

Article

Landslide Susceptibility Assessment at Mila Basin (Algeria): A Comparative Assessment of Prediction Capability of Advanced Machine Learning Methods

Abdelaziz Merghadi ^{1,*} , Boumezbeur Abderrahmane ¹ and Dieu Tien Bui ²

¹ Research Laboratory of Sedimentary Environment, Mineral and Water resources of Eastern Algeria, University of Tebessa, Tebessa 12002, Algeria; a.boumezbeur@univ-tebessa.dz

² Geographic Information System Group, Department of Business and IT, University of South-Eastern Norway, Gullbringvegen 36, N-3800 Bø i Telemark, Norway; Dieu.T.Bui@usn.no

* Correspondence: abdelaziz.merghadi@gmail.com or a.merghadi@univ-tebessa.dz; Tel.: +213-676-735-046

Received: 7 May 2018; Accepted: 7 July 2018; Published: 10 July 2018



Abstract: Landslide risk prevention requires the delineation of landslide-prone areas as accurately as possible. Therefore, selecting a method or a technique that is capable of providing the highest landslide prediction capability is highly important. The main objective of this study is to assess and compare the prediction capability of advanced machine learning methods for landslide susceptibility mapping in the Mila Basin (Algeria). First, a geospatial database was constructed from various sources. The database contains 1156 landslide polygons and 16 conditioning factors (altitude, slope, aspect, topographic wetness index (TWI), landforms, rainfall, lithology, stratigraphy, soil type, soil texture, landuse, depth to bedrock, bulk density, distance to faults, distance to hydrographic network, and distance to road networks). Subsequently, the database was randomly resampled into training sets and validation sets using 5 times repeated 10 k-folds cross-validations. Using the training and validation sets, five landslide susceptibility models were constructed, assessed, and compared using Random Forest (RF), Gradient Boosting Machine (GBM), Logistic Regression (LR), Artificial Neural Network (NNET), and Support Vector Machine (SVM). The prediction capability of the five landslide models was assessed and compared using the receiver operating characteristic (ROC) curve, the area under the ROC curves (AUC), overall accuracy (Acc), and kappa index. Additionally, Wilcoxon signed-rank tests were performed to confirm statistical significance in the differences among the five machine learning models employed in this study. The result showed that the GBM model has the highest prediction capability (AUC = 0.8967), followed by the RF model (AUC = 0.8957), the NNET model (AUC = 0.8882), the SVM model (AUC = 0.8818), and the LR model (AUC = 0.8575). Therefore, we concluded that GBM and RF are the most suitable for this study area and should be used to produce landslide susceptibility maps. These maps as a technical framework are used to develop countermeasures and regulatory policies to minimize landslide damages in the Mila Basin. This research demonstrated the benefit of selecting the best-advanced machine learning method for landslide susceptibility assessment.

Keywords: landslide; susceptibility mapping; machine learning; GIS; Algeria

1. Introduction

The severe landslides affecting the Mila Basin (located in the North-East region of Algeria) have created serious threats not only to the environment and human settlements but also inflicted economic burdens to local authorities by the non-ending reconditioning and restoration projects. In addition, these landslides affect the current landscape evolution of the basin; therefore, predicting

and delineating landslides are crucial tasks to reduce their associated damages. However, landslide prediction and delineation remain challenging tasks in the basin due to the complex nature of landslides.

Fortunately, the advancements achieved in machine learning and Geographic Information Systems (GIS) in the last decade have provided a plethora of quantitative methods and techniques for landslide modeling. Consequently, various models have been proposed and implemented successfully for modeling landslides that help in understanding landslide patterns and their triggering mechanism [1]. The literature reviews showed that physical-based models are capable of delivering the highest prediction accuracy [2]. Nonetheless, for large-scale analysis (similar to this case study), physical-based models require a fair amount of detailed data information to provide reliable results, which is unbelievably expensive [3]. As a result, statistical and machine learning models can be considered a viable option to use. Basically, machine learning methods for landslide are based on the assumption that *“previous, current and future landslide failures do not happen randomly or by chance, but instead, failures follow patterns and share common geotechnical behaviors under similar conditions of the past and the present”* [4]. This requires collecting and preparing an accurate and large database (i.e., a geospatial database of landslide inventory and conditioning factors) with maximum details available. Then, models based on these methods are trained and validated using that database and the resulting models are used to generate landslide occurrence probability grids [2].

Machine learning (ML) is one of the most effective methods for solving non-linear geo-spatial problems like landslides susceptibility, using either regression or classification. In fact, ML has proven to be ideal for addressing large-scale analysis problems where theoretical knowledge about the problem is still incomplete [5]. After all, ML methods do require a significant number of conditioning factors to obtain reliable results. In the literature, several studies have been able to implement and compare machine learning models in landslide susceptibility modeling such as Artificial Neural Networks (NNET) [1,6,7]; Support Vector Machines (SVM) [1,6,8–10]; Decision Trees (DT) [3,11]; Logistic Regression (LR) [1,8,11]; and ensemble methods such as Boosted Trees (BT) [11,12], and Random Forest (RF) [3,8,11]. Despite the availability of some research concerning machine learning techniques and methods, no solid agreement about which method or technique is the most suitable for a landslide-prone area prediction has been identified [13]. Nevertheless, there's *“No free lunch”* (NFL) (according to Wolpert [14], NFL can be explained as: *“any two algorithms are equivalent when their performance is averaged across all possible problems”*) when it comes to machine learning in general and the spatial prediction of landslides in particular due to the high level of uncertainty behind the process. In fact, no single or particular model can be depicted as the most suitable for all case scenarios. Selecting the most suitable method for landslide spatial prediction depends essentially on the underlined scientific goal for the case study [15]. Additionally, the prediction accuracy of landslide modeling is influenced not only by the quality of the landslide inventories and the influencing factors, but also the fundamental quality of the machine learning algorithm used [2]. Therefore, exploring and experimenting with new methods and techniques for spatially predicting this hazard is highly necessary.

The main goal of this study is to investigate and compare five machine learning algorithms, Random Forest (RF), Gradient Boosting Machine (GBM), Logistic Regression (LR), Artificial Neural Network (NNET), and Support Vector Machine (SVM) for landslide susceptibility mapping at the Mila Basin (Algeria). Additionally, this study aims to implement a meta-modeling approach using Sequential Model-Based Optimization (SMBO) for models that configure hyperparameters. Unlike similar studies [1,3,5–12,16], this approach supports automated expensive hyperparameter optimization in order to provide a useful framework with a reproducible and unbiased optimization process. Moreover, it is important to note that the Mila Basin has suffered (and still) from various landslide disaster problems during the last five years; however, no significant attempt has been conducted to understand the phenomenon.

2. Study Area and Data

2.1. Description of the Study Area

The Mila Basin is situated in the northeastern part of Algeria between longitudes of $5^{\circ}55'15.44''$ E and $6^{\circ}49'42.19''$ E and latitudes of $36^{\circ}36'39.01''$ N and $36^{\circ}11'6.82''$ N and covers an area of 2760 km² distributed mostly over the central parts of the Mila and Constantine provinces. Geographically, the study area is fully surrounded by mountainous ranges that belong to different paleogeographic domains and make up the basin substratum, such as M'cid Aicha and Sidi Driss from the North; Djebel Ossmane and Grouz by the South; Djebel Akhal, Chettaba and Kheneg from the East; and Djebel Boucherf and Oukissene by the West (Figure 1). The elevation of the basin varies from 60 m to 1550 m.

The basin is characterized by asymmetrical elongated geometrical form drained by a dense and hierarchical hydrographic network in generally S-N direction [17]. The local climate is semi-arid with a mild winter surrounded by sub-humid fresh climate typical for a mountainous landscape [18]. Annual mean precipitation is around 600 mm/year, in which the precipitation is mainly in the short wet season (usually between October and February). The dry season is long, lasting from March to September. Land use is mostly for bare lands, cereals crops or wild herbs. This low-density vegetation is good for the agriculture industry but encourages land degradation and instabilities by soil erosions.

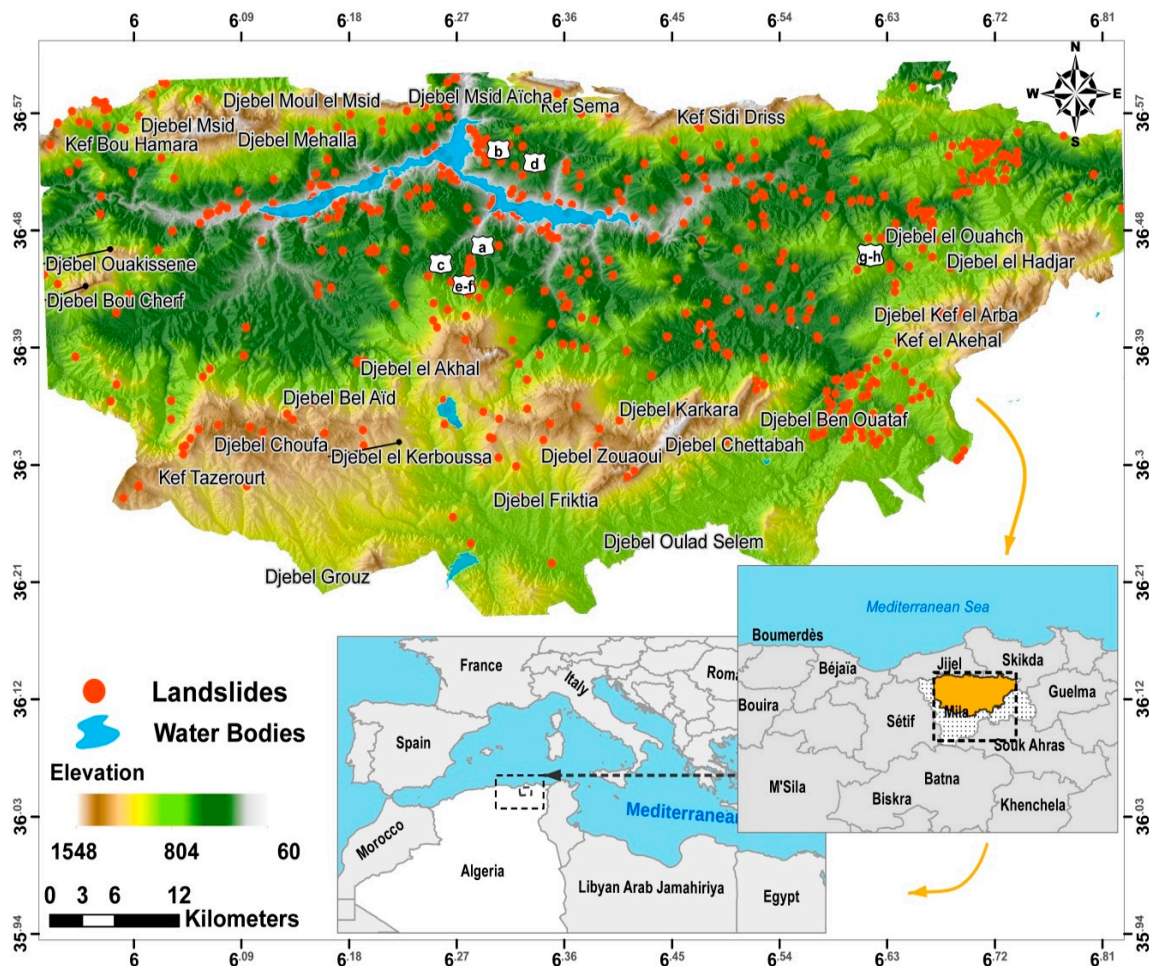


Figure 1. The landslide inventory map and location of the study area.

The local geology consists of different lithostratigraphic units and can be grouped into two groups (called 'series'): (1) Substratum series and (2) Post-nappes series [19] (Figure 2). The Substratum series formulates both the lower base and the substratum of the basin whereas the Post-nappes

series formulates a cover on the top, which has slightly affected by recent tectonic deformations (Table 1). The study area shows a tectonic complexity due to some severe conjugation of folds, faults, and thrusts of different ages and styles. Two general systems of lineaments exist: (1) Diagonal system of NE-SW and NW-SE and (2) Vertical system (also known as “Alpine phase”) of N-S and E-W orientations. The Diagonal lineament system during the late Eocene-Lutetian was directly responsible for creating some important structures (i.e., folds and horst-graben). These structures formulate a base for depositing detritus materials during the Neogene. On the other hand, the Vertical lineament system belongs to a recent compression phase that is responsible for the current morpho-structure of the study area [19].

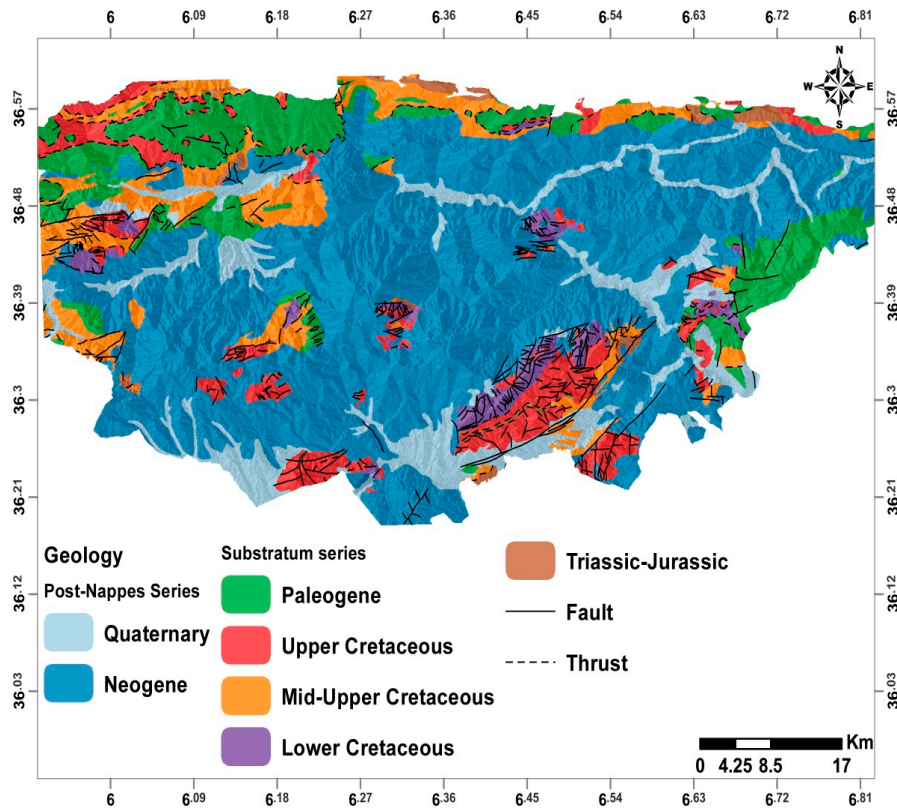


Figure 2. The geological map of the study area.

Table 1. The outcropping geological formations in the study area.

Unit	Period	Epoch	Description
Post-nappes		Quaternary	Alluvium, colluvium, scree, detritus deposits and slopes formations like terraces
		Neogene	Predominantly detritus composed of clay, marl, limestone, conglomerates, sandstone, sand, lacustral limestone and evaporitic formations
Substratum	Paleogene	Eocene	Limestone, cherty limestone, and platted marls
		Paleocene	Opaque to somber marls
	Cretaceous	Upper and Mid-Upper Cretaceous	Marl dominance (variation are ranging from different horizons of gray marly limestone, alternating marl, and limestone, blueish marl, massive bars of limestone, to alternating marl, cherty limestone, and thin micritic limestone all surmounted by grey marls with conglomerate interbeds)
		Lower Cretaceous	Mainly marly limestone and neritic limestone
	Jurassic	Mostly thick carbonate formations (dolostone, limestone, and cherty limestone)	
	Triassic	Evaporitic and clayey deposits	

2.2. Data Used

A key step to successful landslide modeling is preparing an accurate database that serves as the input dataset. For the landslide susceptibility assessment, collecting and constructing a landslide inventory map would be obviously the first and foremost step. In addition to the inventory, selecting landslide related variables to implement is very important [20]. A literature review shows that landslide factors were selected depending on the study case, the scale of the analysis, and data availability [21]. Therefore, a multi-sources geospatial database that includes an inventory map and landslide conditioning factors was constructed.

In this research, the geospatial database was developed and processed in the QGIS, Saga and R software. The database consists of information layers derived from multiple geo-environmental sources (geology, topography, precipitation, landuse, and so forth).

2.2.1. Landslide Inventory Map

In this study, a detailed and reliable landslide inventory map from 1985 to 2017 (Figure 1) for circular and planar failures (both shallow and deep landslides) was constructed using two main sources: (1) historical records provided publicly by the local municipality-hauls (Constantine and Mila) with 531 landslide polygons, and (2) using the Google Earth Pro[®] software. 47 landslide polygons were detected and mapped (from 2000 to 2017). On the other hand, the non-landslide samples were extracted by random sampling a unique 578 sample site (equal to the total number of landslide samples) from public stability maps available at DUC (Direction d'Urbanisme et Construction) using PAW (Plan d'Aménagement de Wilya) and PDAU (Plan Directeur d'Aménagement et d'Urbanisme). Extensive field inspections and Google Earth Pro software were performed to verify the landslide and non-landslide samples (Figure 3).

The mapped landslides are both shallow (depth < 5 m) and deep-seated (depth > 5 m). They mainly occur in the Neogene complex and central middle part of the basin (Figure 2) and are characterized by different volumes ranging from 182 m³ to 620,000 m³. According to the survey campaigns achieved by local authorities (2003–2017), the slopes in the study area fail under the conjunction of both predisposition factors (i.e., geology, lithology geomorphology, and faults) and triggering factors (i.e., intense and persistent meteorological events, human activities, and so forth), resulting in landslides of different sizes and types. Reports suggest that the long and persistent periods of intense to moderate rainfall are the main culprit in triggering and/or reactivating existent deep-seated landslides due to the high amount of water infiltrating underground. On the contrary, short and intense to moderate rainstorms/precipitations indirectly affect slope stability by intensive erosive processes [19,22].

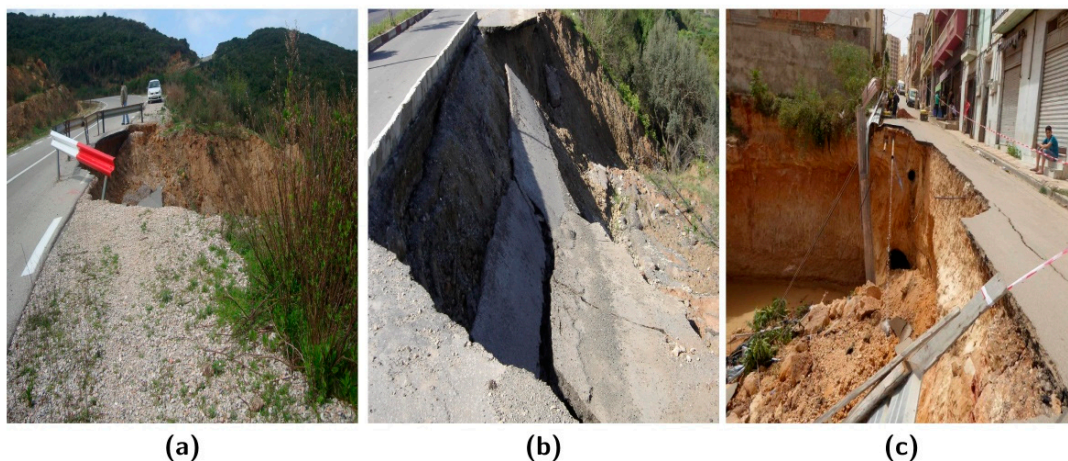


Figure 3. Cont.

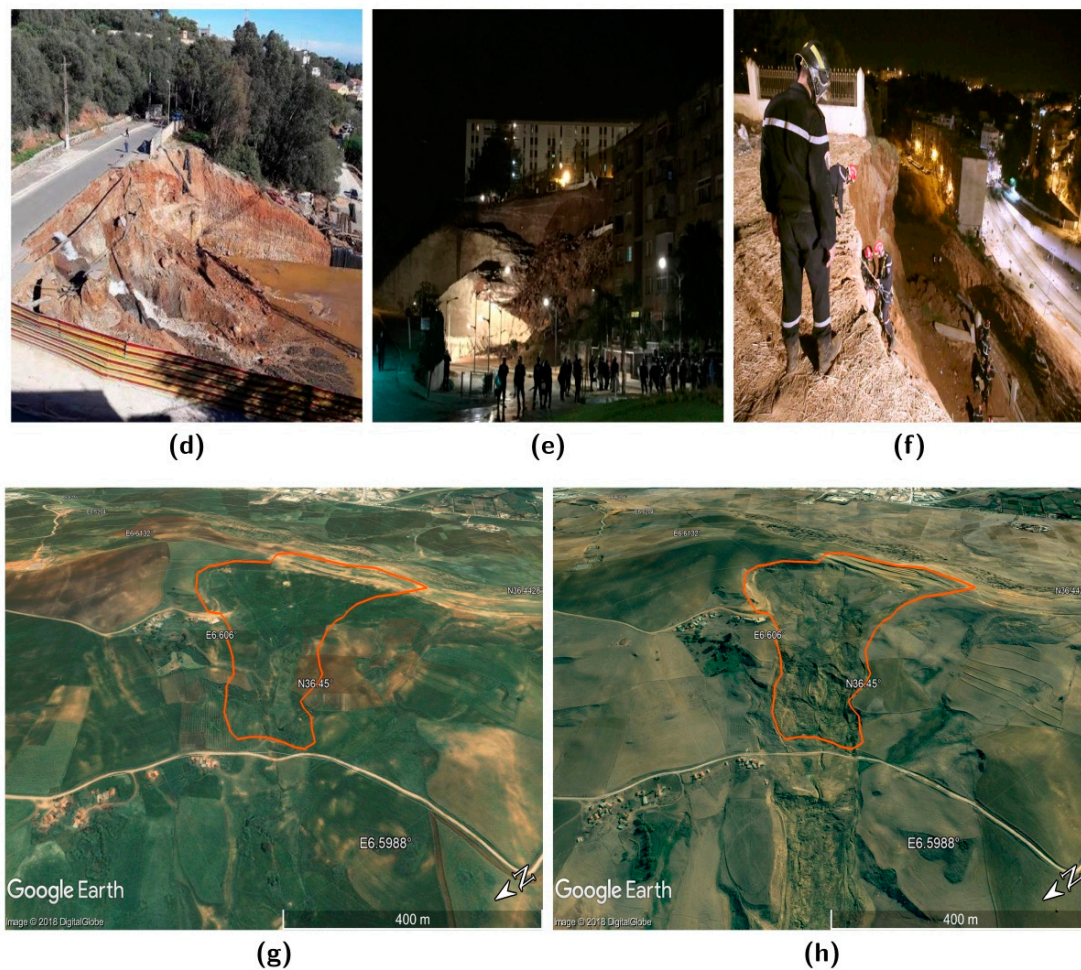


Figure 3. Landslide examples (Source: Mila and Constantine municipalities, Location: see Figure 1): (a) RN 79a, (Type: Deep-Rotational landslide; Date: October 2011); (b) Sibari (Type: Shallow-Planar landslide; Date: February 2008); (c) Mila (Type: Deep-Rotational landslide; Date: September 2013); (d) Grarem (Type: Planar landslide; Date: June 2015); (e,f) Mila (Type: Deep-Rotational landslide; Date: October 2017); (g,h) Didouche Mourad (Type: Deep-Rotational landslide; Date Left: August 2003, Date Right: September 2005).

2.2.2. Landslide Conditioning Factors

Despite the fact that there are no clear guidelines about the proper factors to use for such a kind of analysis [23], 16 conditioning factors (Figure 4) were selected for this case study based on (1) field survey observations; (2) survey campaign reports achieved by local authorities; (3) the most commonly used factors in the literature for landslide susceptibility analysis [1,3,9]; (4) geo-environmental factors of the study area that may directly or indirectly affect landslides, and can be used as predisposing factors [24]; and (5) the scale of the analysis and data availability for the case study [21].

The Digital elevation model (DEM) for the study area with a resolution of 30 m was derived from the NASA Shuttle Radar Topography Mission Global (SRTMGL1) Version 3 (<http://www2.jpl.nasa.gov/srtm>). Using the DEM, five geomorphometric factors were extracted: Altitude (Figure 4A), Slopes (Figure 4B), Aspects (Figure 4C), Topographic Wetness Index (TWI) (Figure 4D), and Landforms (Figure 4E). On the other hand, 7 geological maps at a scale of 1:50,000 scale provided by ASGA (L'Agence du Service Géologique de l'Algérie) were used to derive the lithology map (Figure 4G), stratigraphy map (Figure 4H), and the distance to the faults map (Figure 4N). The rainfall map (Figure 4F) was generated using the annual mean precipitation at 7 meteorological stations during the period of 1985 to 2017 using the Inverse Distance Weighed method. The precipitation data were

provided by ANRH (L'Agence Nationale des Ressources Hydrauliques) and ONM (Office National de Météo). The remaining factors; Bulk Density (Figure 4M), Depth to Bedrock (Figure 4L), Distance to Hydrographic Network (Figure 4O), Distance to roads (Figure 4P), Soil texture (Figure 4J), Landuse (Figure 4K), and Soil types (Figure 4I)—were provided by the Mila and Constantine municipalities.

Detailed classes of all the used factors are shown in Table 2. The reclassification process (the class intervals and the total number of classes) of the continuous factors (altitude, slopes, rainfall, and so forth) was performed automatically using the Geometrical Intervals reclassification method due to the non-uniform distribution of the data in those factors. On the other hand, the categorical factors (Lithology, Stratigraphy, and so forth) remained unmodified.

Table 2. The spatial relationship between the landslide conditioning factors and landslides by frequency ratio.

Conditioning Factors	Class	Class Percentage (%)	Landslide Percentage (%)
Altitude (m)	60–326.047	8.786	19.550
	326.047–597.105	36.055	48.789
	597.105–813.952	28.967	18.512
	813.952–1003.694	18.637	7.785
	1003.694–1722	7.555	5.363
Slopes (°)	0–5.543	26.667	21.107
	5.543–11.394	39.877	37.889
	11.394–18.16987664	23.325	28.374
	18.169–27.101	8.299	10.900
	27.101–78.530	1.831	1.730
Aspects	Flat	0.757	1.038
	1st Quadrant (0° to 90°)	23.709	26.298
	2nd Quadrant (90° to 180°)	28.195	25.260
	3rd Quadrant (180° to 270°)	22.593	21.453
	4th Quadrant (270° to 360°)	24.746	25.952
Topographic Wetness Index (TWI)	0.034–3.550	8.521	3.979
	3.550–5.481	50.807	21.280
	5.481–8.997	31.076	67.647
	8.997–15.402	9.597	7.093
Landforms	Steep slope, fine texture, high convexity	6.920	2.422
	Steep slope, coarse texture, high convexity	25.290	32.007
	Steep slope, fine texture, low convexity	41.067	40.830
	Steep slope, coarse texture, low convexity	26.723	24.740
	Gentle slope, fine texture, high convexity	22.043	19.031
	Gentle slope, coarse texture, high convexity	33.809	34.429
	Gentle slope, fine texture, low convexity	39.618	42.907
Gentle slope, coarse texture, low convexity	4.460	3.633	
Rainfall (mm/Year)	403–593.263	0.070	0.000
	593.263–711.030	3.353	5.190
	711.030–901.294	50.109	48.097
	901.294–1208.684	44.909	45.156
Lithology	Alluvium	1.629	1.557
	Claystone	13.055	16.090
	Colluvium-Detritus Deposits-Scree	16.184	16.263
	Limestone	5.846	6.920
	Marl	10.668	13.668
	Neogene Complex Sandstone	3.173	4.152
Stratigraphy	Quaternary	2.225	1.730
	Neogene	24.557	29.585
	Paleogene	30.166	22.318
	Upper Cretaceous	61.891	61.246
	Upper-Mid Cretaceous	7.943	16.436
	Lower Cretaceous	10.793	18.858
	Triassic-Jurassic	34.420	21.626

Table 2. Cont.

Conditioning Factors	Class	Class Percentage (%)	Landslide Percentage (%)
Soil type	Calcisols	25.679	27.163
	Cambisols	17.017	26.125
	Luvisols	12.090	6.228
	Leptosols	25.701	40.311
	Podzols	30.189	28.893
	Regosols	32.091	24.740
	Vertisols	12.019	6.055
Soil Texture (Texture)	Clay	5.331	7.439
	Sandy Clay	4.057	2.941
	Clay Loam	9.438	7.612
	Silty Clay Loam	8.220	7.612
	Sandy Clay Loam	9.545	7.785
Landuse	Water Bodies	56.050	59.862
	Artificial Surfaces	7.359	6.747
	Forests	19.011	25.433
	Grasslands	1.692	1.384
	CropLand	59.084	60.035
	Bareland	0.796	1.730
Depth to Bedrock (cm) (DepthBR)	49–574.750	19.417	11.419
	574.7502397–761.629	9.797	1.211
	761.6293378–1287.379	15.499	19.723
	1287.379578–2766.481	50.172	58.651
	2766.481936–7479	13.075	8.997
Bulk Density (Kg/m ³) (Bdensity)	1209–1394.941	5.218	5.882
	1394.941–1463.333	3.775	2.249
	1463.333–1521.039	2.464	3.287
	1521.039–1754	1.472	3.979
Distance to Faults (m) (FDist)	0–581	13.958	17.993
	581–4784.550	7.565	6.920
	4784.550–8192	4.753	6.228
Distance to Hydrographic Network (m) (WDist)	0–300	26.129	23.702
	300–750	46.124	41.176
	750–1500	10.521	14.014
	1500–3000	56.840	61.419
	3000–5856	12.376	6.401
Distance to Roads networks (m) (RDist)	0–908.103	7.572	4.498
	908.103–2612.509	8.614	9.862
	2612.509–5811.481	2.624	2.768
	5811.481–11957	1.453	1.038

Geomorphometric factors such as altitude, slopes, and aspects are frequently used in landslide susceptibility analysis due to the crucial effect of terrain types on slope instability either directly or indirectly by (1) increasing or reducing the shear strength; (2) controlling the microclimatic parameters such as exposure to sunlight, wind, rainfall intensity, and the slope material properties; and (3) controlling the landscape forms [12,25]. Furthermore, factors such as landforms and TWI are highly influential to landslide occurrences. The former specifically derive a classification for the landscape based on three-part geometric signature (i.e., slopes, convexity and surface texture) as proposed by Iwahashi and Pike [26]. The latter indicates the effect of topography on the location and the size of the saturated source area of run-off generation, which is highly related to the hydrogeological conditions that influence surface run-off and infiltration [25]. According to Beven and Kirkby [27], TWI can be calculated using the following equation:

$$TWI = \ln \left(\frac{A_s}{\tan \beta} \right) \quad (1)$$

where A_s is the specific catchment area (m^2/m) and β is the local slope in degree.

Geomorphometric factors are not the only factors that may influence landslide occurrence. In fact, other factors such as lithology, stratigraphy, land use, rainfall, soil types, soil texture, depth to bedrock,

bulk density, and proximity factors (distance to faults, distance to hydrographic network, and distance to road networks) are indirectly related to landslide occurrence. They induce (1) shear strength and cohesion, (2) permeability, (3) weathering of slopes materials, (4) erosion of slopes footing, and (5) the saturation of slopes.

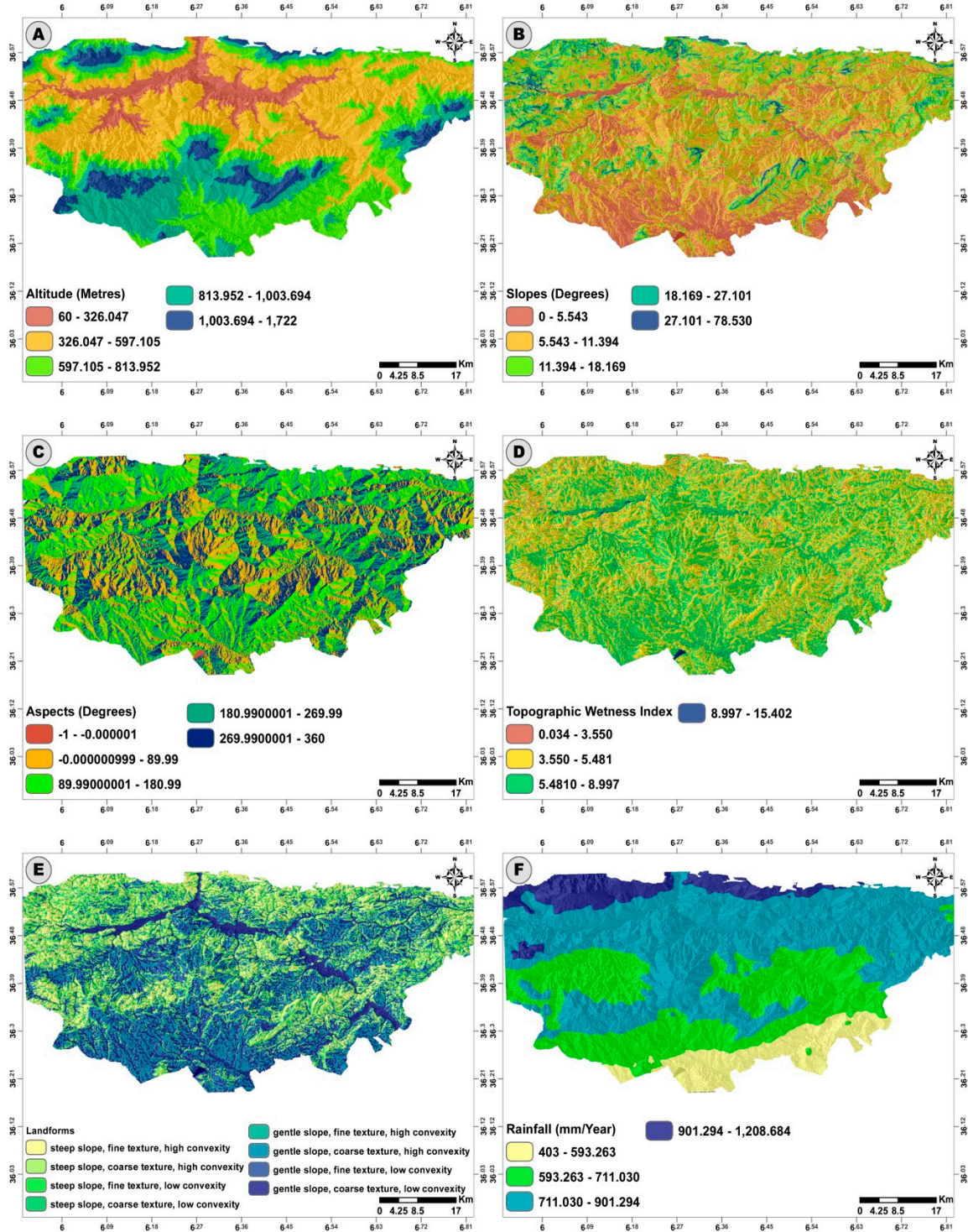


Figure 4. Cont.

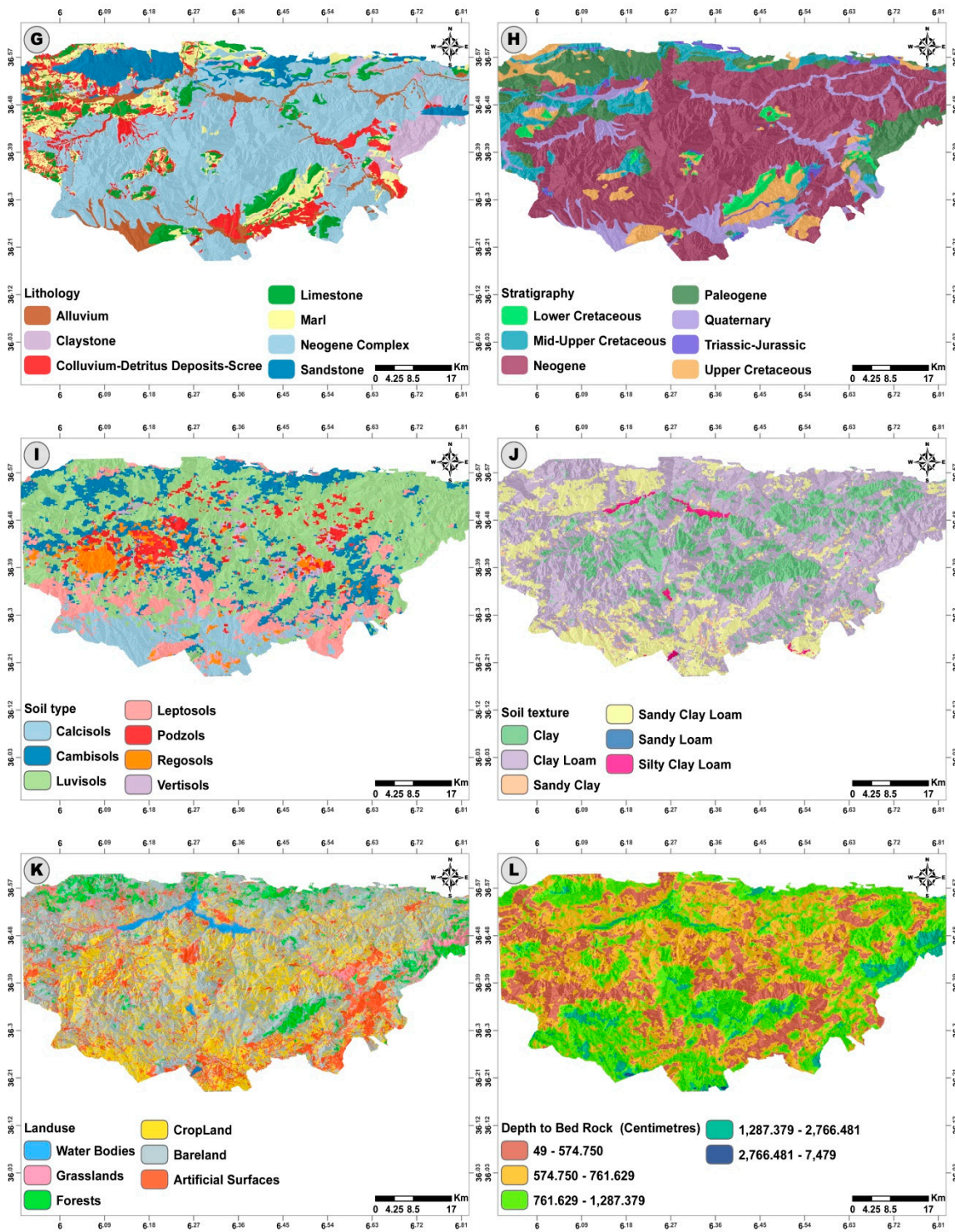


Figure 4. Cont.

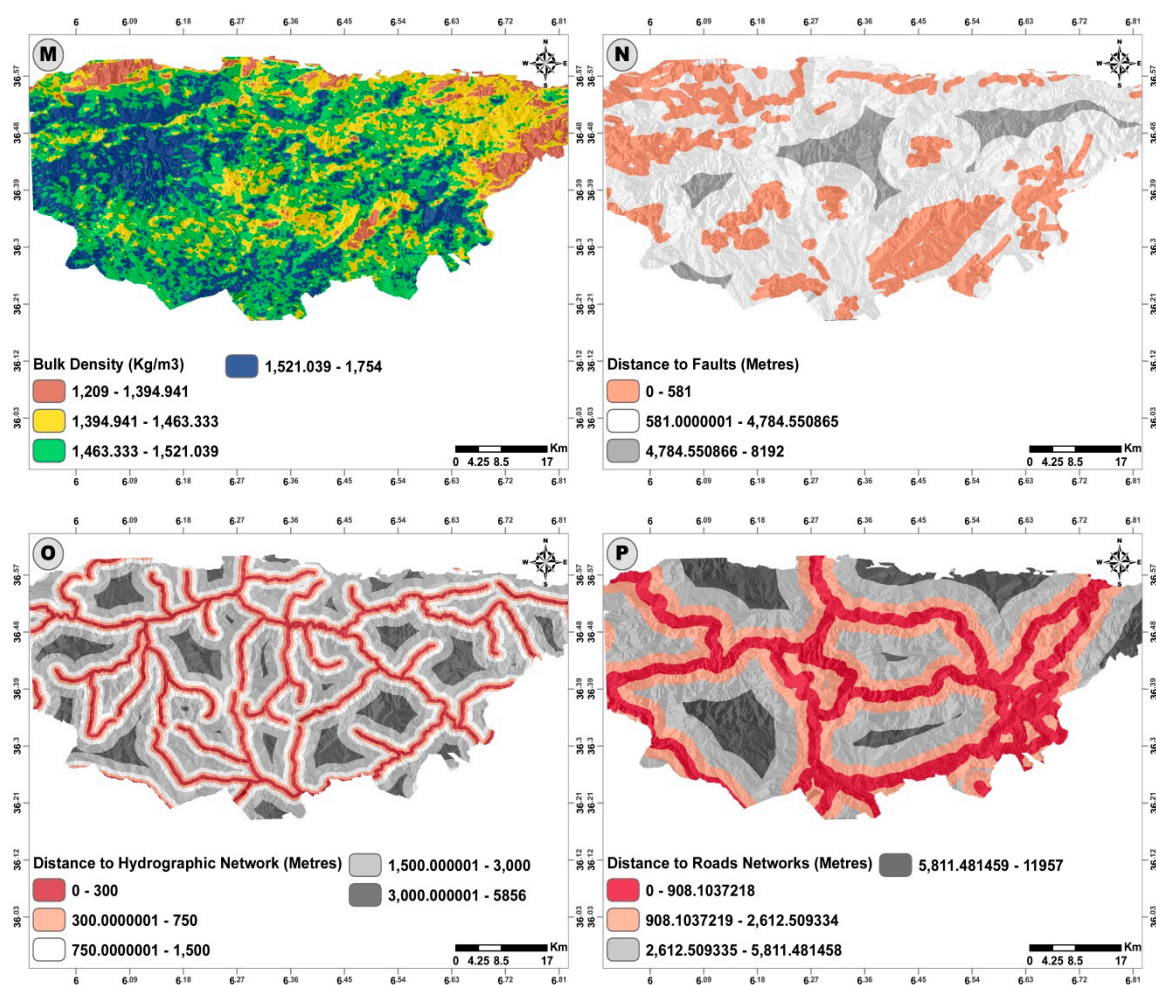


Figure 4. The landslide conditioning factors maps: (A) altitude, (B) slope, (C) aspect, (D) Topographic Wetness Index (TWI), (E) landforms, (F) rainfall, (G) lithology, (H) stratigraphy, (I) soil type, (J) soil texture, (K) land use, (L) depth to bedrock, (M) bulk density, (N) distance to faults, (O) distance to hydrographic network, and (P) distance to road networks.

3. Methods

3.1. Random Forest

RF is an ensemble approach to decision trees such that each tree fits a data subset sampled independently using bootstrapping [28]. RF is known to provide a robust error rate with respect to the outliers in predictors, due to the random selection at each split node depending on the two data objects, namely, Out-Of-Bag (OOB) and proximities. OOB data is used to get both variable importance estimations and an internal unbiased OOB error (the classification error) as trees are added to the forest while bagging is used to randomly select samples of variables as the training dataset for model calibration. For each variable, the function determines the model prediction error if the values of that variable are permuted across the OOB observations. Proximities, on the other hand, are used to replace missing data, locating outliers, and producing illuminating low-dimensional views of the data and can only be calculated after each tree is fitted on for each pair of cases, then normalizing it by dividing over it the total number of fitted trees.

3.2. Gradient Boosting Machine

Gradient Boosting Machine (GBM) or simply Gradient boosting is an ensemble of weak learners, namely, regression trees that benefit from boosting by adding weak learners using a functional gradient descent associated with the whole ensemble to minimize the loss function as much as possible [29]. The rationale behind GBM is that the learning process consecutively introduces weak learners using a functional gradient descent in stage-wise additive approach sequentially allowing the algorithm to enhance the overall accuracy simply by readjusting previous error terms when new weak learners are added.

GBM involves three elements: (1) the loss function to be optimized based on the objective function to be solved; (2) the weak learner to make predictions, specifically a decision tree that is constructed in a greedy manner by choosing the best split points based on specific scores; and (3) an additive model to add weak learners to minimize the loss function, therefore, a weighted combination of classifiers that optimizes the cost using gradient descent in the function space [30].

3.3. Logistic Regression

Logistic regression (LR) is a particular case of the generalized linear model [31] configured to provide a binary form of result. The ability to find the best fitting function to describe the nonlinear relationship between the presence or absence of landslides and a set of conditioning factors combined with practically zero hyperparameters to tune in makes LR so compelling to be a baseline model in susceptibility analysis mapping. Basically, logistic regression relates the probability of landslide occurrence to a link function (in this case “logit”) assumed to contain the conditioning factors on which landslide occurrence may depend, where the relationship between the occurrence and its dependency on conditioning factors can be expressed by the following (Equation (2)):

$$\hat{P} = \frac{1}{1 + e^{-z}} = \frac{e^z}{1 + e^z} \quad (2)$$

where \hat{P} is the probability of a landslide occurrence and has a range of [0, 1] on an S-shaped curve; z is a linear fitting equation that involves the supplied set of landslide-related variables in the form of the following equation (Equation (3)):

$$Z = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (3)$$

where b_0 is the intercept of the model; b_n is the partial regression coefficients; and X_n is the conditioning variable.

3.4. Artificial Neural Network

An artificial neural network or shortly neural network (NNET) is black-box model defined as a “computational mechanism able to acquire, represent, and compute a mapping from one multivariate space of information to another, given a set of data representing that mapping” [32].

Most NNET models are composed of simple and highly interrelated processing units (neurons) that are in permanent connection with each other. Generally, neurons are located in different layers, and NNET are characterized based on the number of layers and the training procedures. Connections between processing units are physically represented by weights, and each neuron has a rule for summing the input weights and a rule for calculating an output value. More than one layer of neurons can be included in the perceptron in order to cope with non-linearly separable problems, and a multilayer perceptron (MLP) can be obtained.

In this study, we are considering an optimization technique that is regarded as one of the best techniques for solving nonlinear optimization problems (in the absence of constraints) similar to, but more sophisticated than standard backpropagation called the Broyden–Fletcher–Goldfarb–Shanno

(BFGS) method after its creators. The BFGS is a “hill-climbing” procedure, which belongs to a class of algorithms that are based on the “Newton” method, but does not require the Hessian matrix of second derivatives of the objective function to be computed. Instead, it is updated by using gradient vectors. These are called quasi-Newton (or secant) methods. Compared to the popular “Backpropagation” used in most landslide susceptibility studies, BFGS performs better for weight adjustment, simply because using a general algorithm from unconstrained optimization seems to be the most fruitful approach [33], which leads to a faster convergence and provides better results with less complication and parameters to tune in.

3.5. Support Vector Machine

Support vector machine (SVM) is one of the new mathematics tools, which is used as a universal constructive learning procedure based on the statistical learning theory rather than loose analogies with natural learning systems [34]. SVMs provide non-linear solutions to regression and classification problems by transforming the input variables in a large-dimension space, whose inner product is given by positive definite kernel functions, then trained using dual optimization techniques with constraints [35]. Typically, SVMs are designed for two-class problems where both positive and negative objects exist. For two-classes classification problems, SVMs seek to find a hyperplane in the feature space that maximally separates the two target classes [36].

4. Used Methodology

This section focuses on presenting the proposed methodology used to conduct this research. The research was performed using five machine learning models, GBM, LR, NNET, RF, and SVM. Model hyperparameters were tuned and configured using Sequential Model-Based Optimization (SMBO). The analysis was programmed from scratch by the authors in the R environment because (1) of the high flexibility that R offers and (2) to reduce the errors and biases that can be introduced either by evaluating models in different software or platforms that may respond differently, whereas the source code for R is available at GitHub. The overall concept of the used methodology of this research is outlined in Figure 5.

4.1. Construction of the Geospatial Database, the Training Dataset, and the Validation Dataset

As the first step, a geospatial database was constructed from 16 factors and a landslide inventory map using various sources in the QGIS, Saga, and R software. Since the implemented models can handle mixed space variables (numeric and categorical) efficiently, there was no need for dummifying the geospatial database (numeric decoding of categorical variables). Only the target class (landslides) was set to the “Yes” label if the samples are landslide positive, otherwise, it was set to “No”. While this database was mainly used as an input dataset to train the landslide susceptibility models, an independent testing dataset needed to be used to properly assess and validate the trained models. Moreover, landslide samples are scarce and hard to obtain, so, in this case, resampling the input dataset into the training and testing sets would be a mandatory task to obtain reliable results [37]. For that reason, the input dataset was randomly resampled using a 5-times-repeated 10-fold cross-validation (CV) approach (Figure 5A).

Accordingly, the process of a 10-fold cross-validation was started by randomly splitting the input dataset into 10 equal sized folds. Then, each of the nine subsets was used to train landslide models whereas the other subset was used to validate the models, and this procedure was carried out 10 times, respectively. The whole process was repeated 5 times, resulting in 50 training-testing pairs. As result, the models were trained 50 times, and then, the performance measures were finally averaged.

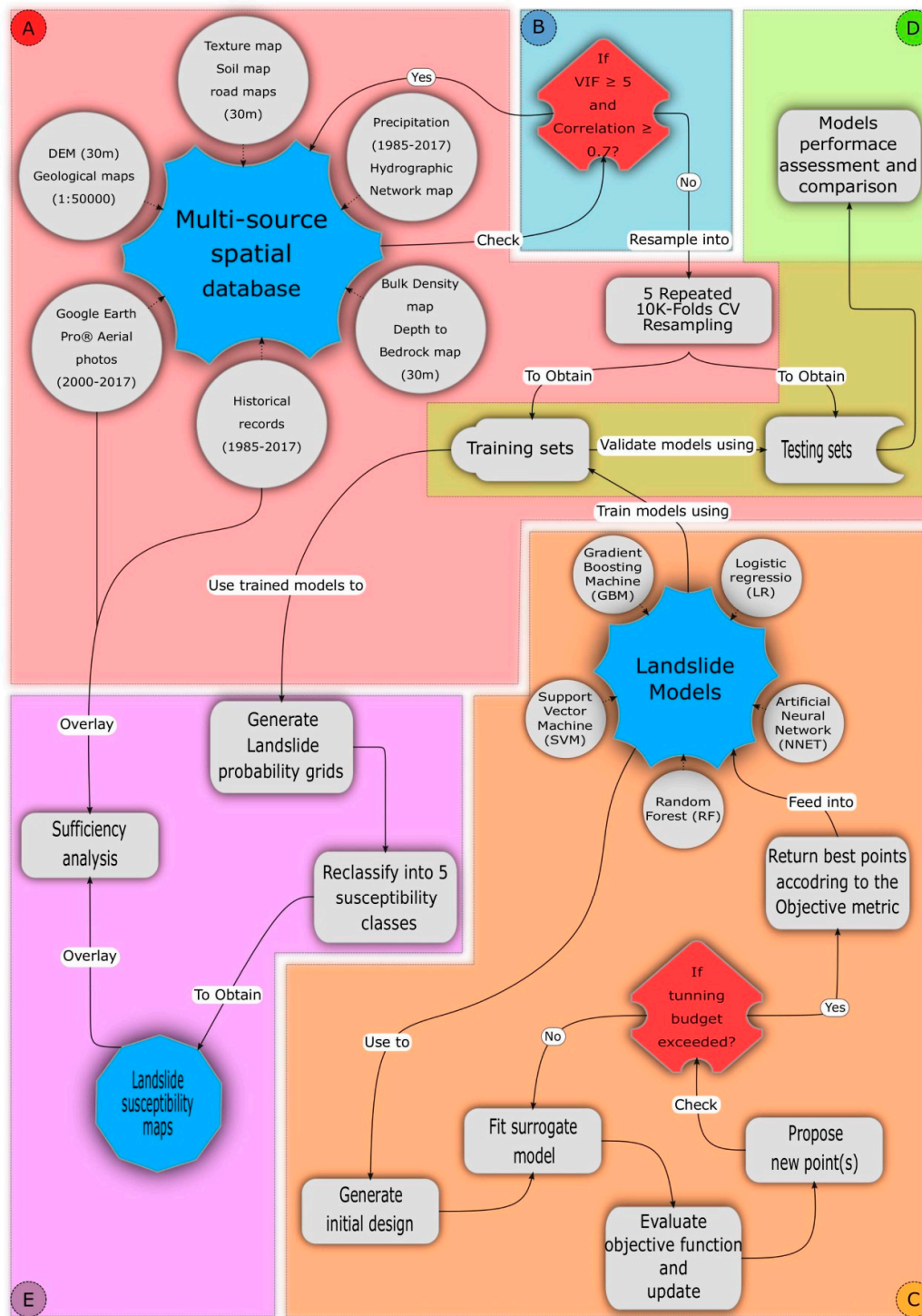


Figure 5. The overall concept of the used methodology for this research: (A) construct a spatial database that will serve as an input dataset for the study from the landslide inventory map and the landslide conditioning factors; (B) Analyzing and optimizing the landslide conditioning factor based on the Pearson correlation and Variance Inflation Factors analysis (VIF) results; (C) Model configuration and implementation using the desired hyperparameters optimization strategy; (D) Model training, validation, and comparison using 5-times-repeated 10 k-folds cross-validations (CV) and the selected performance indicator metrics; (E) susceptibility maps generation and evaluation based on the appropriate assessment strategy.

4.2. Analyzing and Optimizing Landslide Conditioning Factor

It is common for input datasets used in a landslide susceptibility analysis to have a high correlation among certain conditioning factors. This high correlation leads to a faulty modeling with an erroneous system analysis [38]. A possible solution can be performed by a multicollinearity analysis to evaluate the suitability of the underlying assumption used to select the conditioning factors based on the non-independence among factors. To detect and quantify multicollinearity among the used 16 selected variables, Pearson's correlation coefficients [39] can be performed. Nevertheless, in most cases, correlation coefficients are not usually enough, whereas Variance Inflation Factors (VIF) could be implemented. Essentially, Pearson correlations focus on the covariance between each pair of factors divided by the product of their standard deviations (Equation (4)). On the contrary, VIF focuses on the standard error variations of landslide conditioning factors, which imply that the lower, the standard errors, the lower the multicollinearity risk, and the safer the conditioning factor is to implement.

$$r_{x,y} = \sum_{i=1}^n \frac{x_i - \bar{x}}{\sqrt{\sum_{k=1}^n (x_i - \bar{x})^2}} \times \frac{y_i - \bar{y}}{\sqrt{\sum_{k=1}^n (y_i - \bar{y})^2}} \quad (4)$$

where n is the number of samples; x_i , y_i are conditioning factors indexed with i ; \bar{x} is the mean of x_i ; and $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ (analogously, the same applies to \bar{y}).

4.3. Model Configuration and Implementation

Exploring the model's full potential requires correctly tuning a variety of incidental parameter choices and settings [40]. In rare cases, hand-tuning models hyperparameters are enough but in general, there exist methods to do such a task; i.e., Grid search, Random search, Gradient-Based Optimization. However, those methods are widely used and still considered as the main option due to the simplicity and ease of their implementation. Yet, they produce very poor results that lead to (1) costly evaluations (especially when the computational budget is limited); and (2) incorrect assessments about the implemented models, whether they are genuinely bad or simply badly tuned. To avoid the aforementioned problems, we consider a state-of-art technique called Sequential Model-Based Optimization (SMBO) (also known as Bayesian optimization). SMBO can efficiently optimize models by working on a strictly reduced budget for function evaluations and hyperparameters optimization of expensive black-box models. Generally, better results can be achieved using SMBO in fewer experiments compared to traditional techniques (Grid search, Random search, Gradient-Based Optimization) due to (1) the ability to reason about the quality of experiments before they are run [41], and (2) benefiting from the "adaptive capping" to avoid long runs [42].

The main idea behind SMBO is the iterative approximation of the expensive black-box function f using surrogate models (mostly regression models because they are much cheaper to evaluate), which are continuously updated and refined until the evaluation budget is exhausted [43] (usually when the total number of evaluation available is reached or a termination criterion is met). An outline of the SMBO algorithm used in this paper is provided by the "mlr" and "mlrMBO" packages [44] (Figures 5C and 6). The algorithm starts by exploring the parameter space using an initial design D (often constructed in a space-filling fashion). Then, a sequential loop of two alternating stages is evaluated. The first stage is fitting the response surface to the currently available design data. The second stage is optimizing the so-called infill criterion to propose a new promising point x^* for the next expensive evaluation $f(x^*)$ (called y^*). If the optimization budget is exhausted, then the best points associated with the optimal score (in this case the maximum AUC) are returned as a solution for the optimization problem, otherwise, the sequential loop is iteratively repeated.

The overall hyperparameters used for each model are summarized in Table 3 along with their values, short descriptions, and the package used to implement the model.

- 1 Generate an initial design $\mathbf{D} \in \mathbf{X}$;
- 2 Compute $\mathbf{y} = \mathbf{f}(\mathbf{D})$;
- 3 **while** total evaluation budget is not exceeded **do**
- 4 Fit surrogate on \mathbf{D} and obtain $\hat{\mathbf{f}}$ and $\hat{\mathbf{s}}$;
- 5 Get new design point \mathbf{x}^* by optimizing the
infill criterion based on $\hat{\mathbf{f}}$ and $\hat{\mathbf{s}}$;
- 6 Evaluate new point $\mathbf{y}^* = \mathbf{f}(\mathbf{x}^*)$;
- 7 Update: $\mathbf{D} \leftarrow \mathbf{D}(\mathbf{D}, \mathbf{x}^*)$ and $\mathbf{y} \leftarrow (\mathbf{y}, \mathbf{y}^*)$;
- 8 **return** $\mathbf{y}_{min} = \min(\mathbf{y})$ and the associated \mathbf{x}_{min} .

Figure 6. The general sequential model-based optimization approach.

Table 3. The parameters set used by each model along with its respective values.

Model	Package	Parameter	Definition	Value
GBM	"Generalized Boosted Regression Models" Formerly: "gbm" package,	distribution	The loss function	Bernoulli
		Shrinkage	Learning rate	From 0 to 1
		bag.fraction	The fraction of the training set observations randomly selected to propose the next tree	0.5 (default)
		train.fraction	Observations fraction that is used to fit the GBM	1 (default)
		n.trees	Total number of trees	From 2^5 to 2^{10}
		interaction.depth	Maximum depth of variable interactions	From 1 to 8
		n.minobsinnode	Minimum number of observations in the trees terminal nodes	20 (default)
LR	"stats" package,	link	Model link function	logit
NNET	"Feed-Forward Neural Networks and Multinomial Log-Linear" Formerly: "nnet" package,	Maxit	Maximum number of iterations	150 (default)
		MaxNWts	The maximum allowable number of weights	10,000 (default)
		Rang	Initial random weights on [-rang, rang]	0.5 (default)
		Hess	Find the Hessian of the measure of fit at the best set of weights	TRUE (default)
		Size	Number of units in the hidden layer	From 4 to 33
		Decay	Penalty term or weight decay	From 0 to 1
RF	"A Fast Implementation of Random Forests ranger" Formerly: "ranger" package,	Replace	Sample with replacement	FALSE or TRUE
		respect.unordered.factors	Handling of unordered factor covariates	TRUE (default)
		sample.fraction	The fraction of observations to sample	From 0.632 to 1
		num.trees	Number of trees	From 2^5 to 2^{10}
		mtry	Number of variables	From 2 to 8
SVM	"Misc Functions of the Department of Statistics, Probability Theory Group, TU Wien" Formerly: "E1071" package,	kernel	kernel function	radial or polynomial
		Cost	regularization cost	From 2^{-15} to 2^{15} (default)
		gamma (if kernel =: "radial")	kernel width	From 2^{-15} to 2^{15} (default)
		degree (if kernel =: "polynomial")	Polynomial degree	From 1 to 16 (default)

Only “*mtry*”, “*interaction.depth*”, “*n.trees*”, “*num.trees*” and “*size*” have the option to be set by the user according to specific instructions and guidelines. Otherwise, the remaining parameters are exactly bounded to the allowed (or default) values (or range of values) by each package. For the number of variables in each tree (“*interaction.depth*” and “*mtry*”), various heuristics suggested by packages that provide GBM and RF were used to set the optimum value (Table 4). These heuristics suggest that ranges of 1 to 8 and 2 to 8 would be accurate for “*interaction.depth*” and “*mtry*”. The additive nature of GBM allows for the one-way interaction variable in each tree (“*interaction.depth*” = 1), on the contrary, RF does not allow one-way interactions, only two-way interactions or more (“*mtry*” ≥ 2). On the other hand, the total number of trees to fit, “*n.trees*” for GBM and “*num.trees*” for RF, is set to an exponential rate using a base of 2 (2^i , $i = 5, \dots, 11$). By taking into account the instructions of the used packages and some experimental researches e.g., [45]; the total number of trees was set to an optimal value between 2^5 and 2^{10} .

Table 4. The heuristics proposed by the package instructions to set the optimum number of variables for GBM and RF. (N_i : the total number of variables (i.e., 16 in this research)).

Package	Suggested Value	
	<i>mtry</i>	<i>interaction.depth</i>
gbm	N.A	$\sqrt{N_i}$, but often the search space is set between 1 and $\sqrt{N_i}$
ranger	$\sqrt{N_i} = 4$	N.A
xgboost	6	6
H ₂ O	2 to 8	2 to 8
randomForest	$\sqrt{N_i} = 4$	N.A

The number of nodes in the hidden layer (“*size*”) for NNET was set in a range of 4–33 according to empirical suggestions proposed by different authors summarized in Table 5.

Table 5. The heuristics proposed to compute the optimum number of hidden layer nodes for NNET (modified from and Kavzoğlu [46]; N_i : number of input nodes (i.e., the total number of variables of 16 in this study); N_o : number of output nodes; N_p : Number of training samples; k : the noise factor (varies between 4 and 10) is an index number representing the percentage of false measurements in the data or degree of error).

Proposed by	Heuristic	Hidden Nodes
Hecht [47]	$2N_i + 1$	33
Ripley [48]	$(N_i + N_o)/2$	8 or 9
Paola and Schowengerdt [32]	$\frac{2 + (N_i * N_o) + \frac{1}{2} N_o (N_i^2 + N_i) - 3}{N_i + N_o}$	9
Wang [49]	$2 * N_i / 3$	11
Aldrich, et al. [50]	$\frac{N_p}{k(N_i + N_o)} (k = 10)$	7
Aldrich, Van Deventer and Reuter [50]	$\frac{N_p}{k(N_i + N_o)} (k = 7)$	10
Kaastra and Boyd [51]	$\sqrt{N_i * N_o}$	4
	$2N_i$	32

4.4. Model Training, Validation, and Comparison

Different performance metrics can be implemented for quantitative comparison; however, landslide susceptibility problems are strictly classification problems where quality and confidence in probabilities toward landslides are critical. Therefore, using a performance metric to assess prediction robustness is necessary and for this reason, the area under the receiver operating characteristic (ROC) curves (AUC) will be implemented as the only metric for the objective functions in hyperparameter tuning and one of three overall performance indicators of the landslides predictive models. In general, AUC can be interpreted as “the probability of a classifier is able to correctly anticipate the occurrence or

non-occurrence of predefined events" [16]; which is rather convenient, because maximizing the AUC value is the equivalent to maximizing the overall accuracy (Acc) of the classifier. AUC could be quantified [6] as follows: excellent (0.9–1), very good (0.8–0.9), good (0.7–0.8), average (0.6–0.7), and poor (0.5–0.6).

As in most cases, assessing the overall performance and the predictive capabilities of the tuned models based on the predictions robustness is not enough, the accuracy and the reliability of the trained models also need to be assessed. The overall accuracy (Acc), which describes the amount of correctly classified events of both landslide and non-landslide in a float (decimal) range of 1 (correctly classifying all events) to 0 (failing to classify any events), will be used to assess model accuracy. On the other hand, the Cohen kappa index (kappa) will be used to measure the landslide model reliability and can calculate the proportion of the observed agreement beyond that which is expected by chance. According to Landis and Koch [52], the strength of the agreement is given by the Kappa magnitude as 0.8–1.0 for almost perfect, 0.6–0.8 for substantial, 0.4–0.6 for moderate, 0.2–0.4 for fair, 0–0.2 for slight, and ≤ 0 for poor.

Model performance results were evaluated using non-parametric statistical procedures for statistical significance to evaluate and compare the landslide susceptibility models against each other, similarly as in [6]. The Wilcoxon signed-rank test at the 5% significance level was used for each pair of models in order to detect individual differences in model performances. Basically, the Wilcoxon signed-rank test relies on a null hypothesis that there is no difference between the performances of the landslide models. Then, p -value and z -values are calculated and used to determine the probability of rejecting or accepting the null hypothesis [6]. If both the p -value is lower than the significance threshold (p -value < 0.05) and the z -value exceeds its critical values (z -value $< (-1.96)$ or z -value $> (+1.96)$), then it is safe to assume that the null hypothesis is not valid (can be rejected). Therefore, a significant difference between the two compared models exists, otherwise, if that is, p -value ≥ 0.05 and $-1.96 \leq z$ -value $\leq +1.96$, it is safe to assume the opposite.

4.5. Landslide Susceptibility Map Generation and Assessment

Apart from performance metrics, a sufficiency analysis should be performed to assess the sufficiency and accuracy of the predictive models that produce landslide susceptibility maps. This analysis is based on the assumption that: "a model is sufficient and accurate when there is an increase in the landslide density ratio when moving from low to high susceptible classes and high susceptibility classes cover small areas extent" [16]. The sufficiency analysis can be performed by reclassifying the probability grids generated by each model for the study area using Table 6. Then, by overlying the landslide inventory, it is possible to generate summary statistics (i.e., the landslide density distribution and area extent distribution) for each class.

Table 6. The probability intervals for the landslide susceptibility classes.

Susceptibility Class	Very Low	Low	Moderate	High	Very High
Probability Range	From 0 to 0.05	From 0.05 to 0.30	From 0.30 to 0.60	From 0.60 to 0.75	From 0.75 to 1

5. Results

5.1. Analyzing and Optimizing Landslide Conditioning Factor

In a comparative study, constructing the necessary conditioning factors does not necessarily imply that it is suitable for use as an input dataset for models. In fact, it is crucial to check the integrity of the input dataset by performing some sort of analysis (i.e., correlation analysis, and multicollinearity detection) before conducting the modeling, mainly to ensure that (1) the non-independence among conditioning factors and (2) to figure out the suitability of the underlying assumption behind choosing the factors. In this research, 16 conditioning factors were considered by taking into account the aforementioned criteria, and both correlation and VIF analyses were performed.

The Pearson’s correlogram values (Figure 7) are lower than the critical threshold of 0.7, which indicates high collinearity [39]. The highest Pearson’s correlation was between TWI and the Slopes at 0.54. In fact, a high correlation is expected between the generated variables and the source variables (i.e., TWI, Slopes, and Altitude that were derived from the DEM). On the other hand, the VIF results (Figure 8), show that all factors should be used since the highest value is less than the theoretical critical value of 10 [53].

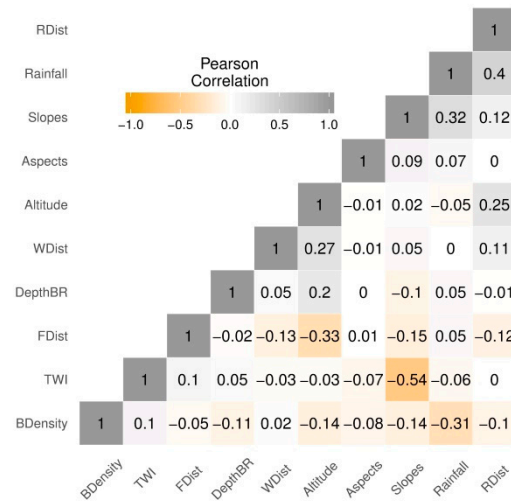


Figure 7. Correlogram based on Pearson correlation matrix of numerical conditioning factors.

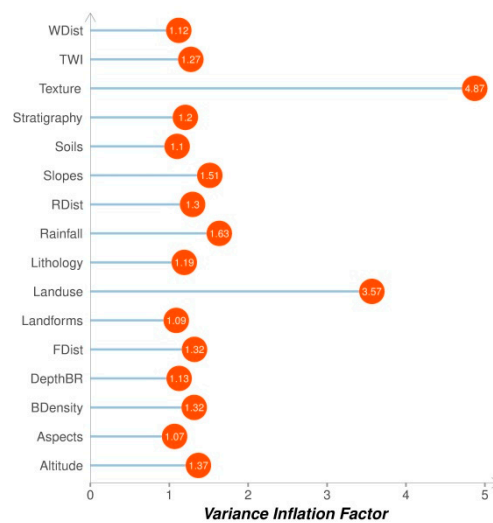


Figure 8. Variance inflation factors analysis results on landslide conditioning factors.

5.2. Model Training

During the training process, the optimal hyperparameters are carefully picked up according to Table 3 using the following procedures:

1. Set a single objective function for each learner using “smoof” [54] with AUC to maximize it as a single performance criterion.
2. Use “lhs” package [55] to set an initial design grid that covers the supplied search space of each model parameter by drawing a Latin Hypercube Sample Design (LHS) using a Column wise Pairwise (CP) algorithm to generate an optimal design with respect to the S optimality criterion [56].

3. During every single iteration, a new point is being proposed through LCB infill optimization of the estimated standard error. This error is usually obtained by a surrogate model that is either kriging-based for a purely numeric space or random forest for a mixed search space.
4. Select and return the optimum values of the desired hyperparameters based on the highest AUC (Table 7).

Table 7. The optimum parameters obtained by the tuning process.

Model	Hyperparameter	Optimal Value
GBM	Shrinkage	0.020
	n.trees	570
	interaction.depth	8
NNET	Size	29
	Decay	0.809
RF	Replace	FALSE
	sample.fraction	0.953
	num.trees	1012
	mtry	5
SVM	kernel	radial
	cost	$2^{8.382}$
	gamma	$2^{-8.398}$
	degree	N/A

5.3. Model Evaluation and Comparison

First, the models are trained using the input dataset and the hyperparameter sets (see Tables 4 and 7), then we evaluate the predictive performance capabilities and the quality of the resulting models using performance indicator metrics like AUC, Acc, and the Kappa index.

The Overall performance results show that all the models have “a substantial agreement” between the observed and the predicted landslides expressed in term of a kappa index ranging between 0.5605 and 0.6405 (Figure 9 and Table 8). The AUC and Acc values range from 0.8575 to 0.8967, and 0.7803 to 0.8203, respectively, indicate that all the models have “very good” predictive capabilities. In particular, the ensembles models that benefit from a divide-and-conquer approach such as RF and GBM yielded significantly better results than traditional methods like NNET, SVM, and LR. In fact, GBM was the highest-ranked model in terms of the performance of the AUC, Acc, and Kappa index with values of 0.8967, 0.8203, and 0.6405, respectively (Table 8). RF held the second highest ranked model with performances similar to GBM with values of 0.8957, 0.8178, and 0.6356 for AUC, Acc, and kappa, respectively. NNET, on the other hand, was able to achieve the highest performance after the ensemble tree models, followed up by SVM. In contrast, the LR performance was consistently lower than the rest of the models in every metric, with values of 0.8575, 0.7803, and 0.5605 for AUC, Acc, and kappa, respectively.

Table 8. The overall performances of the trained landslide models.

Metrics	Model				
	GBM	LR	NNET	RF	SVM
Acc	0.820	0.780	0.809	0.817	0.802
Kappa Index	0.640	0.560	0.619	0.635	0.605

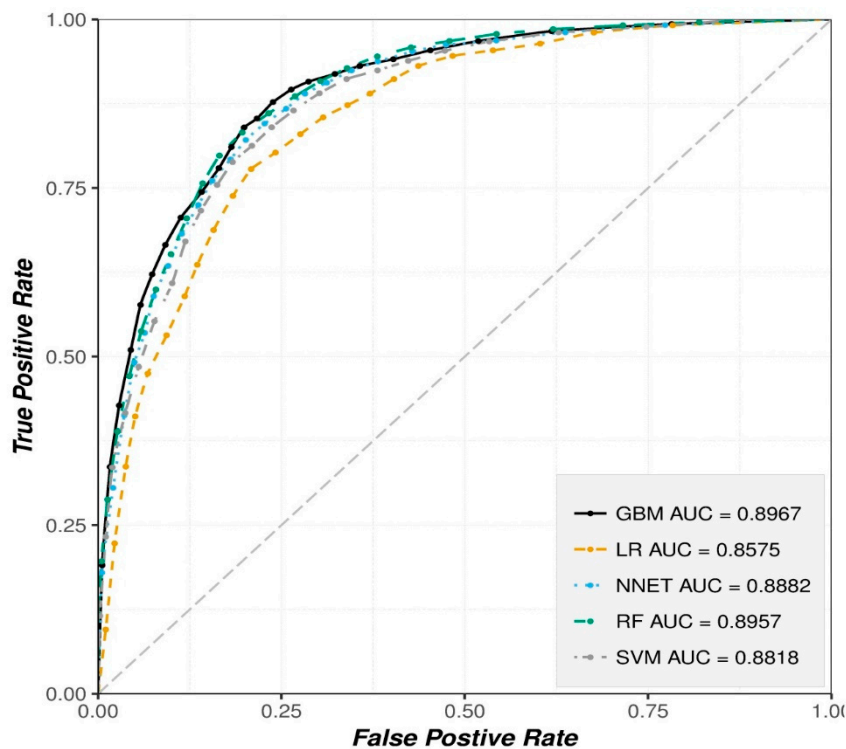


Figure 9. The stacked receiver operating characteristic (ROC) curves of the implemented models.

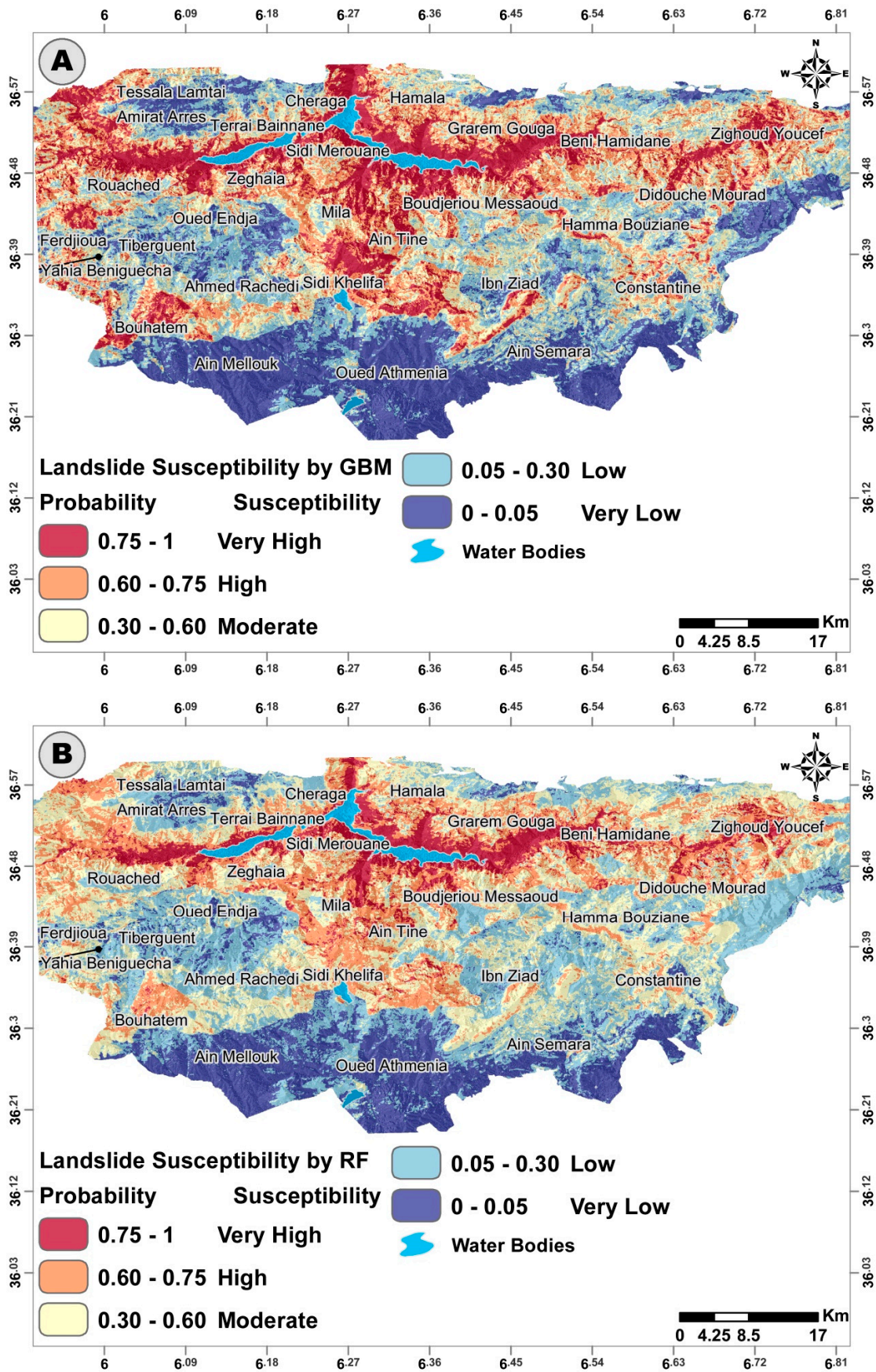
Finally, in order to determine if there are statistically significant differences between the five landslides susceptibility models, a systematic pairwise comparison using the Wilcoxon signed-rank test at the 5% significance level was conducted (Table 9). The results show that there is a systematic difference in the performance results between each pair of models except for the GBM and RF pair, where the difference in performance was found to be statistically insignificant (that is, p -value ≥ 0.05 and $-1.96 \leq z$ -value $\leq +1.96$, so, the null hypothesis was accepted). Overall, it could be concluded that the RF model and the GBM model are the best for the data at hand in this study.

Table 9. The pairwise comparison of the five landslide susceptibility models using the Wilcoxon signed-rank test.

No.	Pairwise Comparison	z Value	p Value	Significance
1	GBM vs. RF	−0.579	0.562	No
2	GBM vs. LR	6.111	0.000	Yes
3	GBM vs. NNET	3.606	0.001	Yes
4	GBM vs. SVM	5.266	0.000	Yes
5	RF vs. LR	6.149	0.000	Yes
6	RF vs. NNET	2.905	0.004	Yes
7	RF vs. SVM	4.025	0.000	Yes
8	SVM vs. LR	5.589	0.000	Yes
9	SVM vs. NNET	−3.223	0.001	Yes
10	NNET vs. LR	5.995	0.000	Yes

5.4. Generating Landslide Susceptibility Map

Once the final models were evaluated and validated, the tuned models were used to successfully predict and generate landslide occurrence in the study area in the form of probability grids, then they were reclassified into five susceptibility classes (Table 6). The implemented models successfully generated susceptibility maps with a fine and smooth prediction surface (Figure 10).



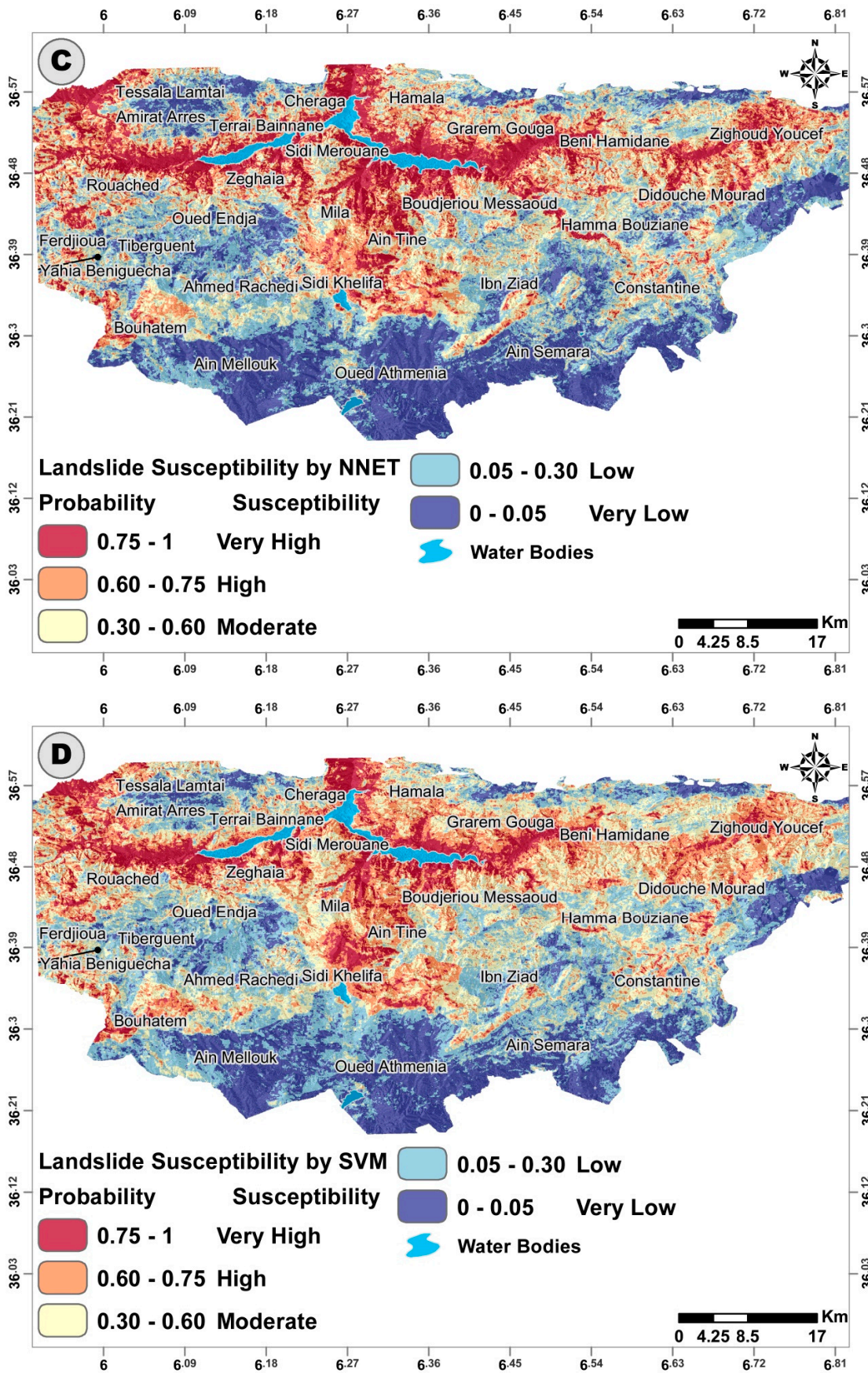


Figure 10. Cont.

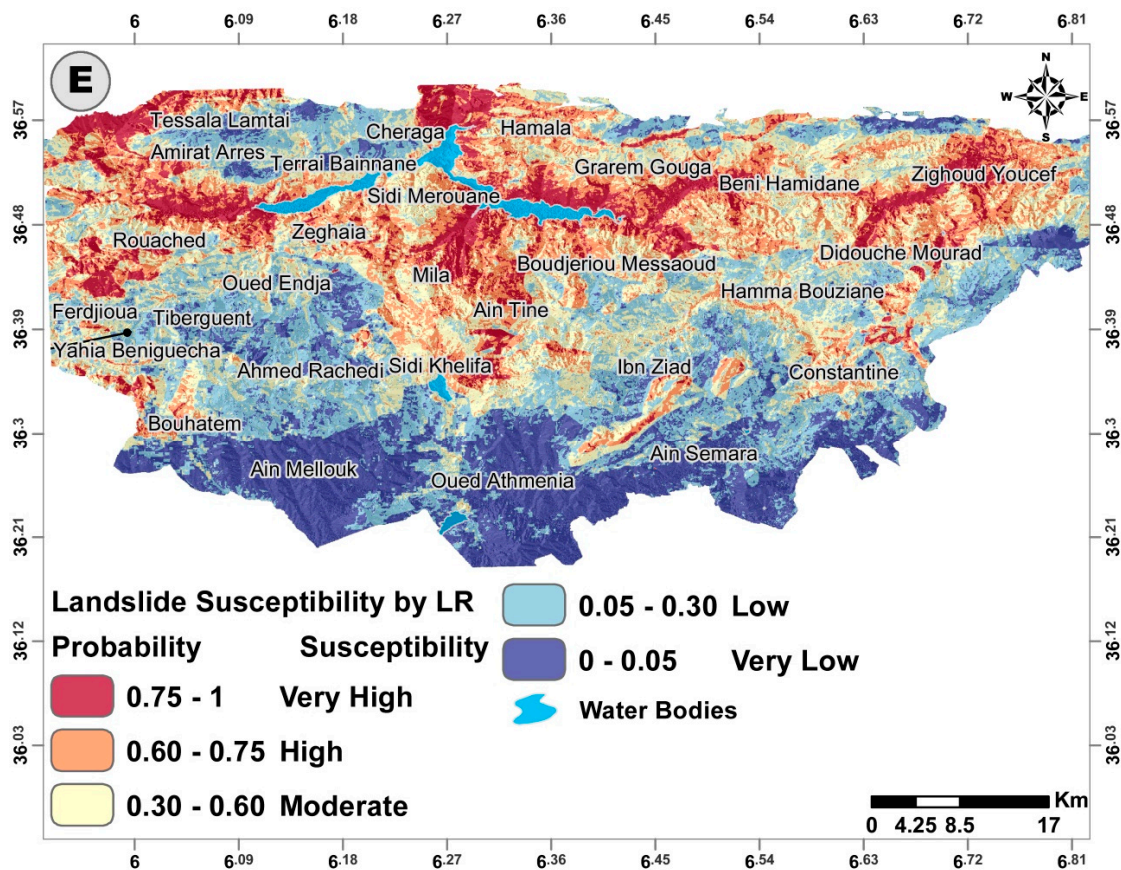


Figure 10. The generated landslide susceptibility maps using: (A) GBM, (B) RF, (C) NNET (D) SVM; and (E) LR.

In the case of a landslide susceptibility assessment, the model evaluation based on the performance metrics is not enough. Models with close or even similar performance results (for example, GBM and RF have no statistical significance in the performance difference in this case study) and they do not necessarily generate similar predictive output surface. The spatial predictive output surface is critical for assessing the quality of landslide susceptibility models. Overall, by performing a sufficiency analysis on the predictive output surface in the form of summary statistics (that is, landslide density distribution and the area extent covered by each susceptibility class), it is possible to gain an insight into the model’s quality by (1) the spatial predictive output surface details and (2) the results of the landslide distribution analysis.

Essentially, by overlapping the landslide inventory and the reclassified susceptibility maps (Figure 10), a sufficiency analysis summary statistic was obtained in the form of a landslide density distribution (Figure 11A) and the total area extent covered by each susceptibility class (Figure 11B). The results are satisfying because they fulfill two spatial conditions: (1) the landslide pixels should be located at the very high and high susceptible classes and (2) the extent of the areas covered by the very high and high susceptible classes should be as small as possible. All models show an increase in the landslide density ratio when moving from low to high susceptible classes, with GBM scoring the best results of approximately 75.61% and 14.52% for landslide density occurrences and the area extent covered by the highest susceptibility class (that is, “very high”). RF scored 74.39% and 6.99% followed by NNET with 68.34% and 14.28%, SVM with 68.17% and 9.90%, and LR with 56.23% and 9.29%.

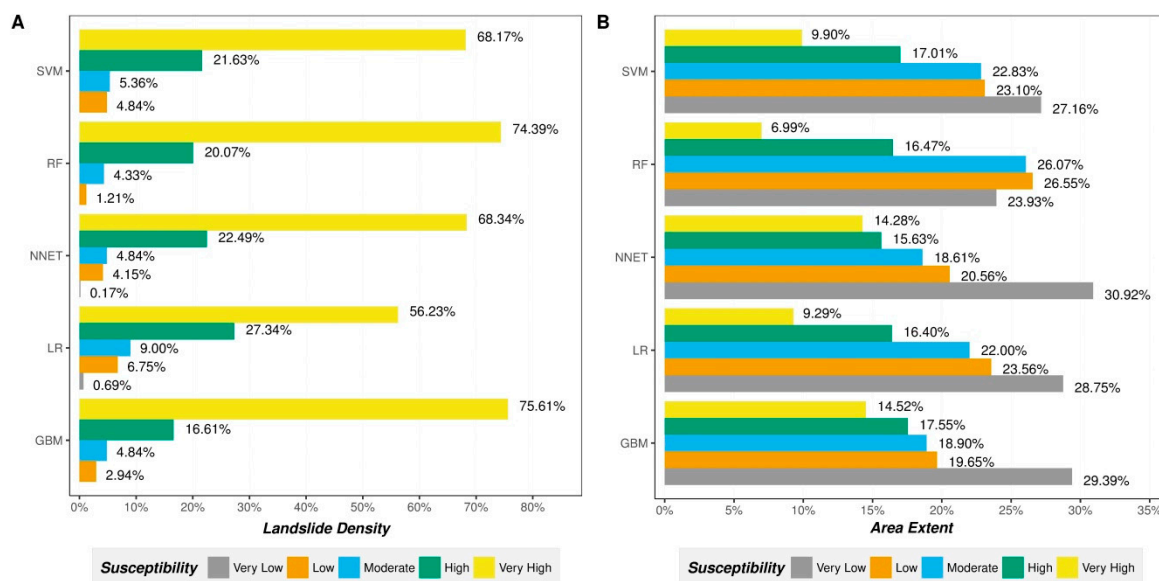


Figure 11. The sufficiency analysis of the susceptibility maps: (A) Landslide density distribution by susceptibility zones; (B) the total area extent covered by susceptibility zones.

A positive indicator of the classification capability of the generated models is that they do not show any landslide events in the “Very Low” susceptibility class (that is, if the landslide density is null, then the class is absent) or they only show a very small percentage (<0.70% of the total landslide events) (Figure 11A). In general, the “Very Low” and “Low” susceptibility classes are grouped pixels with the low probabilities toward landslides, which mean that those pixels have higher confidence probability toward stability. Therefore, having a lower percentage (or even better, an absence) of the lower susceptibility classes indicates a higher confidence in the misclassification error (equal to $1 - Acc$) achieved by those models. Further, they indicate that the misclassification error achieved was near the classification threshold (for binary equal-proportions, the classifications threshold is 0.5) and not at the extremes.

6. Discussions

The most effective way to reduce casualties and economic losses resulting from landslides are landslide risk planning and management; therefore, high-quality landslide susceptibility maps are an important tool [57]. However, it is still a challenge to produce high accuracy landslide susceptibility maps at a regional scale due to the complex nature of landslides and it is widely recognized that the prediction quality of landslides is dependent on the algorithm used. Thus, although various methodologies for producing landslide susceptibility maps have been developed, the prediction accuracy of these methods is still debated [49]. Therefore, in the present study, five classifications algorithms (GBM, LR, NNET, RF, and SVM) were investigated and compared for landslide susceptibility mapping at Mila Basin.

The results obtained in this study (see Table 8 and Figure 9) show that all the implemented models achieved high performance ($AUC > 0.88$, $Acc > 80\%$ and $kappa > 0.60$). However, two ensemble trees models (GBM and RF) yielded the highest prediction results compared to the others. This better performance is confirmed to be statistically significant with the used Wilcoxon signed-rank test. This finding is in agreement with the results from recent studies i.e., in ([58–61]) that reported that the ensemble models outperform single machine learning models. In contrast to GBM and RF, LR consistently yields the lowest results compared to the other implemented models. This finding is in line with the literature where LR achieves the worst, if not the poorest, performance of all models [6,9,11,16,62].

The better-fit and higher performance of GBM and RF compared to LR, NNET, and SVM in this research is due to the divide-and-conquer approach that the assembling technique implements in both models (i.e., benefiting from aggregating weak learners to solve the issue). In fact, the main causes of error in the landslide modeling at the basin scale in this study is due to noise and the uncertainty that existed in the landslide conditioning factor maps (which were collected from various sources and scales). It is still difficult to eliminate the noise and uncertainty, though several fuzzy modeling approaches have been proposed. However, ensemble learning, RF, and GBM, which use the random sampling with replacement strategy, could minimize these due to their diversity and stability [63], which are two key issues of ensemble learning. Thus, both RF and GBM are capable models that work well over noise and uncertainty environments [64] such as landslide modeling; therefore, they are robust and better than the other models in this study.

Generally, GBM models offer similar or even better performance results than RF, but the large number of sensitive parameters and the tendency to easily over-fit makes it difficult to implement it right out the box compared to RF, which is easier to implement and less prone to both over-fitting and outliers. Additionally, some studies [65] have found that GBM performs exceptionally well when the dimensionality is low (≈ 4000 predictors). Above that, RF has the best overall performance. Notably, the results obtained by SVM for typical binary landslide susceptibility problems are very satisfying. Even if it is lower than GBM, RF, and NNET, it is still relevant compared to the results produced by similar studies [1,6,8–10]. NNET, on the other hand, unsurprisingly outperforms SVM and LR, but fails to capture the underlying model of the input data like RF and GBM, simply because neural networks need a large number of observations. However, in the case of landslides, the observation events are scarce and very hard to obtain. On top of samples being scarce, the most recent landslide susceptibility studies [7,9,16] do not benefit from the full potential of NNET by implementing NNET models with vanilla “Backpropagation” or one of its variances for the weight adjustments. In fact, Back-propagation based NNET are extremely slow to converge, which leads to a long execution time and a heavy computational load, not to mention both a large number of parameters to tune in and the special input data preparation required. Unlike Back-propagation NNETs, the implemented feed forward BFGS NNET are faster to converge with fewer hyperparameters to tune in and provided arguably better results than similar studies that implemented NNET [7,9,16].

In the end, it is widely accepted that no single or particular model can be depicted as the most suitable for all case scenarios. For example, the LR model is simple, fast, easy to implement, and is only able to capture the linear relationship between the conditioning factors and the landslide susceptibility. The merit of LR is that it does not compulsorily require a normal distribution data. Additionally, both continuous and discrete data types can be used as an input for the LR model. However, landslides are complex phenomena with non-linear mechanisms. SVMs are useful non-linear classifiers whose goal is not only to correctly classify landslide instances, but also to keep the distance between instances and keep the separation of the hyperplane at a maximum. This makes SVM models appealing for susceptibility evaluation considering the number of hyperparameters to tune in. However, if those hyperparameters are inappropriately set, SVM will often lead to unsatisfactory results. NNET models are very effective for simulating non-linear complex phenomena with multiple conditioning factors (preferably continuous input dataset). However, being a black box model and the large number of samples required to obtain a reliable model are the only downsides to this kind of model. Ensemble tree models (GBM and RF) offer excellent performance with decent interpretability and a moderate number of hyperparameters to tune in but require a considerable time budget (they require a lot of time to converge, especially if used on large-scale analyses). Though some studies (such as in [66]) highly recommend RF and GBM due to the outstanding performance, they suggest that a rather fast and simple model, such as LR would be much better than advanced machine learning models.

All scripts used in this experiment are available in a reproducible repository on Github (https://github.com/aminevsaziz/lsm_in_Mila_basin).

7. Conclusions

This research paper provided a framework for comparing and assessing five machine-learning methods (GBM, LR, NNET, RF, and SVM) for the landslide susceptibility assessment in the Mila Basin. The achieved results demonstrate that there is a significant difference between the implemented models. Even if the obtained results are underlined with a clear objective of comparing and assessing those models, finding the most suitable model for the case study was very challenging as it does not depend solely on the performance results, but also on the high level of uncertainty behind landslide modeling and the limitation and caveats that come with each model.

The two ensemble tree models (RF and GBM) were proven the most suitable models for this case study when comparing them to the remaining models (NNET, SVM, and LR), as they significantly outperformed the rest of the models based on the excellent performance results achieved. Despite that, the remaining three models are considered viable options, as they are adequately capable of satisfactory performance compared to similar studies. Summing up, the obtained landslide susceptibility maps by the implemented models can be used as a preliminary planning framework for planners in the study area or as a technical framework for countermeasures and regulatory policies by decision makers to minimize the damages introduced by either existing or future landslides by the Mila and Constantine municipalities.

Overall, the results of this study have demonstrated the effectiveness of all Five ML technique classifiers, especially ensemble tree models such as the GBM and RF algorithms for the assessment of landslide susceptibility. In terms of future work, we will consider the following issues: (1) exploring other machine learning algorithms; (2) including more landslide observation cases if possible; (3) introduce more richness to the input data pool, such as deformation formations based on InSAR space born imagery.

Author Contributions: A.M. conceived the study, designed, and investigated the analysis, wrote and implemented the analysis computer scripts code and supporting algorithms, validated and ensure the reproducibility of the results and participated in curating the data and writing the manuscript (drafting, reviewing and editing). B.A. participated in designing the analysis, curating the data, and writing the manuscript (drafting). D.T.B. participated in the formal analysis, checked the results, and heavily contributed to writing the manuscript (reviewing and editing) along with the visualization/data presentation.

Funding: This research received no external funding.

Acknowledgments: The authors would like to express their deepest gratitude to the Editor and the reviewers for their helpful and constructive comments on the manuscript and everyone provided assistance in the process of making and enhancing this research especially the DTP (Direction des Travaux Publics) and both the Mila and Constantine municipalities for providing the necessary data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pham, B.T.; Pradhan, B.; Tien Bui, D.; Prakash, I.; Dholakia, M.B. A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of uttarakhand area (India). *Environ. Model. Softw.* **2016**, *84*, 240–250. [[CrossRef](#)]
2. Tien Bui, D.; Pradhan, B.; Lofman, O.; Revhaug, I.; Dick, O.B. Landslide susceptibility assessment in the hoa binh province of vietnam: A comparison of the levenberg–marquardt and bayesian regularized neural networks. *Geomorphology* **2012**, *171–172*, 12–29. [[CrossRef](#)]
3. Chen, W.; Xie, X.; Wang, J.; Pradhan, B.; Hong, H.; Bui, D.T.; Duan, Z.; Ma, J. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena* **2017**, *151*, 147–160. [[CrossRef](#)]
4. Guzzetti, F.; Carrara, A.; Cardinali, M.; Reichenbach, P. Landslide hazard evaluation: A review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology* **1999**, *31*, 181–216. [[CrossRef](#)]
5. Lary, D.J.; Alavi, A.H.; Gandomi, A.H.; Walker, A.L. Machine learning in geosciences and remote sensing. *Geosci. Front.* **2016**, *7*, 3–10. [[CrossRef](#)]

6. Tien Bui, D.; Tuan, T.A.; Klempe, H.; Pradhan, B.; 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* **2016**, *13*, 361–378. [[CrossRef](#)]
7. Dou, J.; Yamagishi, H.; Pourghasemi, H.R.; Yunus, A.P.; Song, X.; Xu, Y.; Zhu, Z. An integrated artificial neural network model for the landslide susceptibility assessment of osado island, Japan. *Nat. Hazards* **2015**, *78*, 1749–1776. [[CrossRef](#)]
8. Goetz, J.N.; Brenning, A.; Petschko, H.; Leopold, P. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput. Geosci.* **2015**, *81*, 1–11. [[CrossRef](#)]
9. Tien Bui, D.; Tuan, T.A.; Hoang, N.-D.; Thanh, N.Q.; Nguyen, D.B.; Van Liem, N.; Pradhan, B. Spatial prediction of rainfall-induced landslides for the lao cai area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. *Landslides* **2017**, *14*, 447–458. [[CrossRef](#)]
10. Dou, J.; Paudel, U.; Oguchi, T.; Uchiyama, S.; Hayakawa, Y. Shallow and deep-seated landslide differentiation using support vector machines: A case study of the chuetsu area, Japan. *Terr. Atmos. Ocean. Sci.* **2015**, *26*, 13. [[CrossRef](#)]
11. Youssef, A.M.; Pourghasemi, H.R.; Pourtaghi, Z.S.; Al-Katheeri, M.M. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at wadi tayyah basin, asir region, Saudi Arabia. *Landslides* **2015**, *13*, 839–856. [[CrossRef](#)]
12. Lombardo, L.; Cama, M.; Conoscenti, C.; Märker, M.; Rotigliano, E. Binary logistic regression versus stochastic gradient boosted decision trees in assessing landslide susceptibility for multiple-occurring landslide events: Application to the 2009 storm event in messina (Sicily, Southern Italy). *Nat. Hazards* **2015**, *79*, 1621–1648. [[CrossRef](#)]
13. Carrara, A.; Pike, R.J. Gis technology and models for assessing landslide hazard and risk. *Geomorphology* **2008**, *94*, 257–260. [[CrossRef](#)]
14. Wolpert, D.H. The lack of a priori distinctions between learning algorithms. *Neural Comput.* **1996**, *8*, 1341–1390. [[CrossRef](#)]
15. Elith, J.; Leathwick, J.R. Species distribution models: Ecological explanation and prediction across space and time. *Ann. Rev. Ecol. Evol. Syst.* **2009**, *40*, 677–697. [[CrossRef](#)]
16. Tsangaratos, P.; Ilija, I. Comparison of a logistic regression and naïve bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size. *Catena* **2016**, *145*, 164–179. [[CrossRef](#)]
17. Mebarki, A. *Le Bassin du Kébir-Rhumel: Hydrologie de Surface et Aménagement des Ressources en eau: Travaux du Laboratoire de Géographie Physique*; Éditeur Inconnu: Nancy, France, 1982; p. 304.
18. Rullan-Perchirin, F. *Recherches sur L'érosion dans Quelques Bassins du Constantinois (Algérie)*. Ph.D. Thesis, Université Panthéon-Sorbonne, Paris, France, 1985.
19. Chettah, W. *Investigation des Propriétés Minéralogiques et Géomécaniques des Terrains en Mouvement Dans la Ville de Mila «Nord-Est d'Algérie»*. Ph.D. Thesis, Université Hadj Lakhdar, Batna, Algeria, 2009.
20. Guzzetti, F.; Mondini, A.C.; Cardinali, M.; Fiorucci, F.; Santangelo, M.; Chang, K.-T. Landslide inventory maps: New tools for an old problem. *Earth Sci. Rev.* **2012**, *112*, 42–66. [[CrossRef](#)]
21. Yalcin, A. Gis-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in ardesen (Turkey): Comparisons of results and confirmations. *Catena* **2008**, *72*, 1–12. [[CrossRef](#)]
22. Zouaoui, S. *Etude Géologique et Géotechnique des Glissements de Terrains Dans le Bassin Néogène de Mila: Glissement de Sibari*. Ph.D. Thesis, Université Hadj Lakhdar, Batna, Algeria, 2008.
23. Ayalew, L.; Yamagishi, H.; Marui, H.; Kanno, T. Landslides in sado island of Japan: Part II. Gis-based susceptibility mapping with comparisons of results from two methods and verifications. *Eng. Geol.* **2005**, *81*, 432–445. [[CrossRef](#)]
24. Van Westen, C.J.; Castellanos, E.; Kuriakose, S.L. Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. *Eng. Geol.* **2008**, *102*, 112–131. [[CrossRef](#)]
25. Conforti, M.; Pascale, S.; Robustelli, G.; Sdao, F. Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the turbolo river catchment (Northern Calabria, Italy). *CATENA* **2014**, *113*, 236–250. [[CrossRef](#)]

26. Iwahashi, J.; Pike, R.J. Automated classifications of topography from dems by an unsupervised nested-means algorithm and a three-part geometric signature. *Geomorphology* **2007**, *86*, 409–440. [[CrossRef](#)]
27. Beven, K.J.; Kirkby, M.J. A physically based, variable contributing area model of basin hydrology/un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrol. Sci. Bull.* **1979**, *24*, 43–69. [[CrossRef](#)]
28. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
29. Natekin, A.; Knoll, A. Gradient boosting machines, a tutorial. *Front. Neurobot.* **2013**, *7*. [[CrossRef](#)] [[PubMed](#)]
30. Mason, L.; Baxter, J.; Bartlett, P.; Frean, M. *Boosting Algorithms as Gradient Descent in Function Space*; Australian National University: Canberra, Australia, 1999.
31. Nelder, J.A.; Baker, R.J. *Generalized Linear Models*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2004.
32. Paola, J.D.; Schowengerdt, R.A. A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery. *Int. J. Remote Sens.* **1995**, *16*, 3033–3058. [[CrossRef](#)]
33. Ripley, B.D. *Pattern Recognition and Neural Networks*; Cambridge University Press: Cambridge, UK, 2007; p. 403.
34. Cristianini, N.; Schölkopf, B. Support Vector Machines and kernel Methods: The New Generation of Learning Machines. *Artif. Intell. Mag.* **2002**, *23*, 3. [[CrossRef](#)]
35. Yao, X.; Dai, F.C. *Support Vector Machine Modeling of Landslide Susceptibility Using a Gis: A Case Study*; The Geological Society: London, UK, 2006.
36. Guo, Q.; Kelly, M.; Graham, C.H. Support vector machines for predicting distribution of sudden oak death in california. *Ecol. Model.* **2005**, *182*, 75–90. [[CrossRef](#)]
37. Molinaro, A.M.; Simon, R.; Pfeiffer, R.M. Prediction error estimation: A comparison of resampling methods. *Bioinformatics* **2005**, *21*, 3301–3307. [[CrossRef](#)] [[PubMed](#)]
38. Dormann, C.F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; Marquéz, J.R.G.; Gruber, B.; Lafourcade, B.; Leitão, P.J.; et al. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **2012**, *36*, 27–46. [[CrossRef](#)]
39. Booth, G.D.; Niccolucci, M.J.; Schuster, E.G. *Identifying Proxy Sets in Multiple Linear Regression: An Aid to Better Coefficient Interpretation*; Research paper INT (USA); U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station: Ogden, UT, USA, 1994.
40. Bergstra, J.; Yamins, D.; Cox, D.D. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In *JMLR Workshop and Conference Proceedings, Proceedings of the 30th International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013*; Sanjoy, D., David, M., Eds.; PMLR: Atlanta, GA, USA, 2013; Volume 28, pp. 115–123.
41. Thornton, C.; Hutter, F.; Hoos, H.H.; Leyton-Brown, K. *Auto-Weka: Combined Selection and Hyperparameter Optimization of Classification Algorithms*; ACM Press: New York, NY, USA, 2013.
42. López-Ibañez, M.; Dubois-Lacoste, J.; Pérez Cáceres, L.; Birattari, M.; Stützle, T. The irace package: Iterated racing for automatic algorithm configuration. *Oper. Res. Perspect.* **2016**, *3*, 43–58. [[CrossRef](#)]
43. Bischl, B.; Richter, J.; Bossek, J.; Horn, D.; Thomas, J.; Lang, M. Mlrmb: A modular framework for model-based optimization of expensive black-box functions. *arXiv* **2017**.
44. Bischl, B.; Lang, M.; Kotthoff, L.; Schiffner, J.; Richter, J.; Studerus, E.; Casalicchio, G.; Jones, Z.M. Mlr: Machine learning in R. *J. Mach. Learn. Res.* **2016**, *17*, 1–5.
45. Kertész, C. Rigidity-based surface recognition for a domestic legged robot. *IEEE Robot. Autom. Lett.* **2016**, *1*, 309–315. [[CrossRef](#)]
46. Kavzoğlu, T. An Investigation of the Design and Use of Feed-Forward Artificial Neural Networks in the Classification of Remotely Sensed Images. Ph.D. Thesis, University of Nottingham, Nottingham, UK, 2001.
47. Hecht, N. IEEE First Annual International Conference on Neural Networks San Diego, California June 21–24, 1987. *IEEE Expert* **1987**, *2*, 14.
48. Ripley, B.D. *Statistical Aspects of Neural Networks*; Springer: Boston, MA, USA, 1993; pp. 40–123.
49. Wang, C. *A Theory of Generalization in Learning Machines with Neural Network Applications*; University of Pennsylvania: Philadelphia, PA, USA, 1994.
50. Aldrich, C.; Van Deventer, J.S.J.; Reuter, M.A. The application of neural nets in the metallurgical industry. *Miner. Eng.* **1994**, *7*, 793–809. [[CrossRef](#)]
51. Kaastra, I.; Boyd, M. Designing a neural network for forecasting financial and economic time series. *Neurocomputing* **1996**, *10*, 215–236. [[CrossRef](#)]

52. Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. *Biometrics* **1977**, *33*, 159–174. [[CrossRef](#)] [[PubMed](#)]
53. Kennedy, P. *A Guide to Econometrics*, 5th ed.; The MIT Press: Cambridge, MA, USA, 2003; p. 500.
54. Bossek, J. Smoof: Single- and multi-objective optimization test functions. *R J.* **2017**, *9*.
55. Carnell, R. *Lhs: Latin Hypercube Samples*; R Package Version 0.14 ed.; R Project: Vienna, Austria, 2016.
56. Stocki, R. A method to improve design reliability using optimal latin hypercube sampling. *Comput. Assist. Mech. Eng. Sci.* **2005**, *12*, 393.
57. Klimeš, J.; Stemberk, J.; Blahut, J.; Krejčí, V.; Krejčí, O.; Hartvich, F.; Kycl, P. Challenges for landslide hazard and risk management in ‘low-risk’ regions, Czech Republic—Landslide occurrences and related costs (ipl project No. 197). *Landslides* **2017**, *14*, 771–780. [[CrossRef](#)]
58. Pham, B.T.; Tien Bui, D.; Prakash, I. Bagging based support vector machines for spatial prediction of landslides. *Environ. Earth Sci.* **2018**, *77*, 146. [[CrossRef](#)]
59. Dang, V.-H.; Dieu, T.B.; Tran, X.-L.; Hoang, N.-D. Enhancing the accuracy of rainfall-induced landslide prediction along mountain roads with a GIS-based random forest classifier. *Bull. Eng. Geol. Environ.* **2018**, *76*, 1–15. [[CrossRef](#)]
60. Chen, W.; Xie, X.; Peng, J.; Shahabi, H.; Hong, H.; Bui, D.T.; Duan, Z.; Li, S.; Zhu, A.X. Gis-based landslide susceptibility evaluation using a novel hybrid integration approach of bivariate statistical based random forest method. *CATENA* **2018**, *164*, 135–149. [[CrossRef](#)]
61. Hong, H.; Liu, J.; Bui, D.T.; Pradhan, B.; Acharya, T.D.; Pham, B.T.; Zhu, A.X.; Chen, W.; Ahmad, B.B. Landslide susceptibility mapping using j48 decision tree with adaboost, bagging and rotation forest ensembles in the guangchang area (China). *CATENA* **2018**, *163*, 399–413. [[CrossRef](#)]
62. Conoscenti, C.; Ciaccio, M.; Caraballo-Arias, N.A.; Gómez-Gutiérrez, Á.; Rotigliano, E.; Agnesi, V. Assessment of susceptibility to earth-flow landslide using logistic regression and multivariate adaptive regression splines: A case of the belice river basin (Western Sicily, Italy). *Geomorphology* **2015**, *242*, 49–64. [[CrossRef](#)]
63. Dai, Q.; Ye, R.; Liu, Z. Considering diversity and accuracy simultaneously for ensemble pruning. *Appl. Soft Comput.* **2017**, *58*, 75–91. [[CrossRef](#)]
64. Brillante, L.; Gaiotti, F.; Lovat, L.; Vincenzi, S.; Giacosa, S.; Torchio, F.; Segade, S.R.; Rolle, L.; Tomasi, D. Investigating the use of gradient boosting machine, random forest and their ensemble to predict skin flavonoid content from berry physical-mechanical characteristics in wine grapes. *Comput. Electron. Agric.* **2015**, *117*, 186–193. [[CrossRef](#)]
65. Caruana, R.; Karampatziakis, N.; Yessenalina, A. An empirical evaluation of supervised learning in high dimensions. In Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland, 5–9 July 2008; pp. 96–103.
66. Vorpahl, P.; Elsenbeer, H.; Märker, M.; Schröder, B. How can statistical models help to determine driving factors of landslides? *Ecol. Model.* **2012**, *239*, 27–39. [[CrossRef](#)]

