5.546	0.402	-5.714	-0.398	0.040	-0.438	2.458	
	2700 - 3000	70203	11.016	0	0.000	0.000	0.000	0.000	0.051	-0.051	0.000	
	3000 - 3300	60098	9.431	0	0.000	0.000	0.000	0.000	0.043	-0.043	0.000	
	3300 - 3600	48131	7.553	0	0.000	0.000	0.000	0.000	0.034	-0.034	0.000	
	3600 - 4012	17023	2.671	0	0.000	0.000	0.000	0.000	0.012	-0.012	0.000	
Aspect	Flat	27	0.004	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.048
	North	43457	6.819	733	12.136	1.780	2.635	0.253	-0.026	0.279	5.553	
	Northeast	70302	11.032	1036	17.152	1.555	2.017	0.193	-0.031	0.224	7.848	
	East	69797	10.953	211	3.493	0.319	-5.224	-0.499	0.035	-0.535	1.598	
	Southeast	81790	12.835	130	2.152	0.168	-8.163	-0.779	0.051	-0.830	0.985	
	South	66167	10.383	151	2.500	0.241	-6.509	-0.622	0.037	-0.659	1.144	
	Southwest	62050	9.737	571	9.454	0.971	-0.135	-0.013	0.001	-0.015	4.326	
	West	74281	11.656	803	13.295	1.141	0.601	0.057	-0.008	0.065	6.083	
	Northwest	113981	17.886	1085	17.964	1.004	0.020	0.001	0.000	0.001	8.220	
	North	55404	8.694	1320	21.854	2.514	4.214	0.406	-0.068	0.474	10.000	
Slope	0 - 5	16831	2.641	58	0.960	0.364	-6.442	-0.442	0.008	-0.450	0.336	0.942
	5 - 10	42140	6.613	189	3.129	0.473	-4.764	-0.327	0.016	-0.343	1.094	
	10 - 15	60684	9.523	260	4.305	0.452	-5.055	-0.347	0.025	-0.372	1.506	
	15 - 20	76501	12.005	260	4.305	0.359	-6.530	-0.448	0.037	-0.485	1.506	
	20 - 25	91215	14.314	281	4.652	0.325	-7.155	-0.491	0.047	-0.538	1.627	
	25 - 30	100910	15.835	472	7.815	0.493	-4.496	-0.309	0.040	-0.349	2.733	
	30 - 35	97785	15.345	762	12.616	0.822	-1.247	-0.086	0.014	-0.100	4.412	
	35 - 40	76163	11.952	989	16.374	1.370	2.004	0.138	-0.022	0.160	5.727	

	40 - 45	44738	7.020	1042	17.252	2.457	5.724	0.396	-0.051	0.447	6.034	
	> 45	30289	4.753	1727	28.593	6.016	11.424	0.799	-0.126	0.926	10.000	
curvature	< -0.6	207934	32.630	2378	39.371	1.207	0.522	0.081	-0.045	0.126	10.000	0.041
	-0.6 - -0.4	38921	6.108	274	4.536	0.743	-0.826	-0.130	0.007	-0.138	1.152	
	-0.4 - -0.2	42124	6.610	288	4.768	0.721	-0.907	-0.143	0.009	-0.152	1.211	
	-0.2 - 0	43119	6.766	298	4.934	0.729	-0.877	-0.139	0.009	-0.147	1.253	
	0 - 0.2	21673	3.401	126	2.086	0.613	-1.358	-0.214	0.006	-0.220	0.530	
	0.2 - 0.4	40798	6.402	292	4.834	0.755	-0.780	-0.123	0.007	-0.130	1.228	
	0.4 - 0.6	37710	5.918	291	4.818	0.814	-0.571	-0.090	0.005	-0.095	1.224	
	> 0.6	204977	32.166	2093	34.652	1.077	0.207	0.031	-0.016	0.047	8.802	
SPI	0 - 300	198751	31.189	1527	25.281	0.811	-0.633	-0.093	0.037	-0.130	4.468	0.037
	300 - 600	57817	9.073	212	3.510	0.387	-2.861	-0.415	0.026	-0.441	0.620	
	600 - 900	48433	7.600	346	5.728	0.754	-0.852	-0.124	0.009	-0.133	1.012	
	900 - 1200	32195	5.052	314	5.199	1.029	0.086	0.012	-0.001	0.013	0.919	
	1200 - 1500	26866	4.216	223	3.692	0.876	-0.400	-0.058	0.002	-0.061	0.652	
	> 1500	273194	42.870	3418	56.589	1.320	0.836	0.120	-0.118	0.238	10.000	
STI	0 - 3	176749	27.736	1444	23.907	0.862	-0.395	-0.066	0.023	-0.089	3.152	-0.019
	3 - 6	737	0.116	0	0.000	0.000	0.000	0.000	0.001	-0.001	0.000	
	6 - 9	910	0.143	2	0.033	0.232	-3.888	-0.638	0.000	-0.638	0.004	
	9 - 12	1236	0.194	4	0.066	0.341	-2.859	-0.469	0.001	-0.470	0.009	
	12 - 15	1416	0.222	9	0.149	0.671	-1.063	-0.175	0.000	-0.175	0.020	
	> 15	456208	71.589	4581	75.844	1.059	0.154	0.022	-0.063	0.085	10.000	
TWI	3.1 - 4.5	112505	24.406	4157	68.825	2.820	10.706	0.600	-0.424	1.024	10.000	-0.005
	4.5 - 5.0	200613	43.519	1140	18.874	0.434	-4.026	-0.225	0.074	-0.299	2.742	
	5.0 - 5.5	178595	38.743	412	6.821	0.176	-11.120	-0.617	0.113	-0.731	0.991	
	5.5 - 6.0	86176	18.694	172	2.848	0.152	-12.259	-0.680	0.051	-0.731	0.414	
	6.0 - 6.5	33512	7.270	73	1.209	0.166	-11.571	-0.642	0.018	-0.660	0.176	
	6.5 - 12.0	25855	5.609	86	1.424	0.254	-8.240	-0.458	0.012	-0.469	0.207	
NDVI	< 0	4951	0.777	77	1.275	1.641	1.553	0.218	-0.002	0.220	0.454	0.122
	0.01 - 0.05	65535	10.284	327	5.414	0.526	-2.013	-0.281	0.023	-0.304	1.927	
	0.06 - 0.1	215682	33.845	1572	26.026	0.769	-0.824	-0.116	0.050	-0.166	9.263	
	0.11 - 0.15	172561	27.079	1697	28.096	1.038	0.116	0.015	-0.006	0.021	10.000	
	0.16 - 0.2	78551	12.326	1093	18.096	1.468	1.204	0.168	-0.030	0.198	6.441	
	0.21 - 0.25	58248	9.140	1006	16.656	1.822	1.882	0.263	-0.038	0.301	5.928	
	0.26 - 0.3	27051	4.245	202	3.344	0.788	-0.748	-0.105	0.004	-0.109	1.190	
	> 0.3	14677	2.303	66	1.093	0.474	-2.339	-0.326	0.005	-0.331	0.389	
Distance to rivers	0 - 100	84256	13.222	1217	20.149	1.524	1.594	0.184	-0.036	0.221	8.631	0.464
	100 - 200	77478	12.158	1410	23.344	1.920	2.469	0.286	-0.060	0.346	10.000	
	200 - 300	73746	11.572	1038	17.185	1.485	1.497	0.173	-0.029	0.202	7.362	
	300 - 400	68072	10.682	633	10.480	0.981	-0.072	-0.009	0.001	-0.010	4.489	
	400 - 500	62947	9.878	452	7.483	0.758	-1.051	-0.122	0.012	-0.133	3.206	
	500 - 600	57850	9.078	421	6.970	0.768	-1.000	-0.116	0.010	-0.126	2.986	
	600 - 700	51229	8.039	273	4.520	0.562	-2.179	-0.252	0.016	-0.269	1.936	
	700 - 800	44679	7.011	216	3.576	0.510	-2.548	-0.295	0.016	-0.310	1.532	
	800 - 900	36894	5.790	171	2.831	0.489	-2.707	-0.313	0.014	-0.326	1.213	
	> 900	80105	12.570	209	3.460	0.275	-4.882	-0.563	0.044	-0.607	1.482	
Distance to Faults	0 - 100	42883	6.729	288	4.768	0.709	-1.201	-0.151	0.009	-0.160	1.325	0.029
	100 - 200	41683	6.541	272	4.503	0.688	-1.301	-0.164	0.009	-0.173	1.252	
	200 - 300	39738	6.236	285	4.719	0.757	-0.972	-0.122	0.007	-0.129	1.312	
	300 - 400	37808	5.933	324	5.364	0.904	-0.351	-0.044	0.003	-0.047	1.491	
	400 - 500	35342	5.546	511	8.460	1.525	1.472	0.185	-0.014	0.199	2.352	
	500 - 600	32411	5.086	594	9.834	1.934	2.299	0.290	-0.022	0.312	2.734	
	600 - 700	30556	4.795	619	10.248	2.137	2.648	0.334	-0.026	0.360	2.849	
	700 - 800	27648	4.339	550	9.106	2.099	2.584	0.326	-0.022	0.349	2.531	
	800 - 900	24551	3.853	424	7.020	1.822	2.092	0.264	-0.015	0.278	1.951	
	> 900	324636	50.943	2173	35.977	0.706	-1.213	-0.154	0.118	-0.272	10.000	
Distance to Roads	0 - 100	44049	6.912	516	8.543	1.236	0.832	0.093	-0.008	0.100	2.383	-0.076
	100 - 200	33962	5.329	554	9.172	1.721	2.133	0.238	-0.018	0.257	2.559	
	200 - 300	29273	4.594	515	8.526	1.856	2.430	0.272	-0.018	0.290	2.379	
	300 - 400	25696	4.032	418	6.921	1.716	2.122	0.237	-0.013	0.251	1.931	
	400 - 500	23064	3.619	313	5.182	1.432	1.410	0.157	-0.007	0.165	1.446	
	500 - 600	21551	3.382	415	6.871	2.032	2.785	0.312	-0.016	0.328	1.917	
	600 - 700	20184	3.167	470	7.781	2.457	3.531	0.396	-0.021	0.418	2.171	
	700 - 800	18726	2.939	364	6.026	2.051	2.822	0.316	-0.014	0.330	1.681	
	800 - 900	17358	2.724	310	5.132	1.884	2.489	0.279	-0.011	0.290	1.432	

	> 900	403393	63.302	2165	35.844	0.566	-2.234	-0.250	0.248	-0.499	10.000	
Lithology	1	13887	2.179	128	2.119	0.972	-0.088	-0.012	0.000	-0.013	0.443	0.078
	2	6060	0.951	0	0.000	0.000	0.000	0.000	0.004	-0.004	0.000	
	3	291515	45.745	2280	37.748	0.825	-0.604	-0.086	0.062	-0.147	7.897	
	4	51169	8.030	426	7.053	0.878	-0.408	-0.057	0.005	-0.062	1.476	
	5	51072	8.014	224	3.709	0.463	-2.422	-0.337	0.020	-0.357	0.776	
	6	208951	32.789	2887	47.798	1.458	1.184	0.164	-0.110	0.273	10.000	
	7	738	0.116	0	0.000	0.000	0.000	0.000	0.001	-0.001	0.000	
	8	2706	0.425	0	0.000	0.000	0.000	0.000	0.002	-0.002	0.000	
	9	11158	1.751	95	1.573	0.898	-0.337	-0.047	0.001	-0.048	0.329	
Rainfall	< 400	200517	31.466	179	2.964	0.094	-11.515	-1.030	0.153	-1.183	0.589	0.074
	400 - 450	200705	31.495	2164	35.828	1.138	0.628	0.055	-0.028	0.083	7.125	
	450 - 500	202753	31.817	3037	50.281	1.580	2.231	0.199	-0.137	0.336	10.000	
	> 500	33281	5.223	660	10.927	2.092	3.598	0.325	-0.027	0.352	2.173	

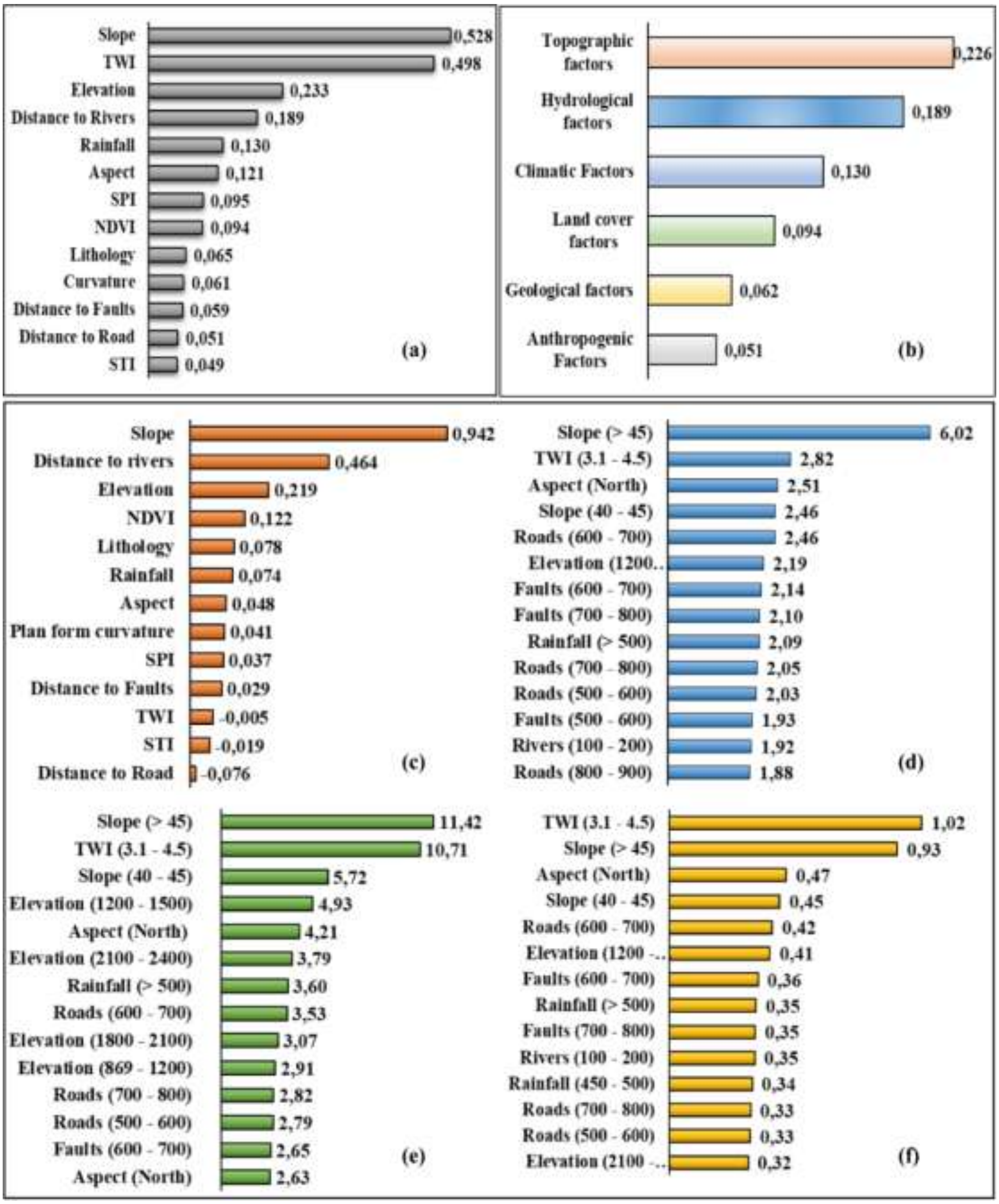


Figure 5: The relative importance of the most landslide conditioning factors using IG method (a) and (b), LR model (c), FR model (d), WF model (e) and WOE (f), for FR, WF and WOE models just the first 14 classes that control landslides are presented

The relative importance of the conditioning factors can be assessed using the corresponding coefficients in the LR model. In this study, all coefficients are positive indicating that they are positively related to the probability of landslide manifestation except STI, TWI and distance to roads. From these results, it can be implied that the slope, distance to rivers and Elevation had the highest coefficients (0.942, 0.464 and 0.219 respectively) which indicate the importance of these factors in the genesis of the landslides.

A logistic regression equation was obtained as shown in eq. 22:

$$Z = -9.012 + 0.219 * Elevation + 0.048 * Aspect + 0.942 * Slope + 0.041 * Curvature + 0.037 * SPI - 0.019 * STI - 0.005 * TWI + 0.122 * NDVI + 0.464 * Distance\ to\ rivers + 0.029 * Distance\ to\ faults - 0.076 * Distance\ to\ roads + 0.078 * Lithology + 0.074 * Rainfall \quad (22)$$

Landslide susceptibility mapping using FR, WF, WOE and LR models: The relationship between landslides and influencing factors was analysed; the degree of spatial correlation between landslides and each factor using the FR, WF, WOE end LR models is shown in table 6. From the equations mentioned above, susceptibility maps of landslides were produced. The four final maps created using the FR, WF, WOE and LR models are grouped in figure 7 to conduct comparative visualisation. The landslide susceptibility maps were reclassified into five classes using the Natural Breaks (Jenks) method in a GIS environment: very low, low, moderate, high and very high (Figure 7 and 8).

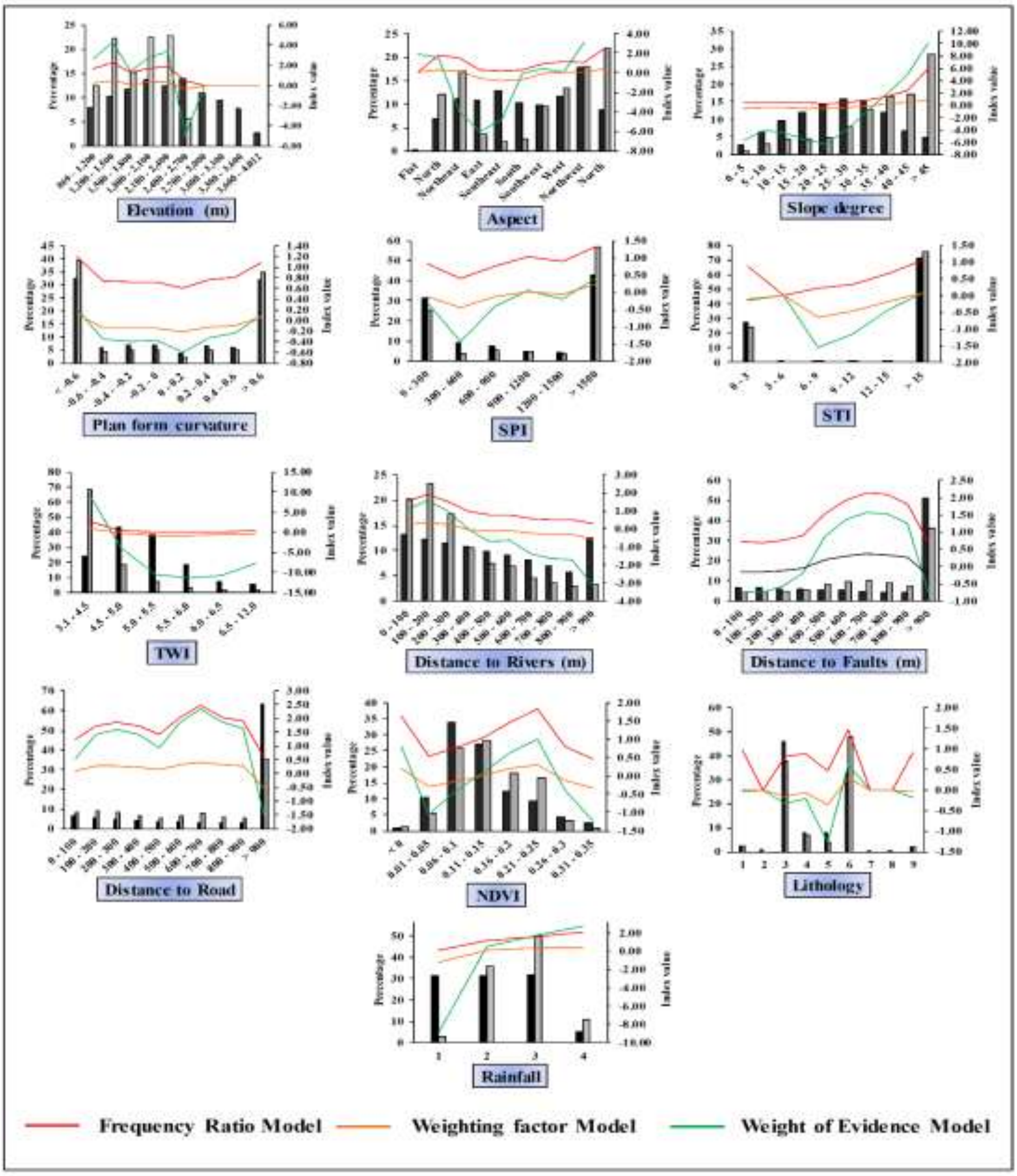


Figure 6: The relationship between landslide and influencing factors class and their FR, WO, WOF index values

In the case of the FR model, it can be observed (Figure 8) that the very low susceptibility class accounts for 23.89% of the study area. The low, moderate and high susceptibility classes account for 27.52%, 28.23% and 14.81% of the study area respectively. The very high susceptibility class accounts for 5.55% of the study area. According to the maps of susceptibility to landslides derived from the WFI model, areas with a very low index represent 19.88% of the total area. Areas with a low, moderate and high index account for 22.56%, 28.34% and 14.60% of the area respectively. The area most threatened by landslides covers 14.62% of the total basin area.

Regarding the landslide susceptibility map generated by the WOE model, 18.28% of the study area belongs to the very low susceptibility class. The low susceptibility class accounts for 22.55% of the study area and the moderate susceptibility class accounts for 28.39% of the study area. The high and very high susceptibility classes account for 19.54% and 11.24% of the study area respectively. Regarding the LR model, figure 7 shows that the very low classification has the maximum area percentage (34.74%) followed by the very high (22.06%), low (18.11%), high (12.69%) and moderate classification (12.39%).

Model performance and evaluation

Statistical rules for spatial effective LHM: In order to verify the results obtained, the four susceptibility maps and the landslide inventory map were compared (Figure 8). For the four models, we see that the majority of the active landslides for validation (25%) falls into the very high susceptibility class: 60.08%, 86.29%, 77.52% and 82.61% for the FR, WFI, WOE and LR susceptibility map respectively. The very low susceptibility class either has very weak or no landslide occurrence in all maps: 0% for the FR, WF and WOE models and 1.21% for the LR model.

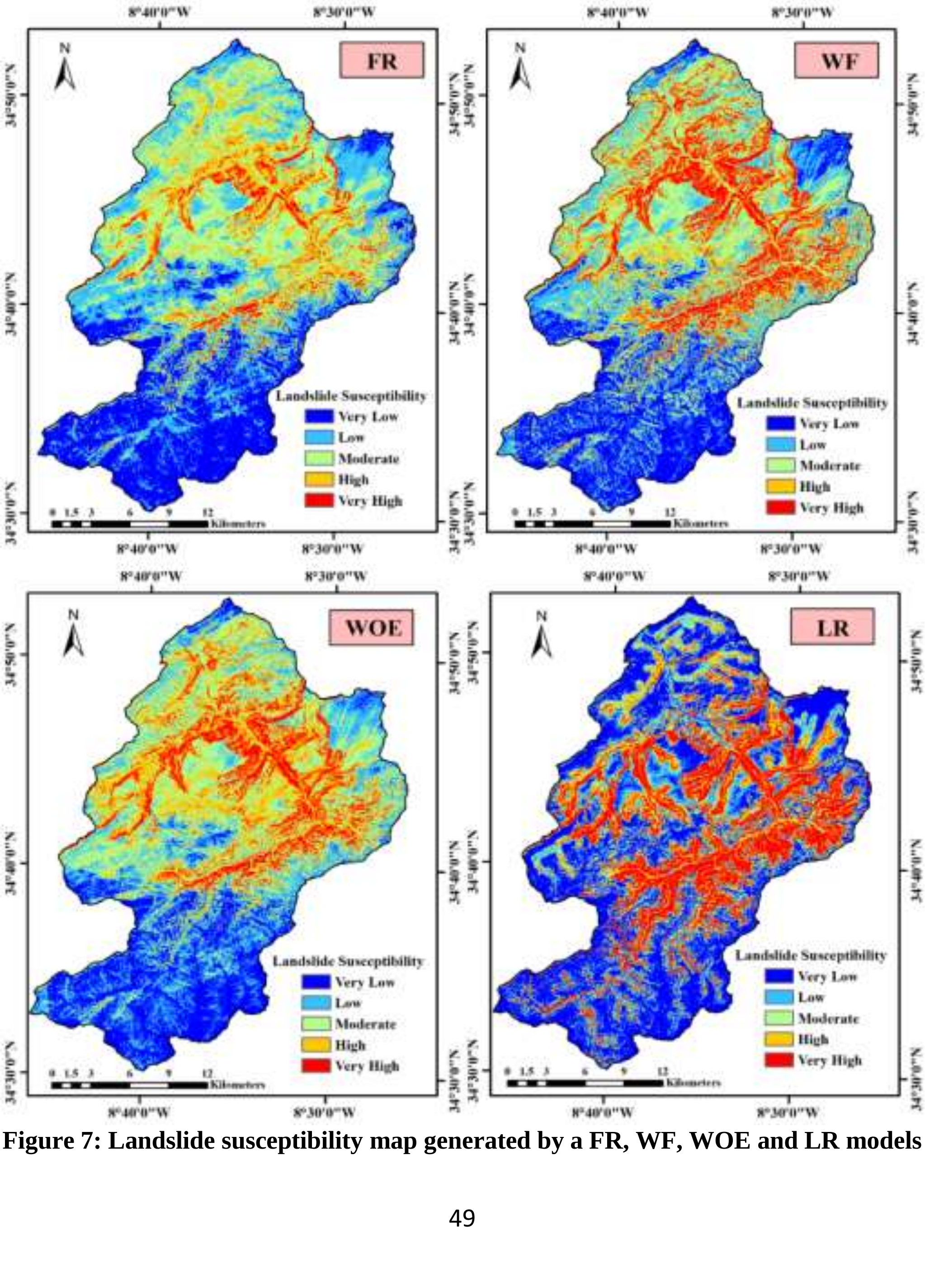


Figure 7: Landslide susceptibility map generated by a FR, WF, WOE and LR models

It is clear from these results that the field-recorded landslide zones have a better fit with the WOE and LR maps than the WF and FR maps. This indicates that the landslide susceptibility prediction is better by LR and WOE than the WF and Fr methods.

Statistical measures: The performance of the landslide models using statistical measures is shown in table 7. It shows that the highest classification accuracy is for the LR model, whose value is 94.40%, although the lowest value is for the WOE model with 89.13%. The classification accuracy is almost equal for the FR model with 90.37% and with WF model as 93.18%.

Regarding the sensitivity for all models, table 7 shows that the LR model has the highest sensitivity (97.71%) indicating a high probability of correctly classifying the landslide pixels to the landslide class. This is followed by the WF model (96.04%) and then the FR model (92.81%) and the lowest sensitivity (89.79%) was found in the WOE model. Concerning specificity, the highest value is for the WOE model (85.42%) indicating that 85.42% of non-landslide pixels are correctly classified to the non-landslide class followed by the LR and WF models (83.68% and 83.13% respectively). The lowest specificity is for the FR model (80.31%).

The Receiver Operating Characteristics Curve (ROC): The estimation of prediction capability for the four landslide susceptibility models is obtained by comparing the landslide training and validation inventory with the susceptibility maps. Then, the rate curves were created (ROC) and the areas under each curve (AUC) were calculated (Figure 9).

The prediction-rate curve, obtained by comparing the landslide training data with the susceptibility map (Figure 9a) showed that the AUC values were 0.75 for the FR and WF models, 0.80 for the WOE model and 0.82 for the LR model. Moreover, in the prediction-rate curve obtained by comparing the landslide validation data with the susceptibility maps (Figure 9b), it was observed that the FR, WF, WOE and LR landslide models present good performance for landslide susceptibility assessment (AUC > 0.8). The LR model achieved the best performance (AUC = 0.88) followed by the WOE model (AUC = 0.82), WF model (AUC=0.81) and FR model (AUC=0.80).

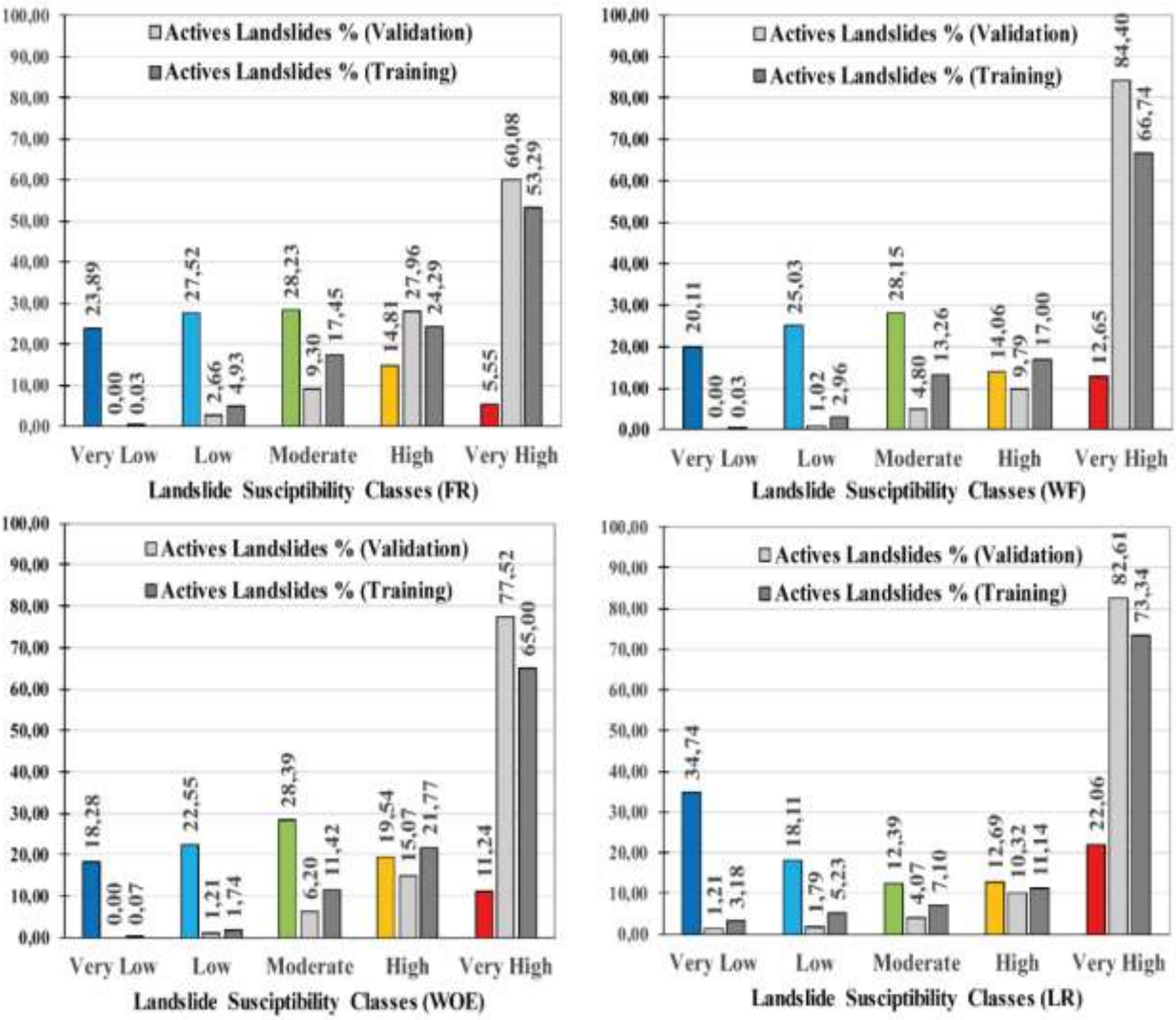


Figure 8: Percentages of different landslide susceptibility classes

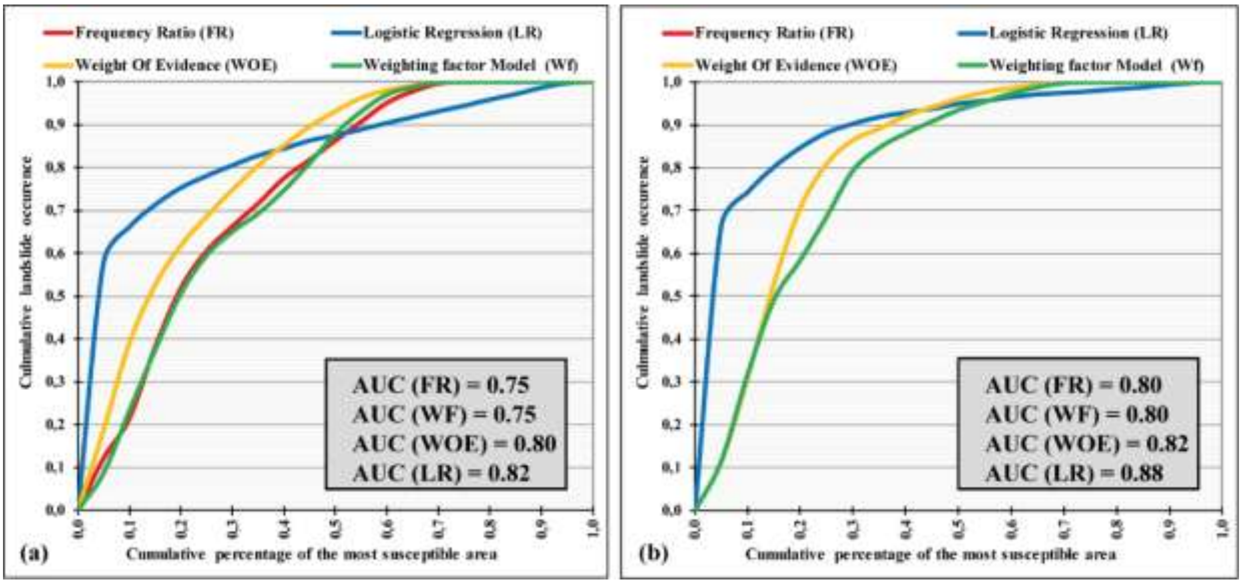


Figure 9: ROC curves for the four landslide susceptibility maps produced by FR, WF, WOE and LR models: (a) landslides training, (b) landslides validation

Table 7
Model performance using validation dataset

Parameters	Models			
	FR	WF	WOE	LR
True positive	1963	1966	2006	1963
True negative	412	483	340	518
False positive	101	98	58	101
False negative	152	81	228	46
Sensitivity	0.928	0.960	0.897	0.977
Specificity	0.803	0.831	0.854	0.836
Accuracy	0.903	0.931	0.891	0.944

The results of the ROC evaluation show that all of the models chosen for the spatial prediction of landslide susceptibility analysis in the Ourika Basin presented judiciously high prediction accuracy. Furthermore, the LR model showed the best result for landslide susceptibility mapping in the Ourika watershed.

Conditioning factors contribution analysis: Analysis of relative importance of conditioning factors contribution to landslides by IG method (Figure 5a and b) shows that topographic factors had the most impact on landslides occurrence in the study area (IG=0.226) with slope and TWI as main factors (IG=0528 and IG=0.498 respectively) followed by hydrological factors (IG=0.189) and climatic factors (0.130).

The importance of topographic and hydrological factors is demonstrated by most previous research in similar areas around the world[9,27]. On the contrary, land cover, geological and anthropogenic factors achieved the lowest values (IG=0.094, IG=0.062 and IG=0.051 respectively).

Otherwise, the assessments of the spatial relationships between the conditioning factors and the landslides susceptibility maps developed by FR, WF and WOE models show that landslides formed principally in areas with a slope greater than 45° and TWI between 3.1and 4.5 (Fig. 5d, e and f). These classes are followed by Aspect (North) and Slope (40-45°) for FR and WOF models and by the slope (40-45°) and elevation (1200-1500m) for WF model. The results based on the LR model showed that slope, distance to rivers, elevation and NDVI are the most important factors which are closely correlated to the occurrence and spatial distribution of landslides compared with other factors (Figure 5c).

The results of our study are largely correlated with previous studies in similar areas. Indeed, the importance of topographic factors in landslides occurrence is shown by several previous studies[2]. In the study area, the slope holds the key role followed by TWI, given that escarpments and soil wetness destabilize slopes.

Ever, the role of recent tectonic and deepening of rivers is important, since these are the factors responsible for the permanent increase of the slopes. Instead, other factors have a minimal role in the manifestation landslides, especially: distance to faults, distance to roads and STI factors.

Generally, the role of faults is difficult to be awarded by accuracy, since the mapping of faults is very old and only a small portion of the structures is mapped (geological map of the study area was carried out in 1971). Finally, the weak participation of anthropogenic factors in the evidence of the landslides will find its explanation in the low population density of the study area in addition to the weakness of the road infrastructure due to the escarpment of reliefs and strong slopes.

Model performance and comparison: This study evaluates the performance of FR, WF, WOE and LR Models for landslide susceptibility mapping in the Ourika area of Morocco. The inventory of previous landslides is a critical step in the assessment of vulnerability to landslides in parallel with an inventory of non-landslides areas, especially for the LR model. In some previous studies, non-landslide data were generated randomly in the study area[8,42] and this method may raise some uncertainties[27]. Hence, in this study, we have mapped areas of non-landslides inventory based on field missions and google earth images to overcome this problem.

On the other hand, to find a more precise model, the Information Gain Method was used for optimisation of the landslide conditioning factors[7,8] and Multicollinearity analysis was used to detect the linearity between the conditioning factors. A total of 13 conditioning factors were selected based on the characteristics of the study area, the literature and their Information Gain value. For validation, we chose to compare the resulting susceptibility maps with the inventory maps (training and validation in the ratio 75% to 25% respectively)[27].

Generally, performance evaluation revealed that the four models have a higher accuracy revealing a good result. Moreover, according to this research, it seems that the LR model (accuracy = 94.40% and AUC=0.88) is the most appropriate algorithm for landslide susceptibility in the Ourika area and similar areas.

Conclusion

Different researchers have proposed several methods for landslide susceptibility mapping. For our case study, we applied the Frequency Ratio, Weighting Factor, Weight of Evidence and Logistic Regression models in order to understand the processes that contribute to the phenomena. The study was conducted in three stages: in the first stage, we completed an inventory of landslides and non-landslide samples; the second step consisted of determining the causative factors and their relationship with landslides.

The last step is the realisation of the landslide's susceptibility maps and application of the validation methods. The 13 causative factors selected in this study were: elevation, aspect, slope angle, curvature, SPI, STI, TWI, distance from rivers, distance from faults, distance from roads, NDVI, lithology and rainfall. A multicollinearity analysis and Information Gain method were executed to choose the appropriate landslide conditioning factors. The 13 factors have positive Information Gain and all of them were included in this analysis.

Finally, the trained FR, WF, WOE and LR models were applied to generate landslide susceptibility maps. The results indicate that these four models may well be useful methods for assessment of landslide phenomena and the realisation of landslide susceptibility maps in similar areas.

In addition, the LR algorithm showed the most stable and reasonable results in this work. In similar area, such a study is of great importance for the protection of property and people that are vulnerable to extreme phenomena. Similar studies would be important using new computer technology to achieve a preventive plan for natural hazards.

References

1. Ait Brahim L., Bousta M., Jemmah I.A., El Hamdouni I., ElMahsani A., Abdelouafi A., Sossey Alaoui F. and Lallout I., Landslide susceptibility mapping using AHP method and GIS in the peninsula of Tangier (Rif-northern morocco). *MATEC Web of Conferences* **149**, https://doi.org/10.1051/matecconf/201814902084 **(2018)**

2. Alkhasawneh M.S. et al, Determination of importance for comprehensive topographic factors on landslide hazard mapping using artificial neural network, *Environ Earth Sci*, **72**, 787–799 https://doi.org/10.1007/s12665-013-3003-x **(2014)**

3. Bennani O., Druon E., Leone F., Tramblay Y. and El Mehdi Saidi M., A spatial and integrated flood risk diagnosis. Relevance for disaster prevention at Ourika valley (High Atlas-Morocco), *Disaster Prevention and Management: An International Journal*, https://doi.org/10.1108/DPM-12-2018-0379 **(2019)**

4. Bourenane H., Guettouche M., Bouhadad Y. and Braham M., Landslide hazard mapping in the Constantine city, Northeast Algeria using frequency ratio, information value, logistic regression, weights of evidence and analytical hierarchy process methods, *Arabian Journal of Geosciences*, **9(2)**, 154 **(2016)**

5. Broeckx J., Vanmaercke M., Duchateau R. and Poesen Jean, A data-based landslide susceptibility map of Africa, *Earth-Science Reviews*, https://doi:10.1016/j.earscirev.2018.05.002 **(2017)**

6. Bui D.T., Pradhan B., Lofman O., Revhaug I. and Dick O.B., Spatial prediction of landslide hazards in Hoa Binh province (Vietnam): a comparative assessment of the efficacy of evidential belief functions and fuzzy logic models, *CATENA*, **96**, 28–40 **(2012)**

7. Bui D.T., Tuan T.A., Klempe H., Pradhan B. and Revhaug I., Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression and logistic model tree, *Landslides*, **13**, 361–378 **(2016)**

8. Chen W., Peng J., Hong H., Shahabi H., Pradhan B., Liu J., Zhu A.X., Pei X. and Duan Z., Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County,

Jiangxi Province, China, *Sci. Total Environ*., **626**, 1121–1135, https://doi.org/10.1016/j.scitotenv.2018.01.124 **(2018)**

9. Chen W., Pourghasemi H.R. and Zhao Z., A GIS-based comparative study of Dempster-Shafer, logistic regression and artificial neural network models for landslide susceptibility mapping, *Geocarto Int.*, **32**, 367–385 **(2017)**

10. Chung C. and Fabbri A., Probabilistic prediction models for landslide, *Photogramm Eng Remote Sens*, **65(12)**, 1388–1399 **(1999)**

11. Chung C.J. and Fabbri A., Validation of Spatial Prediction Models for Landslide Hazard Mapping, *Natural Hazards*, **30**, 451–472 **(2003)**

12. Conforti M. et al, Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo River catchment (northern Calabria, Italy), *Catena*, http://dx.doi.org/10.1016/j.catena.2013.08.006 **(2013)**

13. Cruden D.M., A simple definition of a landslide, *Bulletin of the International Association of Engineering Geology*, **43**, 27, https://doi.org/10.1007/BF02590167 **(1991)**

14. Dai F.C. and Lee C.F., Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong, *Geomorphology*, **42**, 213–228, https://doi.org/10.1016/S0169-555X(01)00087-3 **(2002)**

15. Dang V.H., Hoang N.D., Nguyen L.M.D., Bui D.T. and Samui P., A Novel GIS-Based Random Forest Machine Algorithm for the Spatial Prediction of Shallow Landslide Susceptibility, *Forests*, **11(1),** 118, https://doi.org/10.3390/f11010118 **(2020)**

16. Delcaillau B., Amrhar M., Namous M., Laville E., Pedoja K. and Dugué O., Transpressional tectonics in the Marrakech High Atlas: Insight by the geomorphic, *Geomorphology*, **134**, 344–362 **(2011)**

17. Delcaillau B., Laville E., Amhrar M., Dugué O., Namous M. and Pedoja K., Quaternary evolution of the Marrakech High Atlas and morphotectonic evidences of the Tizi N'Test Fault activity, Morocco, *Geomorphology*, **118**, 262–279 **(2010)**

18. Demir G., Aytekin M. and Akgun A., Landslide susceptibility mapping by frequency ratio and logistic regression methods: an example from Niksar-Resadiye (Tokat, Turkey), *Arab J Geosci*., **8(3)**, 1801–1812 **(2015)**

19. El Alaoui E.L., Fels A., Alaa N., Bachnou A. and Rachidi S., Flood frequency analysis and generation of flood hazard indicator maps in a semi-arid environment, case of Ourika watershed (western High Atlas, Morocco), *Journal of African Earth Sciences*, **141**, 94-106 **(2018)**

20. El Khalki E., Tramblay Y., Saidi M.E., Bouvier C., Hanich L., Benrhanem M. and Alaouri M., Comparison of modeling approaches for flood forecasting in the High Atlas Mountains of Morocco, *Arabian Journal of Geosciences*, **11**, 10-20, https://doi:10.1007/s12517-018-3752-7 **(2018)**

21. Elmoulat M. and Ait Brahim L., Landslides susceptibility mapping using GIS and weights of evidence model in Tetouan-Ras-Mazari area (Northern Morocco), *Geomatics, Natural Hazards and Risk*, 1306-1325, https://doi:10.1080/19475705.2018.1505666 **(2018)**

22. Elmoulat M., Ait Brahim L., Mastere M. and Jemmah A.I., Mapping of mass movements susceptibility in the Zoumi region using satellite image and GIS technology (Moroccan Rif), *Int J Sci Eng Res*., **6(2)**, 210–217 **(2015)**

23. Fawcett T., An introduction to ROC analysis, *Pattern Recognit. Lett.*, **27,** 861–874 **(2006)**

24. Frizon de Lamotte D., Crespo-Blanc A., Saint Bezar B., Fernandez M., Zeyen H., Ayarza P., Robert-Charrue C., Chalouan A., Zizi M., Teixell A., Arboleya M.L., Alvarez-Lobato F., Julivert M. and Micard A., Betics Alboran Sea, Rif Moroccan Meseta, High Atlas, Jbel Saghro, Tindouf basin, In Cavazza W., Roure F., Spakman W., Stampfli G.M. and Ziegler P.A., eds., The TRANSMED Atlas—The Mediterranean Region from Crust to Mantle, Springer, Berlin **(2004)**

25. Guo C., Montgomery D., Zhang Y., Wang K. and Yang Z., Quantitative assessment of landslide susceptibility along the Xianshuihe fault zone, Tibetan Plateau, China, *Geomorphology*, doi:10.1016/j.geomorph.2015.07.012 **(2015)**

26. Hervás J. and Bobrowsky P., Mapping: Inventories, Susceptibility, Hazard and Risk, In Sassa K. and Canuti P., eds., Landslides – Disaster Risk Reduction, Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-540-69970-5_19 **(2009)**

27. Hong H., Liu J. and Zhu A.X., Landslide susceptibility evaluating using artificial intelligence method in the Youfang district (China), *Environ Earth Sci*., **78**, 488, https://doi.org/10.1007/s12665-019-8415-9 **(2019)**

28. Hughes P.D., Fink D., Fletcher W.J. and Hannah G., Catastrophic rock avalanches in a glaciated valley of the High Atlas, Morocco: 10Be exposure ages reveal a 4.5 ka seismic event, *Geol. Soc. Am. Bull.*, **126**, 1093e1104 **(2014)**

29. Juliev M., Mergili M., Mondal I., Nurtaev B., Pulatov A. and Hübl J., Comparative analysis of statistical methods for landslide susceptibility mapping in the Bostanlik District, Uzbekistan, *Science of The Total Environment*, doi:10.1016/j.scitotenv.2018.10.431 **(2018)**

30. Kanungo D.P., Arora M.K., Gupta R.P. and Sarkar S., Landslide risk assessment using concepts of danger pixels and fuzzy set theory in Darjeeling Himalayas, *Landslides*, **5**, 407–416 **(2008)**

31. Bragagnolo L., da Silva R.V. and Grzybowski J.M.V., Artificial neural network ensembles applied to the mapping of landslide susceptibility, *Catena*, **184**, https://doi.org/10.1016/j.catena.2019.104240 **(2020)**

32. Lee S. and Min K., Statistical analysis of landslide susceptibility at Yongin, Korea, *Environ. Geol*., **40**, 1095–1113 **(2001)**

33. Lee S., Ryu J.H., Won J.S. and Park H.J., Determination and application of the weights for landslide susceptibility mapping using an artificial neural network, *Eng. Geol*., **71(3)**, 289–302 **(2004)**

34. Lin L., Lin Q. and Wang Y., Landslide susceptibility mapping on a global scale using the method of logistic regression, *Nat. Hazards Earth Syst. Sci*., **17**, 1411–1424, https://doi.org/10.5194/nhess-17-1411-2017 **(2017)**

35. Mastere M., Van Vliet-Lanoë B. and Ait Brahim L., Land use mapping and its relation to mass wasting and gullying in North-Western Rif (Morocco), *Géomorphologie: Relief, Processus, Environnement*, **3**, 107-124 **(2013)**

36. Meinhardt M., Fink M. and Tünschel H., Landslide susceptibility analysis in central Vietnam based on an incomplete landslide inventory: comparison of a new method to calculate weighting factors by means of bivariate statistics, *Geomorphology*, **234**, 80–97, doi:10.1016/j.geomorph.2014.12.042 **(2015)**

37. Meten M., Bhandary N.P. and Yatabe R., GIS-based frequency ratio and logistic regression modelling for landslide susceptibility mapping of Debre Sina area in central Ethiopia, *Journal of Mountain Science*, **12**, 1355, https://doi.org/10.1007/s11629-015-3464-3 **(2015)**

38. Missenard Y., Zeyen H., Frizon de Lamotte D., Leturmy P., Petit C., Sébrier M. and Saddiqi O., Crustal versus asthenospheric origin of the relief of the Atlas mountains of Morocco, *J. Geophys. Res*., **111**, B03401, https://doi:10.1029/2005JB003708 **(2006)**

39. Muthu K., Petrou M., Tarantino C. and Blonda P., Landslide Possibility Mapping Using Fuzzy Approaches, Landslide Possibility Mapping Using Fuzzy Approaches, *IEEE Transactions on Geoscience and Remote Sensing*, **46(4)**, 1253-1265, https://doi:10.1109/TGRS.2007.912441 **(2008)**

40. Poonam Rana N., Champati Ray P.K., Bisht P., Bagri D.S., Wasson R.J. and Sundriyal Y.P., Identification of landslide-prone zones in the geomorphically and climatically sensitive Mandakini valley, (central Himalaya), for disaster governance using the Weights of Evidence method, *Geomorphology*, **284**, 41–52, doi: 10.1016/j.geomorph.2016.11.008 **(2017)**

41. Poudyal C.P., Chang C., Oh H.J. and Lee S., Landslide susceptibility maps comparing frequency ratio and artificial neural networks: a case study from the Nepal Himalaya, *Environ Earth Sci*, **61**, 1049, https://doi.org/10.1007/s12665-009-0426-5 **(2010)**

42. Pradhan B., A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS, *Comput Geosci*, **51(1)**, 350–365 **(2013)**

43. Pradhan B. and Lee S., Regional landslide susceptibility analysis using back-propagation neural networks model at Cameron Highland, Malaysia, *Landslides*, **7(1)**, 13–30 **(2010)**

44. Pradhan B. and Youssef M.A., Manifestation Of Remote Sensing Data And Gis On Landslide, *Arabian Journal of Geosciences*, 319-326, https://doi:10.1007/S12517-009-0089-2 **(2009)**

45. Saidi M., Daoudi L., Aresmouk M., Fniguire F. and Boukrim S., Les crues de l'oued Ourika (Haut Atlas, Maroc) : Événements extrêmes en contexte montagnard semi-aride, *Comunicações Geológicas*, **97(1)**, 113-128 **(2010)**

46. Teixell A., Arboleya M.L., Julivert M. and Charroud M., Tectonic shortening and topography of the central High Atlas (Morocco), *Tectonics*, **22**, 1051, https://doi:10.1029/2002TC001460 **(2003)**

47. Varnes D.J., Landslides hazard zonation, A review of principales and practice, Paris, 63 **(1984)**

48. Wang L.J., Guo M., Sawada K., Lin J. and Zhang J., Landslide susceptibility mapping in Mizunami City, Japan, A comparison between logistic regression, bivariate statistical analysis and multivariate adaptive regression spline models, *Catena*, **135**, 271–282, http://dx.doi.org/10.1016/j.catena.2015.08.007 **(2015)**

49. Xiong T., Indrawan I.G.B. and Putra D.P.E., Landslide Susceptibility Mapping Using Analytical Hierarchy Process, Statistical Index, Index of Enthropy and Logistic Regression Approaches in the TinalahWatershed, Yogyakarta, *Journal of Applied Geology*, **2(1)**, 66–82, http://dx.doi.org/10.22146/jag.30 **(2017)**.

(Received 05th January 2021, accepted 09th March 2021)