

Article

Landslide Susceptibility Mapping Using Multi-Criteria Decision-Making (MCDM), Statistical, and Machine Learning Models in the Aube Department, France

Abdessamad Jari ^{1,2,*}, Achraf Khaddari ³, Soufiane Hajaj ^{2,*} , El Mostafa Bachaoui ², Sabine Mohammedi ¹, Amine Jellouli ² , Hassan Mosaid ² , Abderrazak El Harti ² and Ahmed Barakat ² 

- ¹ Applied Sciences Ile-de-France Institute, Gustave Eiffel University, 77420 Champs-sur-Marne, France
² Geomatics, Georesources and Environment Laboratory, Faculty of Sciences and Techniques, Sultan Moulay Slimane University, Beni-Mellal 23000, Morocco
³ Laboratory of Geosciences, Faculty of Sciences, Ibn Tofail University, Morocco University Campus, Kenitra PB 133, Morocco; achraf.khaddari@uit.ac.ma
* Correspondence: jari.geominesfpt@gmail.com (A.J.); soufianehajaj13@gmail.com (S.H.)

Abstract: Landslides are among the most relevant and potentially damaging natural risks, causing material and human losses. The department of Aube in France is well known for several major landslide occurrences. This study focuses on the assessment of Landslide Susceptibility (LS) using the Frequency Ratio (FR) as a statistical method, the Analytic Hierarchy Process (AHP) as a Multi-Criteria Decision-Making (MCDM) method, and Random Forest (RF) and k-Nearest Neighbor (kNN) as machine learning methods in the Aube department, northeast of France. Subsequently, the thematic layers of eight landslide causative factors, including distance to hydrography, density of quarries, elevation, slope, lithology, distance to roads, distance to faults, and rainfall, were generated in the geographic information system (GIS) environment. The thematic layers were integrated and processed to map landslide susceptibility in the study area. On the other hand, an inventory of landslides was carried out based on the database created by the French Geological Survey (BRGM), where 157 landslide occurrences were selected, and then RF and kNN models were trained to generate landslide maps (LSMs) of the study area. The generated maps were assessed by using the Area Under the Receiver Operating Characteristic Curve (ROC AUC). Subsequently, the accuracy assessment of the FR model revealed more accurate results (AUC = 66.0%) than AHP, outperforming the latter by 6%, while machine learning models results showed that RF gave better results than kNN (<7.3%) with AUC = 95%. Following the analysis of LS mapping results, lithology, distance to the hydrographic network, distance to roads, and elevation were the four main factors controlling landslide susceptibility in the study area. Future mitigation and protection activities within the Aube department can benefit from the present study mapping results, implicating an optimized land management for decision-makers.



Citation: Jari, A.; Khaddari, A.; Hajaj, S.; Bachaoui, E.M.; Mohammedi, S.; Jellouli, A.; Mosaid, H.; El Harti, A.; Barakat, A. Landslide Susceptibility Mapping Using Multi-Criteria Decision-Making (MCDM), Statistical, and Machine Learning Models in the Aube Department, France. *Earth* **2023**, *4*, 698–713. <https://doi.org/10.3390/earth4030037>

Academic Editor: Christian Conoscenti

Received: 12 July 2023

Revised: 18 August 2023

Accepted: 25 August 2023

Published: 9 September 2023

Keywords: landslide susceptibility; analytic hierarchy process; frequency ratio; random forest; k-nearest neighbor; Aube department



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Landslides are well recognized among the prevalent geological hazards, impacting ecological quality. For example, from 1997 to 2017, landslides caused 378 disasters [1]. In terms of economics, the losses that have been due to landslides were estimated at many billions of dollars [1]. The synergy of human activities and various physical factors favored landslide occurrences in many vulnerable areas [2]. By harnessing the power of statistical, MCDM, and machine learning techniques, we seek to overcome the limited information and provide a comprehensive understanding of landslide susceptibility in the Aube department area. In France, the density of landslides recorded between 1900 and 2020 has varied from

0.2 per 10 km² to over 1 per 10 km². The most notable landslide occurrences include a landslide in March 2014 in Bar-Sur-Aube, specifically below the Ste-Germaine farm, which resulted in the fall of rocks and threatened two buildings and a communal road. Another significant landslide occurred in 2003 upstream of the municipal road connecting Fontaine to Bar-Sur-Aube.

In general, many approaches have been developed for evaluating landslide susceptibility mapping. The use of data-driven methods such as machine and deep learning, statistical methods, and knowledge-driven methods in geographic information system (GIS) environments have become popular in assessing landslide susceptibility. In knowledge-driven methods, the dependence on the investigator's prior knowledge has played a significant role in the results of landslide analysis [3]. The qualitative or heuristic analytic hierarchy process (AHP) method has become one of the best MCDM techniques adopted in landslide susceptibility mapping [4]. AHP facilitates decision-making processes for individuals in positions of authority by simplifying complex decisions into a sequence of comparative evaluations and consolidating the outcomes. AHP methodically breaks down complex decision problems into distinct hierarchical levels, enabling the quantification of subjective viewpoints and their conversion into a cohesive decision model [5]. In addition, numerous statistical models have been used for landslide susceptibility mapping, including logistic regression (LR) [6], weight of evidence (WOI) [7], the frequency ratio (FR) [8], the Modified information value (MIV) [9,10], and the index of entropy (IOE) [11].

Furthermore, data-driven techniques utilize artificial intelligence (AI) algorithms to analyze and predict information by learning from training datasets. These algorithms employ iterative modeling processes to extract valuable insights and patterns from data. Among the adopted machine learning algorithms in landslide susceptibility mapping, we can mention random forest (RF) [12], decision tree (DT) [13], k-nearest neighbor (kNN) [12], support vector machine (SVM) [14], Bayesian network (BN) [15], Multilayer Perceptron (MLP) [16], and deep learning models (DL) [17]. Due to its many benefits, random forests (RFs) have recently received a lot of interest in the machine learning community. RF regularly produces trustworthy findings with exceptional precision, has a quick processing speed, simplifies the model implementation process by requiring less parameter adjustment, and provides exceptional effectiveness in processing high-dimensional data [18]. In addition, the authors often adopt integrated approaches to solve LSM problems within complex areas. For example, knowledge and data-driven models can be more accurate in various geographical contexts [19].

It is important to note that there has yet to be previous research in landslide susceptibility assessment in the Aube department despite numerous landslide occurrences in recent decades. The innovative aspect of this research lies in its use of Multi-Criteria Decision-Making (MCDM), statistical, and machine learning models to develop a comprehensive landslide susceptibility map for the study region. This research aims to bridge the existing gap by providing valuable insights into landslide susceptibility assessment in an area previously unexplored, contributing significantly to both scientific knowledge and practical applications. In order to perform the landslide susceptibility mapping for the Aube department, RF, kNN, AHP, and FR models are implemented and then compared. The three main steps in this investigation are as follows: (i) The selection of eight factors conditioning the occurrence of landslides: lithology, slope, proximity to roads, proximity to the hydrographic network, precipitation, altitude, proximity to faults, and the density of quarries. The density of quarries is considered an anthropic activity to generate different forms of landslide instabilities. (ii) Landslide modeling in a GIS environment. (iii) Lastly, the investigation of the performance of these models using ROC and AUC. By comparing the results of the four models, this study can help regional planners choose powerful models for landslide prediction in the Aube department and other regions worldwide.

2. Study Area and Data

2.1. Study Area

The Aube department is located in Northeastern France, the “Great East region” (Figure 1), and is well recognized as an area of significant landslide risks. The investigated area is about 6004 km², consisting of three arrondissements: Bar-Sur-Aube, Nogent-Sur-Seine, and Troyes. The Aube department lies at an altitude ranging from 17 and 382 m (Figure 1).

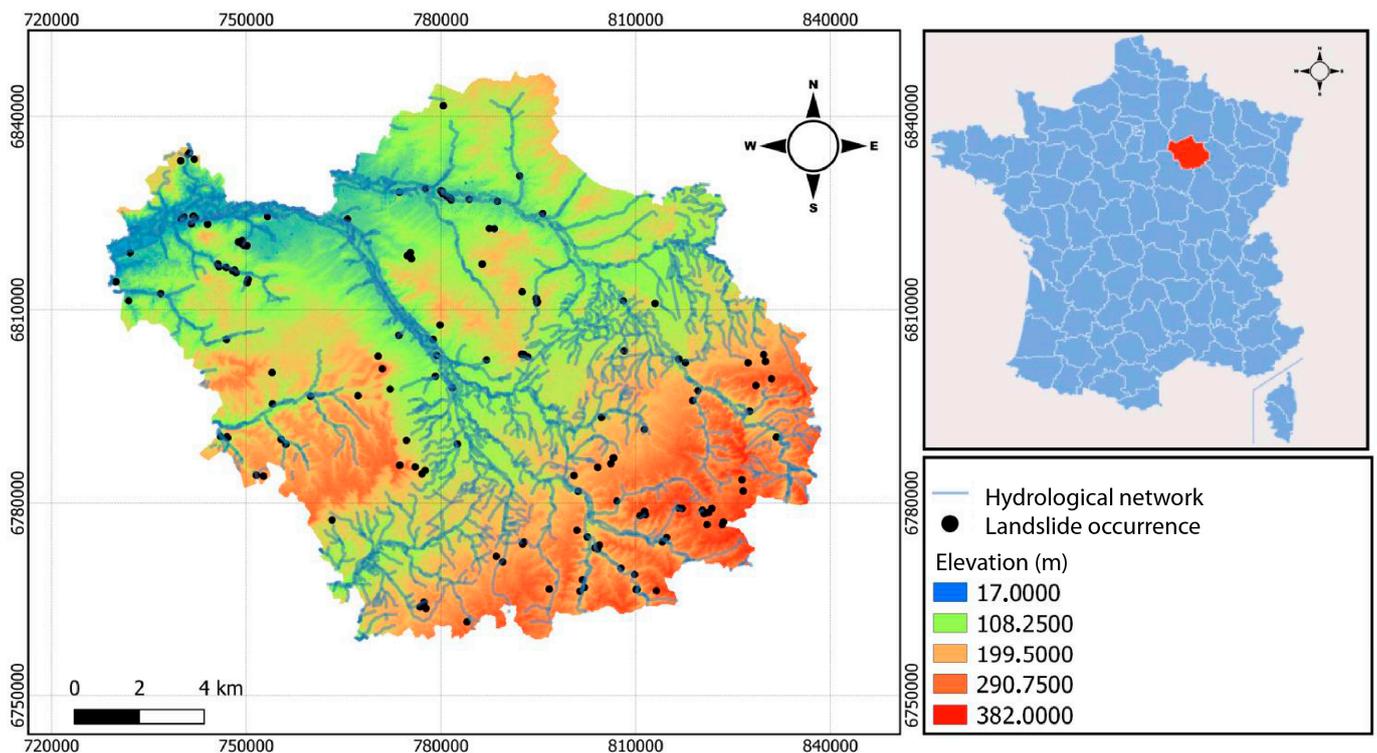


Figure 1. Study area location of the Aube department.

The climatic conditions are characterized by the absence of intense cold or excessive heat. This latter represents a temperate oceanic humid climate zone (Aube, 2016). The annual rainfall of the Aube department varies between 380 and 1200 mm. Numerous landslides have occurred in the area over time (see Figure 1).

Geologically, the study area is mainly formed by a lithologic sedimentary sequence composed of limestone, clay, chalk, and colluvium.

2.2. Landslide Dataset

Landslide occurrence locations were obtained from the points collected by BRGM (in .xls format). We acquired landslide occurrence information from the BRGM website (<https://www.brgm.fr/en/challenges/digital-data-services-infrastructure>) (accessed on 14 October 2021). The georeferenced data were subsequently extracted following the administrative boundaries of the Aube department. This extraction process involved employing the Vector geoprocessing tool “cut” within the QGIS 3.14 software.

2.3. Controlling Factors Thematic Layers

To generate GIS layers of landslide-triggering factors, we used multiple datasets. According to previous studies on landslide-causing factors, many factors could interact to increase the landslide risk, including geology, topography, hydrology, and land management [20–23]. However, the choice of landslide susceptibility controlling factor has no particular guidelines [24]. In the previously defined principal categories, and based on

the study area characteristics, eight principal factors were selected for our study, namely, elevation, lithology, slope, proximity to drainage, proximity to roads, quarry density, proximity to faults, and rainfall. The table below presents each factor and its sources (Table 1). Landslide potential mapping depends on the comprehensive assessment of causal factors, contingent upon relevant data availability within the study area. In this study, we considered eight controlling factors that contribute to landslide occurrences, covering the Aube department's main aspects, including topographical, hydrological, geological, climatological, and anthropogenic aspects. Multicollinearity pertains to the absence of independence among the various factors that contribute to landslides, which can manifest within the used datasets [25]. Variance Inflation Factor (VIF) is the inverse of tolerance. The multicollinearity of landslide factors was evaluated based on VIF (Table 2). All factors expressed tolerable values with VIF less than 10.

Table 1. The current study source of the used data (i.e., integrated control factors).

Triggering Factors	Data Source
Elevation	ASTER-DEM-30 m resolution (https://earthexplorer.usgs.gov), accessed on 5 October 2021.
Lithology	Geological map of the Aube department by BRGM (scale 1:25.000)
Slope	Extracted from ASTER-DEM (30 m resolution)
Precipitations	Obtained from a time series of PERSIANN-CDR (Resolution: 0.04 degrees).
Proximity to roads	Derived from GIS data of the IGN database
Proximity to drainage	Derived from GIS data of the IGN database
Density of quarries	Generated from points collected by BRGM (.xls format)
Faults density	Geological map of the department of Aube, France (scale 1:25.000).

Table 2. The VIF values for the used data (i.e., integrated control factors).

Factors	Elevation	Fault	Hydrology	Lithology	Quarries	Rainfall	Road	Slope
VIF	3.6453	1.1516	1.6088	1.2097	1.3613	3.3052	1.1283	1.0367

Lithology

Several studies have highlighted the crucial significance of lithology in landslide susceptibility mapping [26]. Extensive research supports that rock chemical properties and geotechnical characteristics directly influence the occurrence of landslides. The necessary data about lithologic units in the study area were acquired from the BRGM website (<https://www.brgm.fr/en/challenges/digital-data-services-infrastructure>) (accessed on 14 October 2021). The data were subsequently extracted following the administrative boundaries of the Aube department. Figure 2a shows the lithologic map of the Aube department, where 13 lithologic units are distinguished and gathered further into five main groups based on their proprieties.

Slope

The slope gradient represents the degree of inclination. Various material properties influence it, including cohesion, permeability, deformation behavior, shear strength, and stress distribution. Landslide events often result from the outcome of the interaction between the angle of slope inclination and the material characteristics [27]. We classified the produced slope map into five intervals, namely, 0–5, 5–10, 10–15, 15–20, and 20–60 (Figure 2b).

Rainfall

Rainfall represents a highly critical factor influencing slope instability and landslide occurrence. It has a major influence by altering rock formations along discontinuity planes, leading to decreased cohesion within the rock mass and subsequently increasing the probability of landslides. Rainfall effects were considered after calculating the average

10-year precipitation, using the raster calculator tool and subsequently resampling the resulting layer to a 30 m resolution (Figure 2c).

Road

Road construction activities conducted across slopes can induce destabilization through two distinct mechanisms. Firstly, the terracing of road slopes can augment their gradient, rendering them more susceptible to landslides. Secondly, the construction of roads at foothills and the subsequent removal of lateral support from adjacent slopes substantially increase landslide occurrence probability. The study area map of distance to roads is represented in Figure 2d, where five distance classes were adopted with a minimum of 0 km to a maximum of about 6 km.

Hydrology

The proximity to the hydrological network (drainage), such as stream banks and rivers, can compromise slope stability in numerous ways. Stream erosion, for instance, negatively impacts lateral slope stability and heightens the likelihood of rock ruptures [28]. The study area map of distance to the hydrological network is represented in Figure 2e.

Fault

Fault planes have been demonstrated to create fragility zones, thus intensifying the susceptibility of steep slopes to rupture. Additionally, many studies have demonstrated that local vibrations induced by seismic activity along geological faults can trigger landslides and other destructive phenomena [29]. We employed the Euclidean Distance tool integrated into the QGIS 3.14 software. Subsequently, the obtained features were converted into binary rasters using the rasterization tool to calculate the distance from faults (Figure 2f).

Quarry's density

The quarry's density is an important variable in landslide mapping [30]. The quarries involve extracting valuable materials (e.g., rocks, minerals) using explosive methods. The resulting vibrations can significantly impact the stability of terrains within quarries and their surrounding areas. We utilized the Heat-Map tool within the QGIS 3.14 software to assess quarry density. The generated quarries' density influence factor map is presented in Figure 2g.

Elevation

Elevation is crucial in landslides, impacting the following factors: slope characteristics, water velocity, precipitation, erosion, and gravitational force. These influences are particularly pronounced at higher altitudes, making such areas more susceptible to landslide occurrences. Organic matter, particularly soil nitrogen and carbon, causes hydrogenate activity at higher elevations, which weakens the soil and increases its susceptibility to landslides [31]. To analyze the elevation patterns, the resulting elevation map is categorized into five distinct groups: 17 m, 108.25 m, 199.5 m, 290.75 m, and 382 m (Figure 2h).

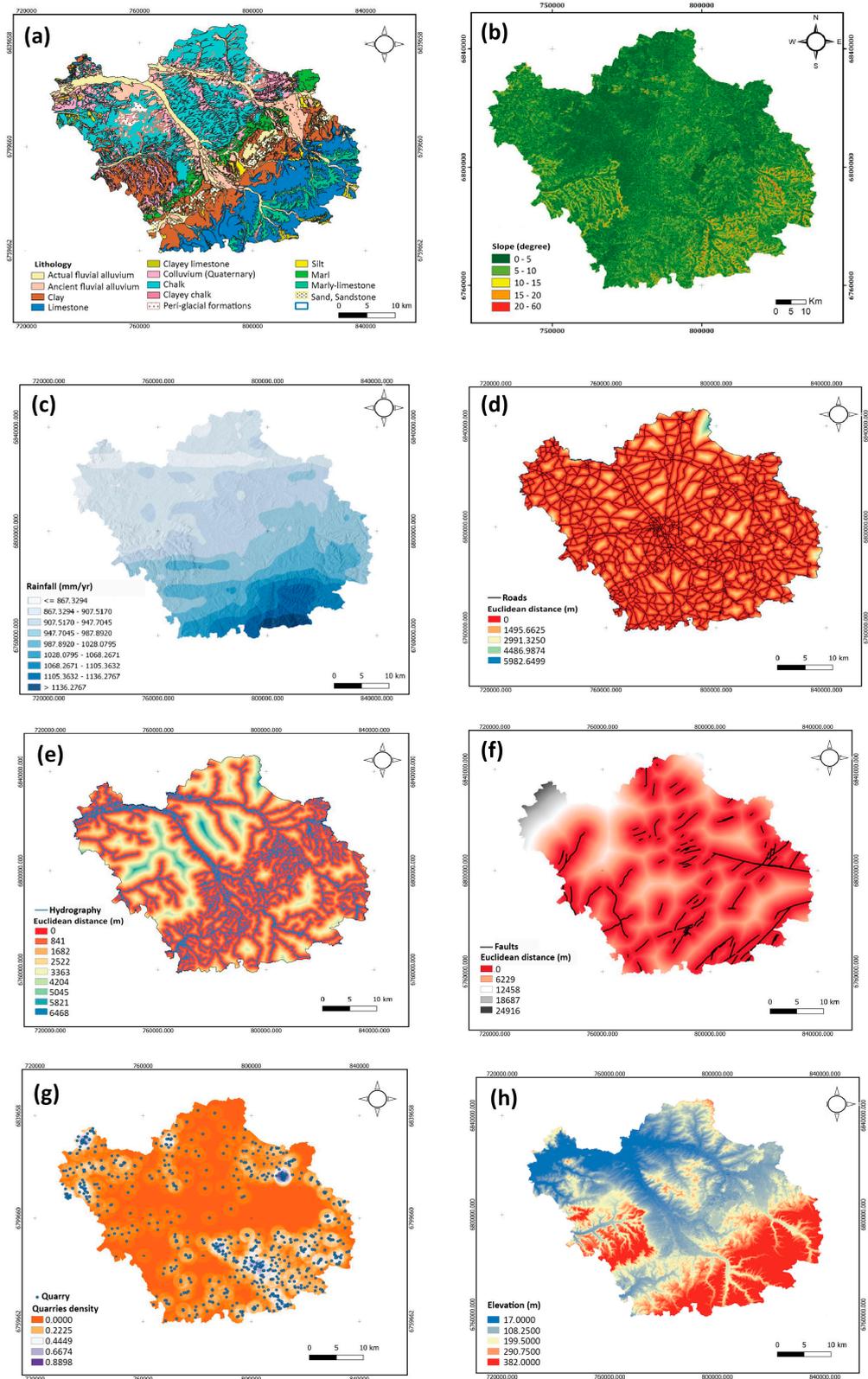


Figure 2. Landslide controlling factors: (a) lithology; (b) slope; (c) rainfall; (d) distance to roads; (e) distance to hydrographical network; (f) distance to faults; (g) density of quarries; and (h) elevation.

3. Methods

The processing steps involved in this study include the generation of thematic layers representing various landslide causative factors using a geographic information system

(GIS) environment. These thematic layers are then integrated and processed to create a comprehensive landslide susceptibility map of the study area in the Aube department, France. Additionally, an inventory of landslides is conducted based on data provided by the French Geological Survey, which is used to train the machine learning models and validation task. Figure 3 shows the flowchart of the present study.

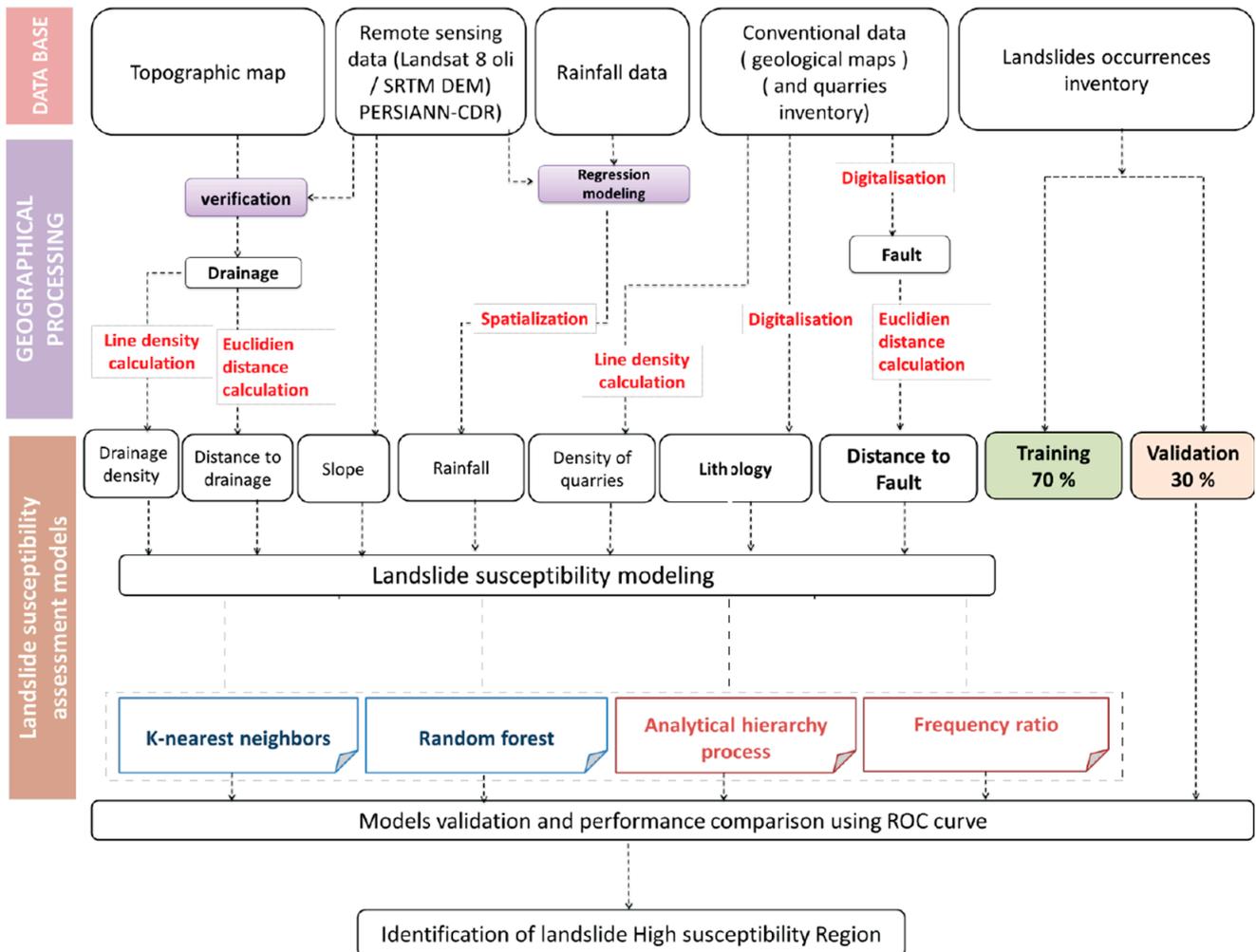


Figure 3. Flow chart.

3.1. Analytic Hierarchy Process (AHP)

The AHP method was introduced by Saaty [5] and is applied to assign normalized weights to each trigger or thematic layer. It is widely accepted as a Multi-Criteria Analysis (MCA) technique for decision making in natural hazard management [32]. The AHP decomposes and makes pairwise comparisons, reducing inconsistencies and providing a priority vector [5]. The analytical hierarchy process involves three main steps: (I) defining the elements or parameters (thematic layers) for the AHP model; (II) generating a pairwise comparison matrix (Table 3) that includes all thematic layers based on a given scale from 1 to 9, where a score of 1 represents equal importance and extreme importance is represented by score of 9 for one theme over the other; and (III) obtaining the thematic layers’ final weight from the normalized eigenvalue or priority vectors associated with the ratio matrix maximum eigenvalue [33]. The following formula is used to calculate the consistency ratio (CR) (Equation (1)):

$$CR = \frac{CI}{RCI} \tag{1}$$

where *RCI* indicates the random coherence index, which is dependent on the order of the matrix, and *CI* indicates the coherence index, which can be calculated according to Equation (2) (see Table 4):

$$CI = (\lambda_{max} - n) / (n - 1) \tag{2}$$

Table 3. Pairwise comparison matrix and normalized principal eigenvector for evaluating landslide causative factors in the AHP.

Factor	Lithology	Slope	Precipitation	Elevation	Distance-Roads	Distance-Drainage	Density of Quarries	Distance-Faults
Lithology	1	2	3	3	4	4	6	8
slope		1	3	4	3	3	5	6
precipitation			1	1/2	2	2	4	5
Elevation				1	3	4	6	8
Distance to roads					1	2	3	3
Distance to hydrography						1	3	1
Quarries density							1	2
Distance to faults								1

Table 4. RCI values for different values of a number of criteria [5].

<i>n</i>	1	2	3	4	5	6	7	8	9	10
(RCI)	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

If the *CR* value is less than or equal to 0.10, it indicates an acceptable level of consistency in the matrix, allowing the analysis to proceed. Conversely, if the *CR* value is greater than 0.10, adjustments to the pairwise comparison matrix are necessary to identify and rectify any inconsistencies [34]. A *CR* value of 0 signifies a perfect consistency level in the pairwise comparison. In the present study, with *n* = 7, the *CR* value is calculated as 0.042, which is strictly less than 0.10. Hence, the pairwise comparison matrix consistency level is deemed tolerable.

3.2. Frequency Ratio (FR)

The frequency ratio model is a commonly employed bivariate statistical model in spatial hazard assessment [35]. It allows for determining the probabilistic relationship between dependent and independent variables. In this study, the frequency ratio model is applied to establish the relationship between each landslide location and a particular class of landslide occurrence factor, utilizing Equation (3):

$$FR = \frac{W/G}{M/T} \tag{3}$$

where *W* is the area of a class of the causal factor, *G* is the total area of the causal factor, *M* is the count of pixels in the class area of the causal factor, and *D* is the total number of pixels of the total study area. Subsequently, the landslide susceptibility index based on the *FR* model is calculated using Equation (4):

$$LSM = \sum_{i=0}^n FRi \tag{4}$$

3.3. k-Nearest Neighbor (k-NN)

The k-nearest neighbor algorithm is a supervised learning algorithm utilized for classification and prediction [36]. It operates based on the proximity principle, which suggests that data points with similar features are close to each other. The kNN algorithm assigns a class to a point based on the nearest neighbors in its classification configuration. It calculates the distance between the target point and its closest neighbors to determine

the class assignment using distance metrics such as the Euclidean or Manhattan distance. By considering the chosen number of neighbors (K), it selects the class with the highest number of votes from these neighbors. The kNN algorithm employs a subordinate density approach and a decision rule to group similar pixels in the feature space. This implies that pixels located nearby in this space are considered part of the same class. This model finds applications in various fields such as pattern recognition, image processing, and data analysis. It enables the classification or prediction of data points based on the characteristics of their nearest neighbors, leveraging the concepts of proximity and similarity.

3.4. Random Forest (RF)

Random forests (RFs) constitute an ensemble of classification and regression trees (CART) pioneered by Breiman [37]. This machine learning algorithm demonstrates comparable or superior accuracy to adaptive boosting while offering faster computational performance [37]. An advantageous feature of RF is its ability to handle continuous and categorical variables. Additionally, RF exhibits robustness against predictor noise, eliminating the need to preselect variables [38]. To regulate model complexity, users typically modify two hyperparameters in RF. Firstly, the number of trees or iterations (*ntree*) influences the number of decision trees; excessively large values may result in overfitting. Secondly, *mtry* determines the number of randomly sampled indicators considered as candidates at each split. We tune the *ntree* and *mtry* parameters in this case study as they profoundly impact our random forest model's performance. In the present study, the model performance is evaluated by systematically tuning the hyperparameters using the grid search method [39] in conjunction with cross-validation (CV) techniques; indeed, grid search could further improve the prediction capability of the ML model [40].

4. Results

The results obtained from the three models AHP, kNN, and RF highlight the following elements, and the most exposed area to landslide risks is in the heart of the study area. This area consists of a clay band with high drainage density. Abrupt variations mark the transitions between this zone and the limestone layer, thus increasing the potential for landslides. Additionally, this area receives significant annual precipitation, ranging between 867 and more than 1100 mm/y. Areas with moderate susceptibility are characterized by flat topography, average altitude, and alluvial lithology.

These zones are typically far from erosion factors such as roads and the hydrographic network. Conversely, regions with low landslide potential are located in the northern part of the study area. These areas are characterized by gentle slopes, lower precipitation levels compared to the southwest zone, and a lithologic composition with lower clay content. Table 5 presents the relation between the influencing factors and landslide occurrences using FR and AHP methods.

Quantitatively evaluating the accuracy of landslide susceptibility maps generated by different classification models (Figure 4) is a crucial step in the process [41]. Additionally, validating the landslide susceptibility model is essential to ensure the practical significance of the resulting maps [42]. For ML-derived maps (Figure 4A,B), most of landslides are located in the very high LS class. A close association with hydrological network and clay geological formation is observed. The AHP-derived map (Figure 4C) shows that the landslide occurrences are more concentrated in the high-elevation area of very high landslide susceptibility. Through the AHP, the geological and topographical control seems clearer. The FR-derived map (Figure 4D) shows that landslide occurrences are located in moderate, high, and very high LS classes. The topographical factor is more dominated as well as the hydrographic network effect. Additionally, various statistical indices, including ROC, are used to assess the predictive performance of the models employed in this study. Based on the validation dataset, the FR model accuracy assessment yields more precise results (AUC = 66.0%) than AHP, outperforming it by 6%. Furthermore, the machine

learning models exhibit superior performance, with RF surpassing kNN by more than 7.3% and achieving an AUC of 95% (Figure 5).

Table 5. Spatial relation between thematic layers and landslides using FR and AHP methods.

Factor	Classes	Area (Pixels)	Area (%)	Landslide (Pixels)	Landslide (%)	FR	AHP	
							Assigned Rate	Weight %
Slope	0–5	3,824,230	57.150	9985	57.378	1.003	2	23%
	0–10	2,167,308	32.400	5482	31.502	0.971	4	
	10–15	515,048	7.700	1295	7.442	0.966	7	
	15–20	132,478	1.970	408	2.345	1.189	8	
	>20	52,169	0.770	232	1.333	1.730	10	
Density of quarries	0–0.07	3,044,553	45.500	6091	35.002	0.768	2	3.5%
	0.07–0.2	2,253,187	33.670	5687	32.680	0.970	4	
	0.2–0.30	814,266	12.160	3573	20.532	1.684	6	
	0.3–0.5	462,708	6.910	1634	9.390	1.358	8	
	>0.5	116,519	1.750	423	2.431	1.388	10	
Distance/ Drainage	0–250	2,863,670	42.790	4358	25.043	0.584	10	5.7%
	250–500	580,094	8.660	1042	5.988	0.691	8	
	500–750	752,296	11.240	1279	7.350	0.653	6	
	750–1000	1,000,062	14.940	2948	16.941	1.134	2	
	>1000	1,495,111	22.340	7781	44.713	2.001	1	
Distance/ fault	0–500	4,371,259	65.328	10,829	62.228	0.951	10	3%
	500–1000	572,063	8.549	1052	6.045	0.707	10	
	1000–1500	575,109	8.595	1537	8.832	1.027	8	
	1500–2000	575,109	8.595	1427	8.200	0.954	6	
	>2000	597,693	8.932	2563	14.728	1.648	4	
Distance/ Road	0–200	1,566,631	23.413	735	4.224	0.180	10	7.7%
	200–400	744,322	11.124	800	4.597	0.413	8	
	400–600	1,069,166	15.979	1560	8.964	0.560	6	
	600–800	1,404,849	20.995	3645	20.946	0.997	2	
	>800	1,906,265	28.489	10,668	61.303	2.152	1	
Precipitation	830–900	1,997,712	29.856	5999	34.473	1.154	6	11%
	900–950	1,829,344	27.339	3193	18.348	0.670	7	
	950–1000	1,639,949	24.509	3123	17.946	0.732	8	
	1000–1060	612,114	9.148	1475	8.476	0.925	9	
	>1060	612,114	9.148	3618	20.791	2.270	10	
Elevation	20	634,093	9.476	3786	21.759	2.297	2	16.9%
	100	3,517,229	52.565	6381	36.672	0.697	4	
	200	1,554,779	23.236	4820	27.701	1.191	6	
	300	837,664	12.519	2048	11.770	0.939	8	
	>300	147,468	2.204	365	2.098	0.951	10	
Lithology	1	1,019,944	15.243	4321	24.830	1.627	8	29.2%
	2	1,595,959	23.851	2528	14.527	0.608	6	
	3	254,879	3.809	386	2.218	0.592	4	
	4	1,610,862	24.074	5736	32.962	1.368	2	
	5	2,209,589	33.022	4428	25.445	0.770	8	

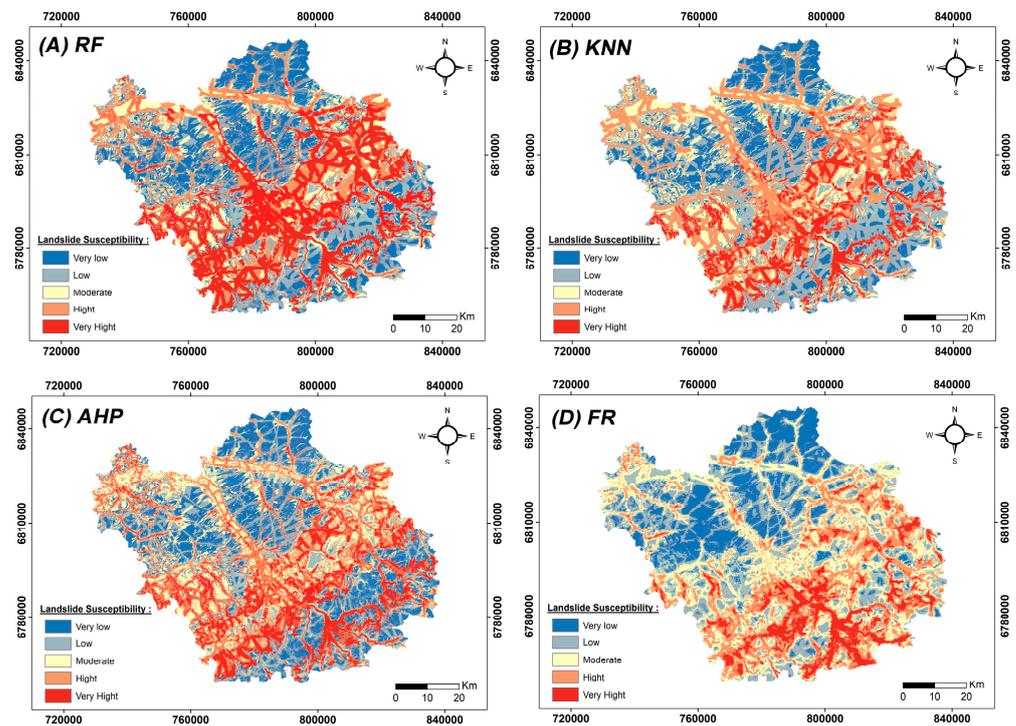


Figure 4. Aube department’s landslide susceptibility maps.

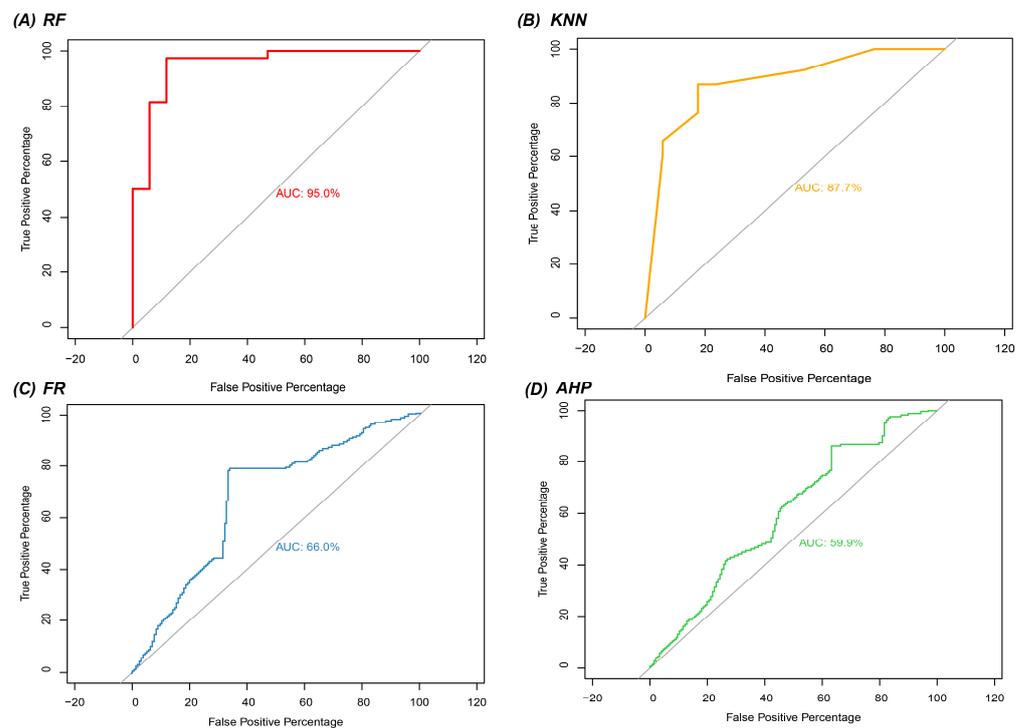


Figure 5. Assessment of the performance of RF (A), kNN (B), FR (C), and AHP (D) using the AUC.

However, when comparing the different models used for landslide susceptibility mapping, it becomes evident that the RF model outperforms the others, demonstrating the highest level of performance. The resulting landslide susceptibility map generated with the RF model exhibits superior accuracy and predictive capability compared to the maps produced by the kNN and AHP models. The RF model proves to be the most reliable and robust in accurately identifying areas prone to landslides. The high precision

and reliability exhibited by the RF model solidify its position as the preferred choice for landslide mapping purposes. Researchers and practitioners can confidently rely on the RF model for effective and accurate identification of landslide-prone areas, aiding mitigation and management efforts.

As shown in Figure 6A, the relative importance of the controlling factors on landslide risk derived from the RF shows a decreasing tendency ordered as lithology, distance to hydrography, elevation, distance to roads, slope, rainfall, distance to faults, and density of quarries. Using the kNN algorithm, the factor's relative importance order from most to least important follows lithology, distance to hydrography, distance to roads, elevation, slope, rainfall, distance to faults, and density of quarries (Figure 6B).

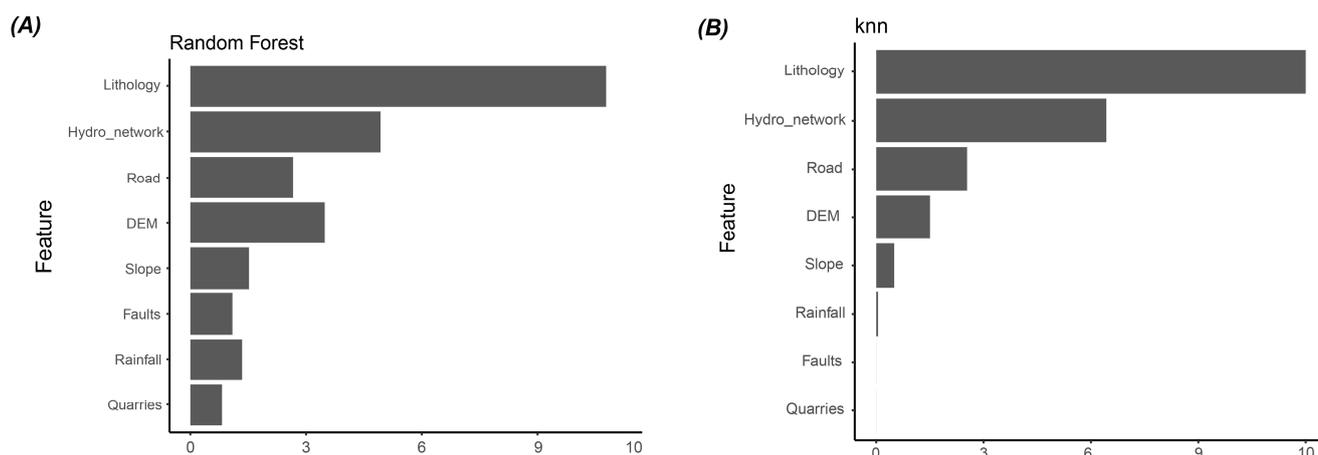


Figure 6. Importance of selected factors using algorithms calculated for Random Forest (A) and kNN (B) algorithms. Abs: Hydro_network: proximity to the hydrological network, DEM: Digital Elevation Model.

Generally, the relevance of landslide controlling factors shows some coincidence, which is observed with the first most important factors, namely lithology and distance to hydrography. The third most important factor for RF is elevation, while the elevation factor is placed in the fourth position for kNN and replaced by distance to roads. Likewise, density of quarries is ranked as the least important factor for the two models. However, the importance of kNN's last factors is considered negligible. At the same time, the RF has considerable importance in our study and in many previous landslide studies that can demonstrate the relative effectiveness of RF model in accurate feature importance computation [37].

5. Discussions

The results of this study underscore the significance of landslides as a main natural hazard capable of causing substantial material and human losses. The Aube Department has witnessed several significant landslide incidents, establishing it as a critical area warranting further investigation. The primary objective of this research was to assess landslide susceptibility in the region, employing a combination of statistical analysis, machine learning techniques, and multi-criteria decision-making methods.

The landslide susceptibility maps obtained in this study underwent a thorough evaluation using the Receiver Operating Characteristic (ROC) curve analysis. This evaluation allowed for a comprehensive assessment of the model's performance in predicting landslides. The accuracy assessment revealed compelling insights into the effectiveness of the different approaches. The FR model demonstrated exceptional accuracy among the evaluated models, as indicated by its considerable AUC value of 66.0%. This finding highlights the FR model's ability to identify areas susceptible to landslides accurately.

On the other hand, the AHP method, while still effective, exhibited slightly lower accuracy, with a 6% difference compared to the FR model. Moreover, the machine learning

models showcased their potential in landslide susceptibility mapping. The RF algorithm stood out among the machine learning models, surpassing the kNN algorithm with a remarkable AUC value of 95%. The RF model's superior performance further solidifies its position as the preferred choice for accurate and reliable landslide mapping.

The high precision and reliability exhibited by the RF model have significant implications for both researchers and practitioners involved in landslide mitigation and management. They can confidently rely on the RF model's robust capabilities to effectively identify areas prone to landslides. By leveraging the accurate information provided by the RF model, mitigation efforts can be better targeted, and appropriate management strategies can be implemented to ensure the safety and well-being of the affected areas. A comprehensive analysis of the landslide susceptibility results showed that several factors played significant roles in influencing landslide occurrence in the study area. Among these factors, lithology, distance to the hydrographic network, distance to roads, and elevation emerged as the main controlling factors shaping the susceptibility to landslides.

The lithologic composition of the area was found to have a profound influence on landslide occurrence. Certain lithologic units displayed higher susceptibility due to their inherent characteristics, such as low cohesion or high permeability. Additionally, proximity to the hydrographic network and roads proved to be crucial factors, as areas near these features exhibited increased vulnerability to landslides. This can be attributed to the potentially destabilizing effects of water flow and human activity. These results (controlling factor importance) correlate well with the findings of other studies, e.g., [43,44].

Additionally, elevation was another crucial factor affecting landslide susceptibility. Steep slopes and areas at higher elevations were more prone to landslides due to gravitational forces and increased instability. These findings highlight the significance of terrain characteristics in determining landslide susceptibility in the study area.

Understanding the dominant controlling factors influencing landslides in the study area provides valuable insights for developing effective mitigation strategies and land-use planning measures. Incorporating this knowledge into land-use policies can help minimize the exposure of vulnerable areas to human settlements and infrastructure. Additionally, identifying these controlling factors can aid in designing and implementing targeted engineering measures and slope stabilization techniques. Overall, these findings contribute to a better understanding of landslide dynamics and support proactive measures to reduce the risks associated with landslides in the region. Otherwise, through the use of many diverse machine learning models, ensemble learning increases prediction accuracy. Another cutting-edge method, deep learning, achieves strong capabilities and adaptability through layered idea hierarchy without requiring human feature extraction [45,46].

6. Conclusions

This study aimed to assess landslide susceptibility in the Aube department (Northeast France) using AHP, FR, RF, and kNN models. Effective landslide controlling factors were carefully carried out, including distance to hydrography, density of quarries, elevation, slope, lithology, distance to roads, distance to faults, and rainfall. The accuracy assessment demonstrated that RF and kNN were the best models for mapping landslide susceptibility in the study area. The derived maps from applying the models were categorized into five classes: very high, high, moderate, low, and very low landslide susceptibility.

The zones of high landslide susceptibility were mainly detected in clays and fluvial alluvium terrain. Based on the analysis of landslide susceptibility results in the study area, the primary factors influencing landslide susceptibility were lithology, elevation, distance to the hydrographic network, and distance to roads. On the other hand, density of quarries, rainfall, distance to faults, and slope presented a relatively minor influence. Validation of the landslide susceptibility maps using AUC revealed that ML models had performed better than statistical ones did. Considering the four models, in ascending order, the performances were 59.9%, 66.0%, 87.7%, and 95.0% for the AHP, FR, kNN, and RF models, respectively. The approach adopted in the present study was highly successful

in assessing the susceptibility of landslides in the Aube department region, especially by using kNN and RF ML models.

Furthermore, the map depicting the landslide susceptibility zones in the Aube department has provided valuable information about potential risks in both the present and future. Thus, it can be a practical resource for planners and decision-makers involved in land-use planning and natural hazard management in the region, as it can help prevent potential hazards by providing appropriate safety measures. The methodology utilized in this study could also be adapted in other regions with similar characteristics to accurately identify the geophysical and geomorphological features.

Author Contributions: Conceptualization and writing—original draft, A.J. (Abdessamad Jari), S.M., S.H. and A.K.; methodology, A.J. (Abdessamad Jari), H.M., A.J. (Amine Jellouli) and A.E.H.; software, A.J. (Abdessamad Jari), S.M., S.H. and A.K.; supervision, E.M.B. and A.E.H.; review and editing, E.M.B., A.E.H. and A.B. All authors have read and agreed to the published version of the manuscript.

Funding: The authors declare no external funding.

Data Availability Statement: The used data in the present manuscript are confidential.

Acknowledgments: The authors thank the French Geological Survey (BRGM) for providing the lithologic map and the landslide inventory. Also, they would thank the support of Gustave Eiffel University and Sultan Moulay Slimane University.

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

References

1. Wallemacq, P.; House, R. *Economic Losses, Poverty & Disasters 1998–2017*; Centre for Research on the Epidemiology of Disasters: Brussels, Belgium; United Nations Office for Disaster Risk Reduction: Geneva, Switzerland, 2018.
2. Zhang, X.; Li, L.; Xu, C. Large-scale landslide inventory and their mobility in Lvliang City, Shanxi Province, China. *Nat. Hazards Res.* **2022**, *2*, 111–120. [[CrossRef](#)]
3. Van Westen, C.J. The modelling of landslide hazards using GIS. *Surv. Geophys.* **2000**, *21*, 241–255. [[CrossRef](#)]
4. El Jazouli, A.; Barakat, A.; Khellouk, R. GIS-multicriteria evaluation using AHP for landslide susceptibility mapping in Oum Er Rbia high basin (Morocco). *Geoenviron. Disasters* **2019**, *6*, 3. [[CrossRef](#)]
5. Saaty, T.L. A scaling method for priorities in hierarchical structures. *J. Math. Psychol.* **1980**, *15*, 234–281. [[CrossRef](#)]
6. Yang, J.; Song, C.; Yang, Y.; Xu, C.; Guo, F.; Xie, L. New method for landslide susceptibility mapping supported by spatial logistic regression and GeoDetector: A case study of Duwen Highway Basin, Sichuan Province, China. *Geomorphology* **2019**, *324*, 62–71. [[CrossRef](#)]
7. Polykretis, C.; Chalkias, C. Comparison and evaluation of landslide susceptibility maps obtained from weight of evidence, logistic regression, and artificial neural network models. *Nat. Hazards* **2018**, *93*, 249–274. [[CrossRef](#)]
8. Pradhan, B. Landslide susceptibility mapping of a catchment area using frequency ratio, fuzzy logic and multivariate logistic regression approaches. *J. Indian Soc. Remote Sens.* **2010**, *38*, 301–320. [[CrossRef](#)]
9. Akinci, H.; Yavuz Ozalp, A. Landslide susceptibility mapping and hazard assessment in Artvin (Turkey) using frequency ratio and modified information value model. *Acta Geophys.* **2021**, *69*, 725–745. [[CrossRef](#)]
10. Kilicoglu, C. Investigation of the effects of approaches used in the production of training and validation data sets on the accuracy of landslide susceptibility mapping models: Samsun (Turkey) example. *Arab. J. Geosci.* **2021**, *14*, 2106. [[CrossRef](#)]
11. 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](#)]
12. Hussain, M.A.; Chen, Z.; Kalsoom, I.; Asghar, A.; Shoab, M. Landslide susceptibility mapping using machine learning algorithm: A case study along Karakoram Highway (KKH), Pakistan. *J. Indian Soc. Remote Sens.* **2022**, *50*, 849–866. [[CrossRef](#)]
13. Yeon, Y.-K.; Han, J.-G.; Ryu, K.H. Landslide susceptibility mapping in Injae, Korea, using a decision tree. *Eng. Geol.* **2010**, *116*, 274–283. [[CrossRef](#)]
14. Dou, J.; Yunus, A.P.; Bui, D.T.; Merghadi, A.; Sahana, M.; Zhu, Z.; Chen, C.-W.; Han, Z.; Pham, B.T. Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan. *Landslides* **2020**, *17*, 641–658. [[CrossRef](#)]
15. Deng, Q.-X.; He, J.; Cao, Z.-J.; Papaioannou, I.; Li, D.-Q.; Phoon, K.-K. Bayesian learning of Gaussian mixture model for calculating debris flow exceedance probability. *Georisk Assess. Manag. Risk Eng. Syst. Geohazards* **2022**, *16*, 154–177. [[CrossRef](#)]
16. Wang, Z.; Liu, Q.; Liu, Y. Mapping landslide susceptibility using machine learning algorithms and GIS: A case study in Shexian County, Anhui Province, China. *Symmetry* **2020**, *12*, 1954. [[CrossRef](#)]

17. Ghasemian, B.; Shahabi, H.; Shirzadi, A.; Al-Ansari, N.; Jaafari, A.; Kress, V.R.; Geertsema, M.; Renoud, S.; Ahmad, A. A robust deep-learning model for landslide susceptibility mapping: A case study of Kurdistan Province, Iran. *Sensors* **2022**, *22*, 1573. [[CrossRef](#)]
18. Lai, J.-S.; Tsai, F. Improving GIS-based landslide susceptibility assessments with multi-temporal remote sensing and machine learning. *Sensors* **2019**, *19*, 3717. [[CrossRef](#)] [[PubMed](#)]
19. Yu, L.; Zhou, C.; Wang, Y.; Cao, Y.; Peres, D.J. Coupling Data-and Knowledge-Driven Methods for Landslide Susceptibility Mapping in Human-Modified Environments: A Case Study from Wanzhou County, Three Gorges Reservoir Area, China. *Remote Sens.* **2022**, *14*, 774. [[CrossRef](#)]
20. Alkhasawneh, M.S.; Ngah, U.K.; Tay, L.T.; Mat Isa, N.A.; Al-Batah, M.S. Determination of important topographic factors for landslide mapping analysis using MLP network. *Sci. World J.* **2013**, *2013*, 415023. [[CrossRef](#)] [[PubMed](#)]
21. Gunzburger, Y.; Merrien-Soukatchoff, V.; Guglielmi, Y. Influence of daily surface temperature fluctuations on rock slope stability: Case study of the Rochers de Valabres slope (France). *Int. J. Rock Mech. Min. Sci.* **2005**, *42*, 331–349. [[CrossRef](#)]
22. Regmi, A.D.; Devkota, K.C.; Yoshida, K.; Pradhan, B.; Pourghasemi, H.R.; Kumamoto, T.; Akgun, A. Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. *Arab. J. Geosci.* **2014**, *7*, 725–742. [[CrossRef](#)]
23. Nwazelibe, V.E.; Unigwe, C.O.; Egbueri, J.C. Testing the performances of different fuzzy overlay methods in GIS-based landslide susceptibility mapping of Udi Province, SE Nigeria. *CATENA* **2023**, *220*, 106654. [[CrossRef](#)]
24. Ayalew, L.; Yamagishi, H. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology* **2005**, *65*, 15–31. [[CrossRef](#)]
25. Chen, W.; Peng, J.; Hong, H.; Shahabi, H.; Pradhan, B.; Liu, J.; Zhu, A.-X.; Pei, X.; Duan, Z. Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China. *Sci. Total Environ.* **2018**, *626*, 1121–1135. [[CrossRef](#)] [[PubMed](#)]
26. Althuwaynee, O.F.; Pradhan, B. Semi-quantitative landslide risk assessment using GIS-based exposure analysis in Kuala Lumpur City. *Geomat. Nat. Hazards Risk* **2017**, *8*, 706–732. [[CrossRef](#)]
27. Lee, S. Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. *Int. J. Remote Sens.* **2005**, *26*, 1477–1491. [[CrossRef](#)]
28. Park, S.; Choi, C.; Kim, B.; Kim, J. Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea. *Environ. Earth Sci.* **2013**, *68*, 1443–1464. [[CrossRef](#)]
29. Ray, P.C.; Parvaiz, I.; Jayangondaperumal, R.; Thakur, V.; Dadhwal, V.; Bhat, F. Analysis of seismicity-induced landslides due to the 8 October 2005 earthquake in Kashmir Himalaya. *Curr. Sci.* **2009**, *97*, 1742–1751.
30. Ageenko, A.; Hansen, L.C.; Lyng, K.L.; Bodum, L.; Arsanjani, J.J. Landslide susceptibility mapping using machine learning: A Danish case study. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 324. [[CrossRef](#)]
31. Bopche, L.; Rege, P.P. Landslide susceptibility mapping: An integrated approach using geographic information value, remote sensing, and weight of evidence method. *Geotech. Geol. Eng.* **2022**, *40*, 2935–2947. [[CrossRef](#)]
32. Panchal, S.; Shrivastava, A.K. Landslide hazard assessment using analytic hierarchy process (AHP): A case study of National Highway 5 in India. *Ain Shams Eng. J.* **2022**, *13*, 101626. [[CrossRef](#)]
33. El Morjani, Z.E.A. *Conception D'un Système D'information à Référence Spatiale Pour la Gestion Environnementale: Application à la Sélection de Sites Potentiels de Stockage de Déchets Ménagers et Industriels en Région Semi-Aride (Souss, Maroc)*; University of Geneva: Geneva, Switzerland, 2002.
34. Fashae, O.A.; Tijani, M.N.; Talabi, A.O.; Adedeji, O.I. Delineation of groundwater potential zones in the crystalline basement terrain of SW-Nigeria: An integrated GIS and remote sensing approach. *Appl. Water Sci.* **2014**, *4*, 19–38. [[CrossRef](#)]
35. Ait Naceur, H.; Igmoulan, B.; Namous, M.; Amrhar, M.; Bourouay, O.; Ouayah, M.; Jadoud, M. A comparative study of different machine learning methods coupled with GIS for landslide susceptibility assessment: A case study of N'fis basin, Marrakesh High Atlas (Morocco). *Arab. J. Geosci.* **2022**, *15*, 1100. [[CrossRef](#)]
36. Fix, E.; Hodges, J. *Discriminatory Analysis, Nonparametric Discrimination*; Tex., Project 21-49-004, Rep. 4, Contract AF-41- (128)-31; U.S. Air Force School of Aviation Medicine, Randolph Field: Montgomery County, OH, USA, 1951.
37. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
38. Díaz-Uriarte, R.; Alvarez de Andrés, S. Gene selection and classification of microarray data using random forest. *BMC Bioinform.* **2006**, *7*, 3. [[CrossRef](#)]
39. Zhu, A.-X.; Miao, Y.; Liu, J.; Bai, S.; Zeng, C.; Ma, T.; Hong, H. A similarity-based approach to sampling absence data for landslide susceptibility mapping using data-driven methods. *Catena* **2019**, *183*, 104188. [[CrossRef](#)]
40. Hajaj, S.; El Harti, A.; Jellouli, A.; Pour, A.B.; Mnissar Himyari, S.; Hamzaoui, A.; Hashim, M. Evaluating the Performance of Machine Learning and Deep Learning Techniques to HyMap Imagery for Lithological Mapping in a Semi-Arid Region: Case Study from Western Anti-Atlas, Morocco. *Minerals* **2023**, *13*, 766. [[CrossRef](#)]
41. Luo, W.; Liu, C.-C. Innovative landslide susceptibility mapping supported by geomorphon and geographical detector methods. *Landslides* **2018**, *15*, 465–474. [[CrossRef](#)]
42. Pham, B.T.; Bui, D.T.; Prakash, I.; Dholakia, M. Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *Catena* **2017**, *149*, 52–63. [[CrossRef](#)]

43. Yu, L.; Cao, Y.; Zhou, C.; Wang, Y.; Huo, Z. Landslide susceptibility mapping combining information gain ratio and support vector machines: A case study from Wushan segment in the Three Gorges Reservoir area, China. *Appl. Sci.* **2019**, *9*, 4756. [[CrossRef](#)]
44. Zhang, T.-Y.; Han, L.; Zhang, H.; Zhao, Y.-H.; Li, X.-A.; Zhao, L. GIS-based landslide susceptibility mapping using hybrid integration approaches of fractal dimension with index of entropy and support vector machine. *J. Mt. Sci.* **2019**, *16*, 1275–1288. [[CrossRef](#)]
45. Zhang, W.; Gu, X.; Tang, L.; Yin, Y.; Liu, D.; Zhang, Y. Application of machine learning, deep learning and optimization algorithms in geoenvironment and geoscience: Comprehensive review and future challenge. *Gondwana Res.* **2022**, *109*, 1–17. [[CrossRef](#)]
46. Zeng, T.; Wu, L.; Peduto, D.; Glade, T.; Hayakawa, Y.S.; Yin, K. Ensemble learning framework for landslide susceptibility mapping: Different basic classifier and ensemble strategy. *Geosci. Front.* **2023**, *14*, 101645. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.