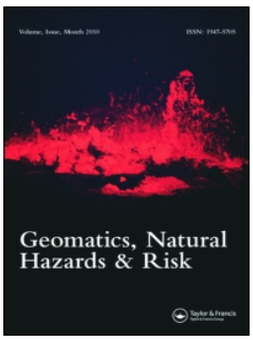

Geomatics, Natural Hazards and Risk

ISSN: 1947-5705 (Print) 1947-5713 (Online) Journal homepage: https://www.tandfonline.com/loi/tgnh20

Predicting landslide susceptibility and risks using GIS-based machine learning simulations, case of upper Nyabarongo catchment

Jean Baptiste Nsengiyumva & Roberto Valentino

To cite this article: Jean Baptiste Nsengiyumva & Roberto Valentino (2020) Predicting landslide susceptibility and risks using GIS-based machine learning simulations, case of upper Nyabarongo catchment, Geomatics, Natural Hazards and Risk, 11:1, 1250-1277, DOI: 10.1080/19475705.2020.1785555

To link to this article: https://doi.org/10.1080/19475705.2020.1785555

Published online: 06 Jul 2020.

GEOMATICS, NATURAL HAZARDS AND RISK
2020, VOL. 11, NO. 1, 1250–1277
https://doi.org/10.1080/19475705.2020.1785555

Predicting landslide susceptibility and risks using GIS-based machine learning simulations, case of upper Nyabarongo catchment

Jean Baptiste Nsengiyumva[a] and Roberto Valentino[b]

[a]Institute of Policy Analysis and Research-Rwanda (IPAR-Rwanda), Kimihurura-Kigali, Rwanda; [b]Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parma, Italy

ABSTRACT

Sustainable landslide mitigation requires appropriate approaches to predict susceptible zones. This study compared the performance of Logistic Model Tree (LMT), Random Forest (RF) and Naïve-Bayes Tree (NBT) in predicting landslide susceptibility for the upper Nyabarongo catchment (Rwanda). 196 past landslides were mapped using field investigations. Thus, the inventory map was split into training and testing datasets. Fifteen predisposing factors were analysed and information gain (IG) technique was used to analyse the correlation between factors and observed landslides. Therefore, the area under receiver operating characteristic (AUROC) with other statistical estimators including accuracy, precision, and root mean square error (RMSE) were employed to compare the models. The AUC values were 78.7%, 80.9% and 82.4% for RF, LMT and NBT models, respectively. Additionally, the NBT produced the highest accuracy and precision values (0.799 and 0.745, respectively). Regarding RMSE values, the NBT model achieved an optimized prediction than RF and LMT models (0.301; 0.428 and 0.364, respectively). The results of the current study may inform further studies and appropriate landslide risk reduction and mitigation measures. They can also be instrumental for policy and decision making in regards with natural risk management.

ARTICLE HISTORY
Received 27 May 2020
Accepted 13 June 2020

KEYWORDS
Landslide; GIS; IPAR-Rwanda; Nyabarongo; disaster-risk

1. Introduction

Landslides are among the greatest deadly natural hazards throughout the world (Chen et al. 2018a; 2018b). Thus, the losses and fatalities induced by landslide hazards are continuously numerous. Most of landslides are generally provoked by climate change effects and anthropogenic factors (UNISDR 2016). Landslides are consequently very common in most countries especially in mountainous areas, which are

CONTACT Jean Baptiste Nsengiyumva j.nsengiyumva@ipar-rwanda.org

either highly susceptible or vulnerable. It is therefore required to put in place strong and appropriate measures to control and minimize all impacts induced by landslides.

As previously confirmed by studies (Claessens et al. 2007; Pellicani et al.2014; Pourghasemi et al.2018), the way to deal with landslide hazards, is to map the particular volume and type of their spatial likelihood in relation with their occurrence within a given area. Hence, this is mostly termed as susceptibility mapping (Corominas et al. 2014; Paulín et al. 2016; Tseng et al.2015). This is normally composed of different aspects such as conditioning factors, landslide categories, failure mechanisms and the coverage of affected areas (Abella and Van Westen 2007). Therefore, any study of landslide susceptibility modeling has to consider those highlighted parameters. The selection of approaches and conditioning factors has to consider categories of landslides, analysis levels, study area features and availability of datasets (Zêzere et al. 2017). Additionally, for any type of landslide, the susceptibility has to be evaluated individually since different categories of landslide hazards present special uniqueness linked to different threshold conditions based on the controlling factors (Tseng et al. 2015).

At present, there exist different classes of landslides in the literature, ranging from simple to very complex (Cruden and Varnes 1996; Nsengiyumva et al. 2019). These include deep-seated, falls, topples, rotational, flows, lateral spreads, complex, shallow and translational landslides among others (Pradhan et al. 2011). Landslides are typically triggered due to natural slope failures that collapses devastatingly. These hazards usually pose a grave threat to lives, properties, environment, and infrastructure. In many cases, landslides are mainly triggered in mountainous and steep regions in prolonged period of intense rainfall events (Godt et al. 2008; Valentino et al. 2014). Thus, precipitation increases the pore pressure in the soil, and the variations in pore pressures are extremely variable due to the hydraulic conductivity, topographic nature, and further soil properties. In addition to soil physical properties, land-cover changes due to anthropogenic factors also affect the rate and spatial dispersal of landslides. Particularly, forests removal, inappropriate land use practices and cultivation on fragile hill and steep slopes are among the major triggers of mass movements (Akgun and Erkan 2016; Bordoni et al. 2015).

Throughout the previous decades, landslide susceptibility modeling has attracted the attention of various scholars around the world (Chen et al. 2017a; Ramani et al. 2011; Zêzere et al. 2017), however, landslides still constitute a global danger. Moreover, numerous methods and techniques exist for susceptibility mapping. They, therefore, range from qualitative approaches to quantitative models (Juliev et al. 2019; Zêzere et al. 2017). The qualitative approaches are mainly grounded on expert's opinion, and they include magnitude frequency, active mapping, Boolean logic, fuzzy logic (Abella and Van Westen 2007; Carrara et al. 1991; Carrara et al. 1999; Cervi et al. 2010). Conversely, the quantitative approaches are built on statistical analysis (bivariate and multi-variate methods) and deterministic theories (SINMAP, TRIGRS and SHALSTAB methods). Additionally, there exist another category of approaches that are identified as semi-quantitative, and they include analytic hierarchy (AHP), the heuristic methods and spatial multi-criteria evaluation (SMCE) (Pisano et al. 2017). Within the very recent years, another category of methods has been introduced for

modeling landslide susceptibility. This is composed of machine learning and data mining techniques such as the logistic regression models (LRM), support vector machines (SVM), artificial neural network (ANN), and decision tree models (DT) (Chen et al. 2017b). Generally, it seems that the machine leaning algorithms enriched the quality and accuracy of generated susceptibility maps (Chen et al. 2019) and these techniques were confirmed to achieve improved performance than classical methods (Chen et al. 2019). Though various methods exist, the prediction of landslide susceptibility is always challenging across the globe (Dou et al. 2018).

The literature emphasizes that various methods have been extensively compared to study susceptibility throughout the world (Chen et al. 2018a; Juliev et al. 2019). Qualitative models as well as data-driven models namely statistical approaches have been explored and applied to study landslides through comparative analysis (Dou et al. 2018; Van Den Eeckhaut et al. 2010). Thus, it was concluded that generated maps were precise and accurate. By comparing qualitative and data-driven techniques, it was ascertained therefore that data-driven approaches generate very objective outcomes and reduce the subjectivity whilst giving weights to conditioning factors. They yield more objective and reproducible outcomes in comparison with qualitative methods (Dou et al. 2018; Yalcin 2008).

Previous studies on landslide susceptibility mapping also compared various data-driven models including multivariate and bivariate techniques (Lanfredi Sofia et al.2018). Different comparative studies showed that multivariate models perform better than bivariate methods. Mostly, the susceptibility analysis using multivariate statistics evaluates the correlation between landslide spatial distribution and controlling factors. Furthermore, the bivariate statistical analysis relates independently each conditioning factor with the landslide distribution. Thus, weights are assigned to conditioning factors based on landslide density. Additionally, within recent comparative studies on landslide susceptibility modelling, statistical models were applied in comparison with machine learning techniques (Chen et al. 2018a; Goetz et al. 2015).

Furthermore, for susceptibility study, various studies made an extensive comparison between data-driven models and deterministic models (Akgun and Erkan 2016; Zizioli et al. 2013; Ciurleo et al. 2017). Typically, deterministic models produced a slight differentiation in modeling landslide susceptibility compared to data-driven models (Paulín et al. 2016; Pourghasemi et al. 2018). Presently, the deterministic models proved rather promising approaches in modelling landslide susceptibility. However, the deterministic methods require extensive soil datasets and they proved not appropriate for large areas or where there is little data (Chen et al. 2017c).

Current literature discloses that different models have been compared to study susceptibility across the globe. However, less comparative analysis was done between recent and novel GIS-based machine learning approaches in predicting landslide susceptibility especially for Africa. The application of modern approaches in predicting spatial probability of landslides is essential in some African areas since the higher accuracy of susceptibility maps may influence land management, planning and protection policies in developing countries (Bui et al. 2017; Bui et al. 2016). Additionally, it is very useful to investigate the comparative analysis among varied models to achieve excellent performance and reasonable results for susceptibility mapping.

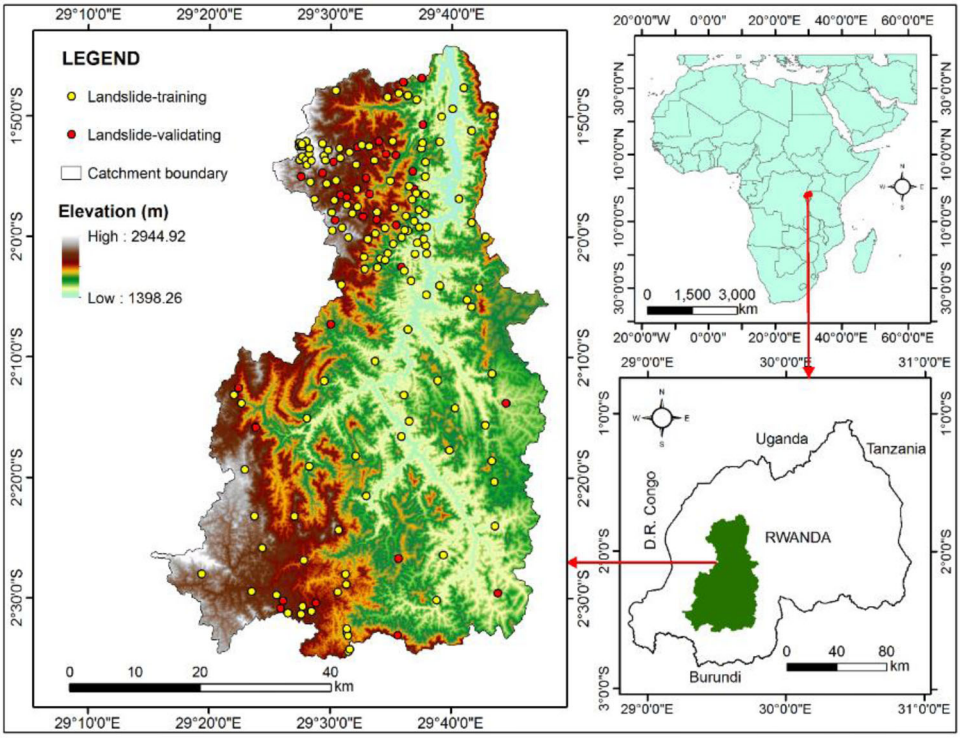

Figure 1. Location of the study area and the landslide inventory map.

Thus, the comparison of methods helps to highlight the advantages and limitations of models in producing landslide susceptibility maps (Ding et al. 2016). As ascertained by Lanfredi Sofia et al. (2018), the absence of comparative analysis of susceptibility methods compromises their reliability and may also cause their misuse. Therefore, to bridge the above mentioned deficiency, the present research intends to explore and analyse the performance and predictive capability of three GIS-based machine learning techniques. These include the random forest (RF), the logistic model (LMT) as well as the naïve-bayes tree (NBT) in predicting landslide susceptibility for the upper Nyabarongo catchment of Rwanda. The present study represents indeed a novel effort of comparing susceptibility modelling analysis using machine leaning simulations for Africa and the study area. Furthermore, this study evaluated and compared the results using different statistical estimators including the receiver operating characteristics (AUROC), root mean square error (RMSE), accuracy and precision.

2. Study area

The current study was conducted in the south-western Rwanda, and covered the entire upper Nyabarongo catchment (Figure 1). The area is mostly dominated by hills and valleys, and it is among the steepest areas in the country (Ndayisaba et al. 2016; Nsengiyumva et al. 2018). This region has a tropical climate though the temperature tends to decrease because of high altitude. Thus, this situation influences rainfall aspects and the average annual precipitation varies between 1000 mm to 1600 mm/year

(Nsengiyumva et al. 2018; Nyesheja et al. 2019). It is located in a heavy rainfall area comparing to other parts of Rwanda.

Geomorphologically, the upper Nyabarongo catchment belongs to the Congo-Nile ridge region of Rwanda. Generally, the study area represents the hilly land of Rwanda (Figure 1) and extends over landslide-prone and hilly areas with high elevation stretching between 1398.26 m and 2944.92 m above sea level. Largely, the study area has an elevation rising from east to western part and this becomes a major cause of landslide hazards.

This mountainous terrain is located within the east of the Kivu Lake covering a total area of about 3,743.5 km^2 with a perimeter of 410 km. Therefore, the area is positioned within 1°50′–2°20′ S latitude, 29°10′–29° 40'E longitude (Figure 1). Entirely, the upper Nyabarongo catchment is dominated by six types of land use classes, namely forestland, cropland, built up land, grassland, wetland and water bodies.

Topographically, the upper Nyabarongo catchment is a landslide-prone zone which presents an appropriate individuality to explore and make comparison of the machine learning simulations in predicting susceptibility. Landslides have recently become the highest frequent and devastating hazard in the area under investigation (MIDIMAR 2016, 2018). They therefore induce massive devastations and fatalities. Also, as reported by Rwandan Government through MINEMA, from 2011 to May 2015, 124 human lives were lost due to natural hazards including landslides, with injuries and about 897 houses completely demolished (MIDIMAR 2018). In 2016 (May alone), landslides killed 35 people, and 26 were injured while 67 road segments and 29 bridges were completely destroyed (Nsengiyumva et al. 2018). From January to December 2018, landslides and other rainfall-induced hazards took about 234 lives with 218 injuries, demolished 15,264 family houses and 9,412 of crops in hectares, damaged infrastructure facilities (31 and 52 facilities for road sections and bridges respectively), completely destroyed 87 school rooms and killed 797 livestock (MIDIMAR 2018). Most of these damages were recorded in Nyabarongo vicinity. This background information about the physical settings of the upper Nyabarongo catchment zone makes it an ideal case study. Despite the high frequency, intensity and magnitude of landslide hazards, no comparative analysis using novel simulation techniques (LMT, RF and NBT) has previously been done to predict the spatial likelihood of landslides in the upper Nyabarongo-Catchment of Rwanda. Therefore, this background information confirms the rationale for authors to select the area for simulating the three models.

3. Datasets and methodology

3.1. The inventory map

It was confirmed by previous studies that a landslide inventory is indispensable for susceptibility modelling (Corominas et al. 2014; Pham et al. 2017; Van Tien et al. 2018). It is therefore made of a compilation of locations where landslides occurred in past and their features. This comprises of areas of previous and current landslides and can display locations, time of occurrence, landslide types, frequency and intensity

of occurrence, scale and extent, mechanisms of failure, causal-related factors, damages and effects induced (Calvello and Pecoraro 2018; Chen et al. 2019). Generally, it is assumed that circumstances that caused past landslides, may be the same to trigger future landslide occurrences (Nsengiyumva et al. 2018). Thus, past landslide locations help to identify and recognize the correlation between historical landslide events and predisposing factors (Chen et al. 2018a; Chen et al. 2018b; Nsengiyumva et al. 2018; Youssef et al. 2016).

In this study, information about locations of past landslides and non-landslide locations was collected from different sources including both primary and secondary data sources. These included reports from ministries and other agencies, databases, websites (www.minema.gov.rw) and extensive field investigation from February to December 2019 in the catchment zone to validate landslide locations. Therefore, a total of 196 past landslide events (1997-1999), were mapped as points (x, y coordinates) within the upper Nyabarongo Catchment of Rwanda (Figure 1) using the GPS devices. These landslides were characterized by an average length of about 67 m. Moreover, their extent was ranging between 14 and 7089 m^2, with an average extension of about 473 m^2. The slide surface depth ranged between 0.90 m and 1 m. The slide surface depths vary according to the location of the landslides, characteristics of the soil, land use/cover types and other anthropogenic aspects.

Through field visits, the authors have also detected that big part of past landslides in Nyabarongo catchment occurred on embankments alongside roads and in cut-slopes, within poorly cultivated and uncovered lands among others (Figure 2e). The area is mostly affected by both translational and rotational landslides as well as shallow landslides involving soil at different depths. To complement the field datasets and other collected information, authors have conducted few interviews with local residents and indigenous knowledge holders within the hazard-prone zones in the area under investigation.

The authors have randomly split the inventory map into training dataset for model building and validation/testing dataset for model performance validation. As confirmed by Dou et al. (2019) the model performance is validated by splitting the dataset into two parts. Nonetheless, there is no universal rule for selecting the ratio of testing and training dataset (Pradhan and Lee 2010). The current study employed a random proportion of 75% landslide locations (equivalent to 147 landslide locations) to build the models whereas 25% (49 landslide events) were used for model performance evaluation and susceptibility map validation (Figure 1g). Equally, random non-landslide sites were collected within the study area to derive the three final susceptibility maps. The inventory was extracted and split into training and validating points using GIS software environment, ArcMap 10.3 (Youssef et al. 2016; Zêzere et al. 2017).

3.2. Landslide conditioning factors

For any susceptibility study, it is mandatory to consider predicting factors which may be natural or man-made. The factors normally inform what may have led to slope instability in the past. Studies confirmed that landslides can be triggered by similar causes that lead to instability in the past (Abella and Van Westen 2007).

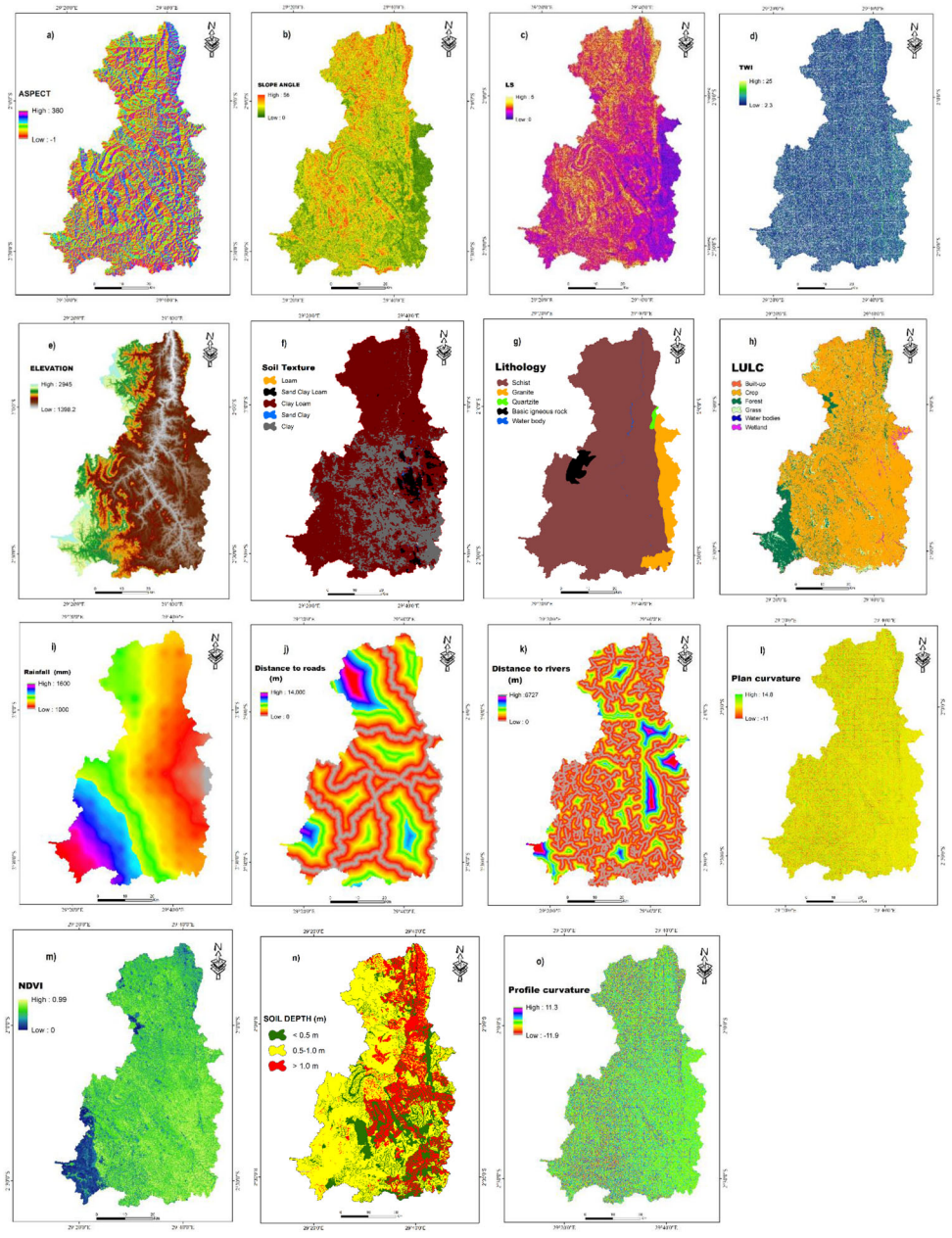

Figure 2. Landslide conditioning factors: (a) Aspect; (b) Slope angle; (c) LS, (d) TWI, (e) elevation, (f) Soil texture, (g) Lithology, (h) LULC, (i) Rainfall, (j) Distance to roads, (k) Distance to rivers/streams, (l) Plan curvature, (m) NDVI, (n) Soil depth, (o) Profile curvature.

For the current study, fifteen landslide factors were selected and employed to predict susceptibility namely the normalized difference vegetation index (NDVI), slope angle, distance to roads, slope aspect, elevation, profile curvature, plan curvature, soil depth, lithology, soil texture, land use/land cover (LULC), distance to rivers, wetness index of the topography (TWI), topographic factor (LS) and precipitation. These were

Table 1. Summarized spatial database for susceptibility modeling.

No.	Dataset/Factors	Data Source	Spatial resolution/ Desription	Data structure/ format
1	Landslide inventory map	Fieldwork in Rwanda in 2018 (x,y coordinate points), Secondary data sources	1.400.000 scale	Vector dataset
2	Shuttle Radar Topography Mission (STRM) digital elevation model (DEM)	United States Geological Survey Earth Explorer: http://earthexplorer.usgs.gov/	30 × 30m	Raster
3	Rwanda Land cover land use (LCLU) Normalized Difference Vegetation Index (NDVI)	Regional Centre for Mapping of Resources for Development (RCMRD): http://apps.rcmrd.org/landcoverviewer/ Landsat-8 OLi images provided by the United States Geological Survey (USGS)	30 × 30m	Raster
4	Lithology	Geological map of Rwanda (Rwanda Ministry of Environment): www.moe.gov.rw	30 × 30m	Raster
5	Soil datasets (Soil texture, Soil depth)	Rwanda Ministry of Agriculture (MINAGRI: www.minagri.gov.rw), Rwanda Agriculture Board (RAB: http://www.rab.gov.rw)	30 × 30m	Raster
6	Road network datasets (Distance to roads)	Rwanda Transport Development Agency/ Ministry of Infrastructure (http://www.rtda.gov.rw)	30 × 30m	Raster
7	River/stream networks (Distance to rivers)	Rwanda Ministry of environment (www.moe.gov.rw)	30 × 30	Raster
8	Precipitation datasets (mean annual precipitation: mm/year)	Rwanda meteorological Agency Meteorological stations data (meteorwanda.gov.rw)	Mean annual rainfall for 20 years (1998–2018)	Raster
9	Catchment boundaries/shapefiles	Rwanda Ministry of Environment (www.moe.gov.rw) Rwanda Environment Management Authority (www.rema.rw)	1.400.000 Scale	Vector
10	Study area Shapefiles/ boundaries (updated shapefiles of 2015)	Ministry of Environment, Rwanda Water and Forestry Authority (www.rwfa.gov.rw) Rwanda Land Management and Use Authority (www.rlma.rw)	1.400.000 scale	Vector

regarded as causal factors of landslide occurrence in this case study and will be defined in detail in the following section. The literature confirms that no universal rule is followed in selecting landslide conditioning factors for susceptibility modelling (Chen et al. 2017d; Pham et al. 2017).

To predict susceptibility using machine learning approaches, authors have to identify the actual predisposing factors that might have contributed to slope instability in the area (Nsengiyumva et al. 2018). This is therefore significant since it can help to

generate reasonable results (Guzzetti et al. 2000). The selection of conditioning factors (Table 1) was based on data availability, landslide categories, modelling methods and objectives of the study. Moreover, the Rwanda risk management plan, disaster management policy, risk atlas, field investigations and the previous studies on landslide (MIDIMAR 2015) were referred to in order to deduce the fifteen conditioning factors used in this study.

The 30 m spatial resolution STRM DEM was obtained from the United States Geological Survey (USGS), using ArcMap 10.3 of GIS environment (Maes et al. 2018). From this dataset, seven factors were produced including elevation, TWI, plan curvature, slope angle, aspect, profile curvature and the LS. Elevation is confirmed very important for susceptibility modelling since it indicates the deviations of heights in maximum and minimum terrains (Chen et al. 2019). It has strong correlation with landslide occurrence as it impacts on the topographic nature, temperature and vegetation structures, moisture and anthropogenic activities (Chen et al. 2019). All these conditions are, therefore, linked with the stability of the slope in line with susceptibility. Slope angle and slope aspect were considered in this study as they are important and common contributing factors to the occurrence of hazards. Consequently, for the present study area, no historical landslide events were recorded in the very low slope angles (Figure 2b). For slope aspect, it indicates the moisture of the topography based on precipitation patterns and solar radiation (Pham et al. 2018). Thus, it was confirmed by previous studies that slope category influences characteristics of slope failure (Chen et al. 2017b). Furthermore, curvature was used due the fact that it controls the water flows which influence the occurrence of landslide hazards within the area (Chen et al. 2018a). The literature confirms therefore that concave slopes are more unstable than convex slopes (Montrasio et al. 2012; Pourghasemi et al. 2018)

Generally, the curvature is considered for susceptibility prediction due to its values which denote various erosivity levels of water, runoff settings and topographical features of the area (Dou et al. 2019). For the present study both profile and plan curvatures were used to derive susceptibility maps for the upper Nyabarongo Catchment in Rwanda (Figure 2l and o). Additionally, the TWI which is also considered as a widely-applied landslide causal-related factor was used for this study (Figure 2d). The TWI was calculated using Equation 1 as follows:

$$TWI = \ln\left(\frac{\alpha}{\tan\beta}\right) \tag{1}$$

Where β = the slope angle (radiant) and α denotes the flow accumulation towards a point. Thus, the values of TWI for the Nyabarongo were computed using GIS software environment, ArcMap 10.3. The LS factor describes the length and steepness the slope in a given area (Amanambu et al. 2019). LS can explain the impact of topography on the occurrence of landslides. This factor has been considered for susceptibility maps' generation for this study. It is therefore expressed using Equations 2, 2a, 2 b and 3:

$$L_{i.j} = \frac{(A_{i.j-in} + D^2)^{m+1} - A_{i.j-in}^{m+1}}{D^{m+2}.X_{i.j}^{m}.(22.13)^{m}} \tag{2}$$

$$m = \frac{\beta}{1+\beta} \tag{2a}$$

$$\beta = \frac{\sin\theta/0.0896}{3(\sin\theta)^{0.8} + 0.56} \tag{2b}$$

$$S_{i.j} = \begin{cases} 10.8\sin\theta_{i.j} + 0.03, & \tan\theta_{i.j} < 9\% \\ 16.8\sin\theta_{i.j} - 0.50, & \tan\theta_{i.j} \geq 9\% \end{cases} \tag{3}$$

With $L_{i.j}$ denoting the length factor of the slope for the grid cell (i.j); D stands for the size of the grid-cell (m); $X_{i.j}$ is given by ($\sin a_{i.j} + \cos a_{i.j}$); $a_{i.j}$ means the direction of aspect for the grid-cell (i.j); also $A_{i.j-in}$ stands for the flow accumulation at the inlet (m^2) of a grid (i.j). Additionally, the length of the slope m is related to β ratio of rill erosion to interrill erosion; and θ means the slope in degrees (Amanambu et al. 2019).

LULC was proved a very imperative susceptibility modelling factor (Pisano et al. 2017). Studies have confirmed that land use is highly correlated with mass wastes (Guzzetti et al. 2000; Persichillo et al. 2017). To apply this factor for the present study, the authors produced the updated LULC map of 2017 with 30 m spatial resolution (Figure 2h). This was derived using datasets obtained from landsat-8 OLI (U.S.G.S.) through the global visualization toolset. To achieve this, the authors used the Envi 5.3 software environment and the likelihood (maximum) classification method was applied. After radiometric adjustments and all other necessary corrections, the authors classified the LULC map based on the prior RCMRD classification for the central-eastern Africa region. Furthermore, the current study applied the U.S.G.S. method, type one for the classification. Thus, the area was categorized into six land cover/land use types (Figure 2h). Additionally, the accuracy assessment was conducted using sixty points randomly selected for each land use type from the ground reference data. These were then overlaid to a classified map image for verification and validation. Overall, the suitable accuracy of 91.6% was reached for the study area. The derived LULC map revealed that the cropland class occupies about 67.9% of the study area (Figure 2h).

Moreover, lithology plays a big role in analysing the slope stability. In many cases, landslides occur within lithological zones with lowest strength and higher moisture content (Chen et al. 2017a). The lithology factor was derived from the available geology map of Rwanda (1:100,000) with 30 m spatial resolution (RNRA 2015). The soil data were acquired from the Rwanda Ministry of Agriculture (MINAGRI 1995; RAB 2000). These datasets were generated from 1995 and 2000 national soil studies (Hengl et al. 2015). The datasets were used to derive the two conditioning factors used for this study namely soil texture and soil depth (Figure 2n). For rainfall conditioning factor, the authors utilized mean annual rainfall map for 20 years (1998-2018) produced from meteorological data of the investigated area.

These data were provided by the Rwanda Meteorological Agency (meteorwanda.gov.rw). It is largely confirmed that the likelihood of a landslide to happen largely depends on heavy or prolonged rainfall (Chen et al. 2017c). This is a fundamental

landslide trigger based on its capability to raise the levels of ground water as well as water pore pressure increases (Ding et al. 2016).

The NDVI factor was used for this study (Figure 2m, Table 1). This has become very prevalent in landslide susceptibility mapping studies across the globe (Chen et al. 2017d). NDVI measures the vegetation level in the study under investigation. Authors used landsat-8 to extract NDVI factor as expressed by Equation 4:

$$\mathrm{NDVI} = \frac{(\mathrm{NIR}-\mathrm{R})}{(\mathrm{NIR}+\mathrm{R})} \tag{4}$$

where NIR = the near infrared band, R = the red band within the electromagnetic spectrum. Therefore, NDVI values ranges between −1 and 1 with positive values representing vegetated ground. For this study, distance to roads and distance to rivers were considered as landslide factors for this study (Figure 2j and k; Table 1).

3.3. Landslide predicting factors selection

It is very important to use appropriate techniques in selecting proper factors for susceptibility modelling. As confirmed by studies (Chen et al. 2018a; Zêzere et al. 2017), conditioning factors selection is a very useful phase that helps to avoid noisy factors that may cause the model confusion. The quality of the final maps is not only dependent of employed models but also on the qualitative status of used datasets (Van Westen et al. 2013). This step increases the performance and predictive fitness of the models. Thus, different techniques were introduced in the literature for conditioning factors selection. They include information gain ratio, consistency, gain ratio, chi-square statistics and others (Pham et al. 2017). For the current research, the information gain (IG) was used to assess the capability of the predicting factors. Thus, the selection of predicting factors reduces inappropriate and useless input datasets to increase the modelling accuracy (Chen et al. 2018a).

The IG has been extensively applied for various researches (Bui et al. 2017; Chen et al. 2018b). The information gain value for any conditioning factor is determined using Equations 5 and 6 below:

$$\mathrm{IG}(\mathrm{Y}, \mathrm{X_i}) = \mathrm{H}(\mathrm{Y}) - \mathrm{H}(\mathrm{Y}|\mathrm{X_i}) \tag{5}$$

with H(Y) standing for the value of entropy for Yi whereas H (Y|Xi) = the Y entropy after relating the landslide conditioning factor values (Bui et al. 2016).

$$\mathrm{Info}([\mathrm{P_1}, \mathrm{P_2}, \ldots . \mathrm{P_n}]) = \mathrm{entropy}(\mathrm{P_1}, \mathrm{P_2}, \ldots . \mathrm{P_n}) \tag{6}$$

where P1, P2 … $\mathrm{P_n}$ represent the instance numbers for each factor's class, and the value of $\mathrm{P_i}$ is given by the division of its value by the sum of all $\mathrm{P_i}$. Generally, the landslide factors are not equally correlated to the landslides (Table 2).

Table 2. Significance of landslide predisposing factors using IG method.

No.	Conditioning factor	Average merit (AM)	Standard Deviation (SD)
1	Land use land cover	0.106	±0.011
2	Slope angle	0.088	±0.006
3	Elevation	0.055	±0.008
4	Distance to roads	0.051	±0.008
5	Rainfall	0.039	±0.004
6	Aspect	0.023	±0.012
7	Lithology	0.016	±0.009
8	TWI	0.011	±0.005
9	Soil texture	0.009	±0.013
10	NDVI	0.007	±0.009
11	LS	0.004	±0.007
12	Distance to rivers	0.003	±0.013
13	Soil depth	0.001	±0.015
14	Plan Curvature	0.000	±0.000
15	Profile curvature	0.000	±0.000

3.4. Landslide susceptibility modeling

Various approaches exist to map landslide susceptibility (Chen et al. 2017a, b; Chen et al. 2018a). Only three of these methods (LMT, RF, and NBT) have been selected because they were suitable for the present study area. Therefore, the authors applied the conditioning factors and the three models for this study based on the data availability, objectives of the study, landslide type and on the study area scale. Briefly, the comparison of the three established models for a landslide-prone area increased the knowledge on GIS-based machine learning techniques in predicting landslides for African continent as whole and for the study area in particular.

3.4.1. Logistic model tree (LMT)

The LMT is a landslide susceptibility mapping model that has been extensively applied for susceptibility studies (Chen et al. 2019; Dou et al. 2019). This model combines a linear logistic regression with a decision tree in order to leverage their advantages (Chen et al. 2018b; Karabulut and Ibrikci 2014). Thus, at any given tree node, authors have to employ the LogitBoost algorithm to fit functions of the logistic regression (Chen et al. 2018a; Karabulut and Ibrikci 2014). For landslide susceptibility modelling, it is advisable to determine the probability for each class of the conditioning factor using Equation 7. This is therefore done based on the principle that there are x vectors and C classes in each susceptibility dataset (Karabulut and Ibrikci 2014):

$$p(c|x) = \frac{e^{F_{c(x)}}}{\sum_{n=1}^{c} e^{\left(F_{n(x)}\right)}} \tag{7}$$

Where Fc(x) denotes the linear regressions functions, C = the number of classes. Therefore, the fitness of Fc(x) is performed by applying the least squares technique. Moreover, the total sum of Fc(x) for all classes must be equal to 0 as expressed with Equation 8:

$$\sum_{n=1}^{c} F_c(x) = 0 \tag{8}$$

The logistic model tree uses the functions of logistic regression to estimate the probability value for each class of the conditioning factor (Karabulut and Ibrikci 2014). It is therefore a probability model capable of handling uncertainties. To estimate the fitness using LMT, the LogitBoost applies maximum likelihood for the determination of the least possible deviations between both observed and predicted values (Chen et al. 2018a).

3.4.2. Random Forest model (RF)

The RF model consists of an ensemble learning approach that associates different decision trees for landslide to spatially predict susceptibility for a given area (Ayala-Izurieta et al. 2017; Dou et al. 2019). As previously stated by studies on landslide mapping, the RF method can be characterized as a collection of decision and random trees (Chen et al. 2018a). Each tree is dependent on the values of random vectors equally distributed among all forest's trees.

For landslide susceptibility mapping, each node of normal trees can be split using the perfect split for all landslide predicting factors (Chen et al. 2017d). However, for the RF model, every node is divided by using the best split in a subset of factors selected randomly by the node. Therefore, based on the RF algorithm, the smaller the value, the better the split for the node in susceptibility modelling (Kausar and Majid 2016). Naturally, a random vector i_k within the RF algorithm, is independently produced from the prior random vectors across all the trees whereby every tree is generated by random vector i_k and training datasets. The outcomes of this method are represented by the groups of tree classifiers h (x, i_k), k = 1, 2, … n at input x vector (Chen et al. 2019). For the present study, i_k represents the conditioning factors for susceptibility simulation. The RF entailed two categories of trees including both non-landslide and landslide, and each of them was established from fifteen random features.

Basically, the generalization error (GE) is defined using Equation 9 in a RF algorithm (Chen et al. 2019; Kausar and Majid 2016)

$$GE = P_{x,\, y(mg(x,\, y)<0)} \tag{9}$$

whereby x,y denote the conditioning factors and P_{xy} indicates the probability over both x and y space, while mg = the margin function as expressed using Formula 10:

$$mg(x, y) = av_k|(h_k(x) = y) - max_{j \neq y^{av}k}|(h_k(x) = j) \tag{10}$$

At a random vector, the margin function (mg) determines to which extent the number of votes exceeds the average. Thus, I(*)= the indicator function (Chen et al. 2019).

The Random Forest method has been a predominant approach for determining suitable but unseen patterns among huge datasets (Chen et al. 2018a).

3.4.3. Naïve-Bayes tree model (NBT)

The Naïve-Bayes tree model (NBT) is considered a hybrid algorithm (Chen et al. 2018a). On each tree's leaf node, NBT consists of decision tree classifiers and Naïve-Bayes. Furthermore, NBtree is a generative classifier for susceptibility prediction (Tsangaratos and Ilia 2016). The NBT was introduced in 1996 and has become one of the commonly applied machine learning approaches. It is mostly built on a tree classification like in hierarchy (Pham et al. 2017; Pham et al. 2016).

Generally, Bayesian classification consists of a procedure of estimating the new observation probability that belongs to a predefined category, by using a probability-based method (Chen et al. 2019). This method is built on Bayes' theory that takes all attributes as independent to maximize the posterior probability for determining the classification (Pham et al. 2018). Furthermore, the NBT is considered as a classification tree method, but it comprises both leaves and nodes. The previous studies confirmed that the performance of NBT is far better than Naïve Bayes and decision tree (Dou et al. 2019).

The probability is therefore expressed by Equation 11:

$$P(C_j|X) = \frac{P(X|C_J) * P(C_j)}{P(X)} \tag{11}$$

$p(C_j|X)$ means the probability of the observations that are unkown while X belongs to C_j category which is known as the posteriori probability; also, p(X|Cj) = the category Cj given probability, and observation that are unknown belong to this category, p(Cj) denotes the prior probability the unknown observation X to be observed in category C_j, p(X) represents the prior probability of the unknown observation and X is the same for each category C_j.

For k landslide related variables, y_j stands for the analysis of the Boolean output for susceptibility analysis, and describes both landslide and non-landslide prediction. Tien Bui et al. (2016) ascertained that Equation 12 may be applicable to decide and choose the class with maximum posterior probability.

$$Y_j = \text{argmax}P(Y_j)\prod_{i=1}^{k} P\left(\frac{X_i}{Y_j}\right) \tag{12}$$

Where $P(y_j)$ stands for y_i prior probability which may be calculated following the ratio of the observations with y_j class output in the training datasets, j = the non-landslide or landslide for susceptibility modelling while k = the overall number of cases. Besides, $P(x_i/y_j)$ denotes the conditional probability which is computed using Equation 13.

$$P\left(X_i|Y_j\right) = \frac{1}{\sqrt{2\pi\alpha}} e^{\frac{-(x_i-\eta)^2}{2\alpha^2}} \tag{13}$$

Where η = the mean and α represents the standard deviation of x_i.

Generally, for the mapping of landslide susceptibility using NBT model, a tree growing follows the selection of the attribute measure following the concept of

entropy which is taken as the degree of disorder (Chen et al. 2017a). Assume that given cases are represented by D and |D| is the total of all the cases. The cases can therefore be categorized into m classes whereby: D_i (i = 1, 2, … .m). Thus, |Di| = the number of the cases belonging to the D_i class. To calculate the expected entropy for D classification, Formula 14 is applied:

$$\text{Entropy}(D) = -\sum_{i=1}^{m}(|D_i|/|D|)\log_2[(|D_i|/|D|)] \tag{14}$$

with the partitioning of D set on attribute A (having z number of values), the expected entropy is summarized with Equation 15:

$$\text{Entropy}_A(D) = -\sum_{j=1}^{z}\frac{|D_j|}{|D|} * \text{Info}(D_j) \tag{15}$$

At this stage, the difference between Entropy (D) and Entropy A (D) is considered as the Information Gain (InforGain), and its value helps to determine the split using Equation 16:

$$\text{InfoGain}(A) = \text{Entropy}(D) - \text{Entropy}_A(D) \tag{16}$$

Nonetheless, the InfoGain may cause biases for attributes with many values and the related number of splits may be not reasonable (Chen et al. 2017b). To avoid the bias, the authors are therefore obliged to use SplitInfo in decision tree to normalize the InfoGain (Pham et al. 2016). The SplitInfo is therefore computed using Equation 17 below:

$$\text{SplitInfo} = -\sum_{j=1}^{z}\frac{|D_j|}{|D|} * \log_2\left(\frac{|D_j|}{|D|}\right) \tag{17}$$

The SplitInfo is a type of Entropy related to the split point of a given attribute. From this, the Information Gain Ratio in decision tree is therefore defined using Equation 18 as follows:

$$\text{Gain Ratio}(A) = \frac{\text{InfoGain}(A)}{\text{SplitInfo}(A)} \tag{18}$$

3.4.4. Model-performance evaluation and validation

As confirmed by studies on landslide (Chen et al. 2017c; Chen et al. 2018b), a final derived map is not suitable unless it is validated. It is required to use appropriate approaches for susceptibility map validation (Van Den Eeckhaut et al. 2005). For the present study, the authors used the receiver operating characteristic (ROC) and other statistical estimators namely the accuracy, precision and the root mean square error (RMSE) to evaluate the three produced susceptibility maps. ROC portrays the

percentages of true positive against the false negative percentages to rate the past landslides cumulatively in a decreasing order. This helps to find the success rates using the areas under ROC curves (AUROC) (Ahmed and Dewan 2017). The AUROC is used for detecting the models predictive capabilities.

In case of the poor prediction (poor modelling), the AUROC values become smaller or equal to 50 whereas the better prediction is obtained for AUC values closer to 100 (Ahmed and Dewan 2017). Furthermore, the higher the AUROC value, the better the model's performance, and an AUROC value of 100 depicts an excellent and outstanding performance (Bui et al. 2017). The AUROC curves are regarded as one of the best and common techniques for validating and comparing models in recent studies (Begueria 2006; Chen et al. 2018a; Zêzere et al. 2017). Thus, it has been extensively applied as a common tool to assess the performance capability of the models (Chen et al. 2017d). To assess and compare the three susceptibility methods, the authors applied statistical estimators using Equations 19, 20, 21 and 22 (Chen et al. 2018b; Tharwat 2018):

$$\mathrm{AUC} = \frac{(\sum \mathrm{TP} + \sum \mathrm{TN})}{(\mathrm{P} + \mathrm{N})} \tag{19}$$

$$\mathrm{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \tag{20}$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{21}$$

Whereby P represents the landslide number, N stands for the non-landslide number. Both TP (which is the true positive) and TN (which means the true negative) represent the numbers of correctly classified pixels; and both FP (which is the false positive) and FN (which is the false negative) portray the numbers of incorrectly classified pixels (Chen et al. 2017a). Therefore, for AUC, accuracy, and precision, a better predictive ability of the model is shown by a higher value (Tharwat 2018). Furthermore, when the obtained value is closer to 1, the derived susceptibility map is confirmed accurate and reliable. For the current study, RMSE value is defined as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}} \sum (\mathrm{y} - \mathrm{y}')^2} \tag{22}$$

where N = total number of data, y is the observed output, y'= the predicted output. Therefore, the closeness of RMSE value to 0, indicates the competency of the model to forecast susceptibility (Ercanoglu and Gokceoglu 2004; Nefeslioglu et al. 2008).

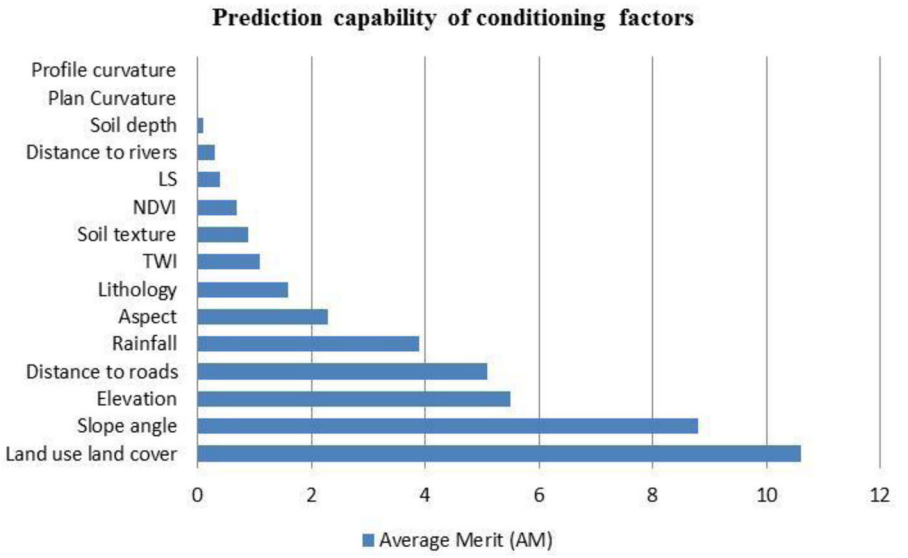

Figure 3. Prediction capability of the fifteen landslide conditioning factors in the present study.

4. Results and discussion

4.1. Landslide conditioning factor analysis

To generate the landslide susceptibility maps for the upper Nyabarongo catchment (Rwanda), authors have applied IG technique to select conditioning factors. The factor's selection based on their weights to fit with the models for susceptibility mapping (Figure 3). Thus, factors with weights higher than zero were considered for the susceptibility analysis. In contrast, factors with less weight (below or equal to zero value) were excluded from susceptibility modelling.

The capability of 15 factors using IG method is illustrated in Figure 3 and Table 2. The analysis disclosed that only 13 predicting factors have positive correlation with the landslide hazard spatial occurrences in Nyabarongo based on their positive values (Average Merit >0). Therefore, from these factors, land use/land cover becomes the peak for landslide prediction capability (AM = 0.106). This can be explained by many past locations of landslide events recognized within the Nyabarongo Catchment due to land cover settings and inappropriate land use practices (Figure 4). It is therefore in conformity with other previous research works (MIDIMAR 2015; Nsengiyumva et al. 2018). The next factor was slope angle which has also the best correlation with landslides in this research (AM = 0.088). Slope was confirmed to be among significant predicting factors of landslides (Ahmed and Dewan 2017). For other factors including elevation (AM = 0.055), distance to roads (AM = 0.051), rainfall (AM = 0.039), aspect (AM = 0.023), lithology (AM = 0.016), TWI (AM = 0.011), soil texture (AM = 0.009), NDVI (AM = 0.007), LS (AM = 0.004), distance to rivers (AM = 0.003), soil depth (AM = 0.001) respectively showed significant spatial relationship to the landslide susceptibility modelling.

Similar factors have been largely used in various studies related to susceptibility modelling (Chen et al. 2019; Nsengiyumva et al. 2018; Pham et al. 2018). However,

Figure 4. Fresh Landslide occurrences triggered by heavy rainfall in the study area (**Source**: Field visits by the researchers, June 2018-August 2019).

the other two conditioning factors namely plan and profile curvatures have no positive correlation since they are tested as low or null prediction capability (Average Merit = 0). Thus, both factors were not considered for the current susceptibility modelling to achieve improved accuracy of the final output (Bui et al. 2017; Bui et al. 2016). Recent studies confirmed that prediction ability of a given factor largely depends on the used landslide model (Chen et al. 2017b; Zêzere et al. 2017). Nevertheless, further studies may be useful to analyse the appropriate approaches for selecting predicting factors and more improvement of their predictive capability.

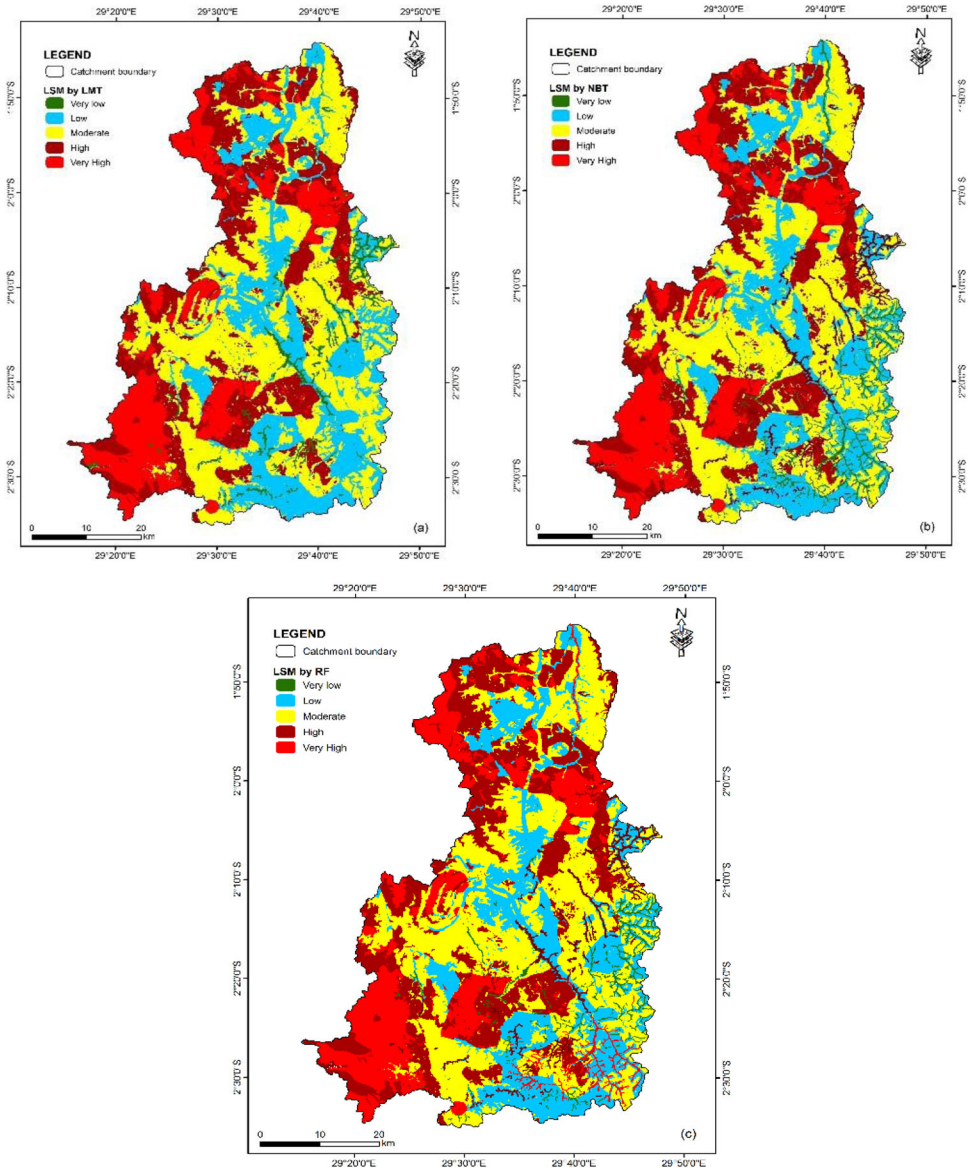

Figure 5. Landslide susceptibility maps derived using: (a) LMT, (b) NBT and (C) RF models.

4.2. Generation of landslide susceptibility maps

For this study, the three models (NBT, LMT and RF) were used to study susceptibility in the upper Nyabarongo catchment of Rwanda (Figure 5). The susceptibility models were built employing the abovementioned data (training datasets) together with testing and validating datasets. Thus, the three models were applied in determining the susceptibility indexes for all pixels in the area under investigation. The susceptibility maps were therefore derived with five classes. Besides, the natural breaks method was employed to make classification of the three derived maps (very low,

Table 3. Different landslide susceptibility classes in percentages using RF, LMT and NBT models.

Susceptibility class	RF (%)	LMT (%)	NBT (%)
Very low	3.2	3.63	4.18
Low	24.12	25.13	25.21
Moderate	32.61	33.10	30.68
High	21.05	22.02	22.70
Very high	19.02	16.12	17.23
Total	100	100	100

low, moderate, high and very categories) (Figure 5). Moreover, the natural break has become the most widely applied method in classifying susceptibility maps (Bui et al. 2016; Carrara et al. 1999). It is therefore confirmed one of the best appropriate techniques for modelling susceptibility based on the data distribution histogram (Chen et al. 2017d).

Overall, the analysis generated the three landslide susceptibility maps for the study area (Figure 5) and it can be seen that the five susceptibility classes are differently dispersed across the entire catchment zone (Table 3). The distribution of landslide susceptibility across the entire study area, confirms unconditionally that the upper Nyabarongo catchment of Rwanda is highly susceptible to landslides. For the RF model, the findings reveal that the category of very low susceptibility falls into 3.20% of the total catchment and 24.12% falls into the low susceptibility class. The categories from moderate, high and very high susceptibility represent 32.61%, 21.05% and 19.02% of the area respectively (Figure 5, Table 3).

For susceptibility map derived using LMT, 3.63% of the upper Nyabarongo catchment falls into very low susceptibility category and 25.13% belongs to low susceptibility category. Moreover, the categories of moderate, high and very high susceptibility represent 33.10%, 22.02% and 16.12% of the entire catchment, respectively. Regarding the susceptibility map derived by NBT model, it can be detected that very low and low susceptibility classes account for 4.18% and 25.21% of the upper Nyabarongo catchment, respectively. 30.68% of the upper Nyabarongo catchment area falls within the moderate susceptibility category. Moreover, 22.70% and 17.23% of the area fall into the high and very high susceptibility categories, respectively (Table 3).

Overall, 17.46% of the total area under study falls into the very highly susceptible class and 22.02% falls into the high susceptibility category, while 3.67% represents the very low category/stable zone. These results confirm that the upper Nyabarongo catchment area is very prone given its geomorphological settings, high presence of landslide influencing factors as well as different anthropogenic factors. Obviously it is not excluded that the high presence of landslides in this area, could be influenced by different causes, including anthropogenic, geological, climatological and environmental aspects that have not been directly taken into account in the selected models (Bui et al. 2017; Zêzere et al. 2017). This catchment qualifies to make susceptibility modelling achievable. As shown by previous studies (Chen et al. 2018b; Persichillo et al. 2017), susceptibility mapping has to be conducted in order to deal with the prevailing natural hazards. Susceptibility mapping is, therefore, a critical stage within the landslide risk management cycle. This helps to identify high-risk zones and predominant causal-related factors. The selected methods used the GIS and remote sensing (RS)

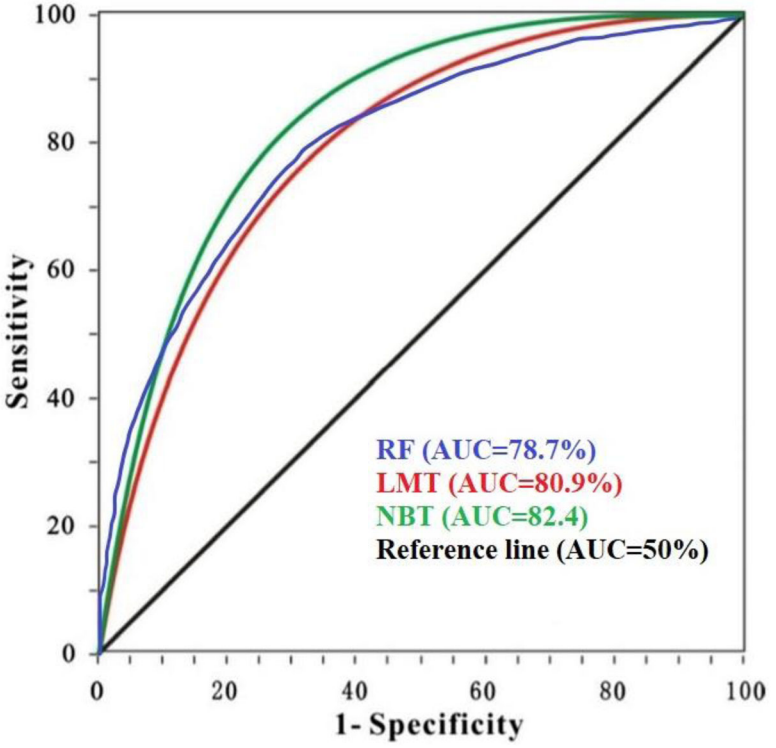

Figure 6. The performance of landslide models (RF, LMT and NBT) using ROC curves.

techniques. From Table 3, it can be observed that big part of the upper Nyabarongo catchment falls into moderate susceptible area for all the tree models.

The results from the models helped to respond to the critical scientific questions of making judgments on which method is appropriate to successfully predict susceptibility within the area under study. The landslide hazard situation in the Nyabarongo catchment of Rwanda is mostly aggravated by some human activities including unplanned and poor settlements, improper land use practices, lack of storm water drainage, absence of rainwater harvesting mechanisms, high level of vulnerability and exposure to landslides. Furthermore,due to the accelerated population growth, there is a high rise in pressure on land, whereby local residents continue to invade the fragile mountainous environment and ecosystem, cut the hills to settle and earn their living.

Analytically, the northern and western parts of the modelled area proved to be highly susceptible to landslides on the three derived susceptibility maps (Figures 4 and 5). In contrast, the very low susceptible areas were mostly detected in the southeastern part of the catchment and this is due to a number of reasons including the topographic nature and presence of the conditioning factors (Figure 5). Further studies may be proposed to complement these findings using other novel machine learning and ensemble techniques and such researches would generate more reasonable results in line with landslide risk mitigation and sustainable environmental management in Rwanda.

Table 4. Model performance evaluation with statistical estimators.

	Susceptibility models		
Statistical estimator	RF	NBT	LMT
Accuracy	0.733	0.799	0.762
Precision	0.692	0.745	0.724
RMSE	0.428	0.301	0.364

4.3. Models comparison and validation

As previously explained, the AUROC curves and the three statistical measures (accuracy, precision and RMSE) were applied to assess the complete performance of the three approaches and to validate the derived maps (Figure 6, Table 4).

From the findings of this research, it was disclosed that all the three methods reasonably produced accurate landslide susceptibility maps. Therefore, it can be observed from the analysis of the AUROC (Figure 6) that the susceptibility prediction rates were 79.8%, 80.6% and 81.2% respectively for RF, LMT and NBT models. Therefore, the NBT and LMT susceptibility models attained very good performance in modelling landslide susceptibility within the study area with AUROC $\geq$ 80% (Figure 6) (Yilmaz and Ercanoglu 2019). Moreover, the results analysis confirmed the NBT model to be the best predictor of landslide susceptibility within the upper Nyabarongo catchment zone. It has outperformed the RF and LMT models. However, it is very clear that all the models produced reasonable outputs and they proved promising methods for susceptibility modelling within the area under investigation.

Furthermore, three statistical estimators (Accuracy, precision and RMSE) were used to compare and evaluate the models (Table 4). The results disclosed that the NBT method achieved the highest performance for accuracy and precision values (0.799 and 0.745 respectively), followed by the LMT model (Accuracy = 0.762 and precision = 0.724). The RF model achieved the lowest accuracy and precision values (0.733 and 0.692 respectively). For RMSE, the NBT model also achieved the best performance with 0.301 of RMSE values, followed by LMT model (0.364) and RF model (0.428 RMSE value). Generally, the NBT model performs better than LMT and RF models for the accuracy, precision and RMSE. Despite this slight discrepancy, the three used models produced reasonable results and proved suitable for susceptibility mapping in the landslide prone-areas. Moreover, the overall validation of the findings indicated a sensible agreement between the derived maps and the observed data on past landslide locations.

The current study offers a great contribution to the knowledge of landslide susceptibility prediction, mainly for the mountainous zones of the central-eastern African region, and specifically for Rwanda (Nahayo et al. 2019). Through susceptibility mapping, it is appropriately used for land management and policy, there is a high possibility of mitigating landslide hazards to turn into disasters. However, there is a serious shortage of landslide database and past landslide records for Nyabarongo catchment, and this gap should be addressed through the adoption of adequate and regular system for recording landslide events to serve as a database for future inventory building and landside quantitative modelling. This study also represents a novel contribution of using GIS-based machine learning techniques for the Eastern African

region as a whole and for the study area in particular. Furthermore, scholars, decision and policy makers would deploy considerable efforts for mitigation and preventive measures especially for areas modelled as highly and very highly susceptible. The study results may be informative and instrumental for sustainable land use planning, rational environment management and resilience building.

5. Conclusion and policy implications

For this study, the authors employed three different GIS-based machine learning methods to map landslide susceptibility for the upper Nyabarongo catchment in Rwanda. The RF, NBT and LMT models were employed and explored in various susceptibility studies, but their exploration through comparative analysis has never been done before for the whole Africa in general and for the upper Nyabarongo catchment of Rwanda. The methods and the predicting factors used in the present study were chosen based on the availability of data, the objectives of the study as well as the size and environmental settings of the study area. The produced maps were categorized into five classes of very high, high, moderate, low and very low susceptibility based on the natural break method. The final results were compared and validated using the AUC/ROC, accuracy, the RMSE and precision measures. According to the produced results, it can be confirmed that the NBT and LMT models achieved the highest prediction capability, but the RF model also proved a promising approach for susceptibility prediction though it yielded lower values for AUROC and the statistical measures.

Therefore, the results showed that NBT model produced the highest value of AUC (82.4%), followed by the LMT method (80.9%) while the RF model produced the least AUC value (78.4%). Additionally, for statistical measures of accuracy and precision, the NBT method produced the most reasonable results (0.799 and 0.745 respectively). Also, for RMSE estimator, the NBT model outperformed other two models with 0.301 RMSE value. Moreover, these prediction values confirm that the derived susceptibility maps for this study are effectively reliable. Overall, the three employed approaches proved real promising models for spatially predicting landslide susceptibility for Eastern-Africa region. Conclusively, these results may be suitable for further studies and for appropriate landslide risk reduction and mitigation for Rwanda and for different parts across globe with similar topographical settings. They can inform policy and decision making in regards with natural risk management.

Acknowledgements

Sincere appreciations are addressed to the Institute of Policy Analysis and Research-Rwanda (IPAR-Rwanda) for providing logistics and other different form of support towards the completion of this research. Additionally, the authors are thankful to the support in data and information provision by the Ministry in Charge of Emergency Management (MINEMA) in Rwanda and the Ministry of Environment (MoE). The authors thank the Editor and the two anonymous reviewers for their inputs and contribution to the work.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability

All datasets used by this research are available upon request from the corresponding author.

References

Abella EC, Van Westen C. 2007. Generation of a landslide risk index map for Cuba using spatial multi-criteria evaluation. Landslides. 4(4):311–325.

Ahmed B, Dewan A. 2017. Application of bivariate and multivariate statistical techniques in landslide susceptibility modeling in chittagong city corporation, bangladesh. Remote Sens. 9(4):304.

Akgun A, Erkan O. 2016. Landslide susceptibility mapping by geographical information system-based multivariate statistical and deterministic models: in an artificial reservoir area at Northern Turkey. Arab J Geosci. 9(2):165.

Amanambu AC, Li L, Egbinola CN, Obarein OA, Mupenzi C, Chen D. 2019. Spatio-temporal variation in rainfall-runoff erosivity due to climate change in the Lower Niger Basin, West Africa. CATENA. 172:324–334.

Ayala-Izurieta JE, Márquez CO, García VJ, Recalde-Moreno CG, Rodríguez-Llerena MV, Damián-Carrión DA. 2017. Land cover classification in an ecuadorian mountain geosystem using a random forest classifier, spectral vegetation indices, and ancillary geographic data. Geosciences. 7(2):34.

Begueria S. 2006. Validation and evaluation of predictive models in hazard assessment and risk management. Nat Hazards. 37:315–329.

Bordoni M, Meisina C, Valentino R, Bittelli M, Chersich S. 2015. Site-specific to local-scale shallow landslides triggering zones assessment using TRIGRS. Nat Hazards Earth Syst Sci. 15(5):1025–1050.

Bui DT, Tuan TA, Hoang N-D, Thanh NQ, Nguyen DB, Van Liem N, Pradhan B. 2017. 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. 14(2):447–458.

Bui DT, Tuan TA, Klempe H, Pradhan B, Revhaug I. 2016. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides. 13(2):361–378.

Calvello M, Pecoraro G. 2018. FraneItalia: a catalog of recent Italian landslides. Geoenviron Disasters. 5(1): 5–13.

Carrara A, Cardinali M, Detti R, Guzzetti F, Pasqui V, Reichenbach P. 1991. GIS techniques and statistical models in evaluating landslide hazard. Earth Surf Process Landforms. 16(5): 427–445.

Carrara A, Guzzetti F, Cardinali M, Reichenbach P. 1999. Use of GIS technology in the prediction and monitoring of landslide hazard. Natural Hazards. 20(2/3):117–135.

Cervi F, Berti M, Borgatti L, Ronchetti F, Manenti F, Corsini A. 2010. A.: Comparing predictive capability of statistical and deterministic methods for landslides susceptibility mapping: a case study in the northern Apennines (Reggio Emilia Province, Italy). Landslides. 7(4): 433–444.

Chen W, Peng J, Hong H, Shahabi H, Pradhan B, Liu J, Zhu A-X, Pei X, Duan Z. 2018a. Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China. Sci Total Environ. 626:1121–1135.

Chen W, Pourghasemi HR, Panahi M, Kornejady A, Wang J, Xie X, Cao S. 2017a. Spatial prediction of landslide susceptibility using an adaptive neuro-fuzzy inference system combined

with frequency ratio, generalized additive model, and support vector machine techniques. Geomorphology. 297:69–85.
Chen W, Pourghasemi HR, Zhao Z. 2017b. A GIS-based comparative study of Dempster-Shafer, logistic regression and artificial neural network models for landslide susceptibility mapping. Geocarto Int. 32(4):367–385.
Chen W, Sun Z, Han J. 2019. Landslide susceptibility modeling using integrated ensemble weights of evidence with logistic regression and random forest models. Appl Sci. 9(1):171.
Chen W, Xiaoshen X, Jianbing P, Himan S, Haoyuan H, Dieu TB, Zhao D, Shaojun L, and A-Xing Z 2018a. GIS-based landslide susceptibility evaluation using a novel hybrid integration approach of bivariate statistical-based random forest method. Catena. 164:135–149.
Chen W, Xie X, Wang J, Pradhan B, Hong H, Bui DT, Duan Z, Ma J. 2017d. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. CATENA. 151:147–160.
Chen W, Xie X, Peng J, Wang J, Duan Z, Hong H. 2017c. GIS-based landslide susceptibility modelling: a comparative assessment of kernel logistic regression, Naïve-Bayes tree, and alternating decision tree models. Geomatics Nat Hazards Risk. 8(2):950–973.
Chen W, Zhang S, Li R, Shahabi H. 2018b. Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naïve Bayes tree for landslide susceptibility modeling. Sci Total Environ. 644:1006–1018.
Ciurleo M, Cascini L, Calvello M. 2017. A comparison of statistical and deterministic methods for shallow landslide susceptibility zoning in clayey soils. Eng Geol. 223(7):71–81.
Claessens L, Knapen A, Kitutu M, Poesen J, Deckers JA. 2007. Modelling landslide hazard, soil redistribution and sediment yield of landslides on the Ugandan footslopes of Mount Elgon. Geomorphology. 90(1–2):23–35.
Corominas J, Van Westen C, Frattini P, Cascini L, Malet JP, Fotopoulou S, Catani F, Van Den Eeckhaut M, Mavrouli O, Agliardi F, et al. 2014. Recommendations for the quantitative analysis of landslide risk. Bull Eng Geol Environ. 73:209–263.
Cruden DM, Varnes DJ. 1996. Landslide types and processes. In: Turner AK, Schuster RL, editors. Landslides: investigation and mitigation. Washington, DC: National Academy Press; p. 36–75.
Ding Q, Chen W, Hong H. 2016. Application of frequency ratio, weights of evidence and evidential belief function models in landslide susceptibility mapping. Geocarto Int. 32(6):1–639.
Dou J, Yamagishi H, Zhu Z, Yunus AP, Chen CW. 2018. TXT-tool 1.081-6.1 A Comparative Study of the Binary Logistic Regression (BLR) and Artificial Neural Network (ANN) Models for GIS-based spatial predicting landslides at a regional scale. In Landslide Dynamics: ISDR-ICL landslide interactive teaching tools; Springer, Cham; p. 139–151.
Dou J, Yunus AP, Tien Bui D, Merghadi A, Sahana M, Zhu Z, Chen C-W, Khosravi K, Yang Y, Pham BT. 2019. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. Sci Total Environ. 662:332–346.
Ercanoglu M, Gokceoglu C. 2004. Use of fuzzy relations to produce landslide susceptibility map of a landslide prone area (West Black Sea Region, Turkey). Eng Geol. 75(3–4):229–250.
Godt J, Baum R, Savage W, Salciarini D, Schulz W, Harp E. 2008. Transient deterministic shallow landslide modeling: requirements for susceptibility and hazard assessments in a GIS framework. Eng Geol. 102(3–4):214–226.
Goetz J, Brenning A, Petschko H, Leopold P. 2015. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. Comput Geosci. 81:1–11.
Guzzetti F, Cardinali M, Reichenbach P, Carrara A. 2000. Comparing landslide maps: A case study in the upper Tiber River Basin, central Italy. Environ Manage. 25(3):247–263.
Hengl T, Heuvelink GBM, Kempen B, Leenaars JGB, Walsh MG, Shepherd KD, Sila A, MacMillan RA, Mendes de Jesus J, Tamene L, et al. 2015. Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. PLoS One. 10(6):e0125814.

Juliev M, Mergili M, Mondal I, Nurtaev B, Pulatov A, Hübl J. 2019. Comparative analysis of statistical methods for landslide susceptibility mapping in the Bostanlik District, Uzbekistan. Sci Total Environ. 653:801–814.

Karabulut EM, Ibrikci T. 2014. Effective automated prediction of vertebral column pathologies based on logistic model tree with SMOTE preprocessing. J Med Syst. 38(5):50.

Kausar N, Majid A. 2016. Random forest-based scheme using feature and decision levels information for multi-focus image fusion. Pattern Anal Appl. 19(1):221–236.

Lanfredi Sofia C, Oliveira S, Pereira S, Zêzere J, Corsini A. 2018. A comparison between bivariate and multivariate methods to assess susceptibility to liquefaction-related coseismic surface effects in the Po Plain (Northern Italy). Geomatics Nat Hazards Risk. 9(1):108–126.

Maes J, Parra C, Mertens K, Bwambale B, Jacobs L, Poesen J, Dewitte O, Vranken L, de Hontheim A, Kabaseke C, et al. 2018. Questioning network governance for disaster risk management: Lessons learnt from landslide risk management in Uganda. Environ Sci Policy. 85:163–171.

MIDIMAR, 2015. National Risk Atlas of Rwanda.Available online:. http://midimar.gov.rw/index.php?id=76&tx_pagebrowse_pi1%5Bpage%5D=3&cHash= f1359c48518cebbd859e-c0e04d7c02f3, Accessed date: 30th November, 2018

MIDIMAR, 2016: The National Contingency Matrix Plan for Rwanda. http://minema.gov.rw/uploads/tx_download/NATIONAL_DISASTER_CONTINGENCY_MATRIX_.pdf. Accessed 14 September, 2018.

MIDIMAR. 2018. Rwanda Rapid Post Disaster Needs Assessment (PDNA). Kigali-Rwanda: Ministry in Charge of Emergency Management. pp. 232.

MINAGRI. 1995. National Soils Map of Rwanda at a Scale of 1.50,000. Rwanda Ministry of Agriculture, Kigali, Rwanda. https://popups.uliege.be/1780-4507/index.php?id=10902. Accessed 14th September, 2018.

Montrasio L, Valentino R, Losi GL. 2012. Shallow landslides triggered by rainfalls: modeling of some case histories in the Reggiano Apennine (Emilia Romagna Region, Northern Italy). Nat Hazards. 60(3):1231–1254.

Nahayo L, Ndayisaba F, Karamage F, Nsengiyumva JB, Kalisa E, Mind'je R, Mupenzi C, Li L. 2019. Estimating landslides vulnerability in Rwanda using analytic hierarchy process and geographic information system. Integr Environ Assess Manag. 15(3):364–373.

Ndayisaba F, Guo H, Bao A, Guo H, Karamage F, Kayiranga A. 2016. Understanding the spatial temporal vegetation dynamics in Rwanda. Remote Sens. 8(2):129.

Nefeslioglu HA, Duman TY, Durmaz S. 2008. Landslide susceptibility mapping for a part of tectonic Kelkit Valley (Eastern Black Sea region of Turkey). Geomorphology. 94(3–4):401–418.

Nsengiyumva JB, Luo G, Amanambu AC, Mind'je R, Habiyaremye G, Karamage F, Ochege FU, Mupenzi C. 2019. Comparing probabilistic and statistical methods in landslide susceptibility modeling in Rwanda/Centre-Eastern Africa. Sci Total Environ. 659:1457–1472.

Nsengiyumva JB, Luo G, Nahayo L, Huang X, Cai P. 2018. Landslide susceptibility assessment using spatial multi-criteria evaluation model in Rwanda. IJERPH. 15(2):243.

Nyesheja EM, Chen X, El-Tantawi AM, Karamage F, Mupenzi C, Nsengiyumva JB. 2019. Soil erosion assessment using RUSLE model in the Congo Nile Ridge region of Rwanda. Phys Geogr. 40(4):322–339.

Paulín GL, Pouget S, Bursik M, Quesada FA, Contreras T. 2016. Comparing landslide susceptibility models in the Río El Estado watershed on the SW flank of Pico de Orizaba volcano, Mexico. Nat Hazards. 80(1):127–139.

Pellicani R, Van Westen CJ, Spilotro G. 2014. Assessing landslide exposure in areas with limited landslide information. Landslides. 11(3):463–480.

Persichillo MG, Bordoni M, Meisina C. 2017. The role of land use changes in the distribution of shallow landslides. Sci Total Environ. 574:924–937.

Persichillo MG, Bordoni M, Meisina C, Bartelletti C, Barsanti M, Giannecchini R, D'Amato Avanzi G, Galanti Y, Cevasco A, Brandolini P, et al. 2017. Shallow landslides susceptibility assessment in different environments. Geomatics Nat Hazards Risk. 8(2):748–771.

Pham BT, Bui DT, Pourghasemi HR, Indra P, Dholakia M. 2017. Landslide susceptibility assessment in the Uttarakhand area (India) using GIS: a comparison study of prediction capability of naïve bayes, multilayer perceptron neural networks, and functional trees methods. Theor Appl Climatol. 128(1–2):255–273.

Pham BT, Pradhan B, Bui DT, Prakash I, Dholakia M. 2016. A comparative study of different machine learning methods for landslide susceptibility assessment: a case study of Uttarakhand area (India). Environ Model Software. 84:240–250.

Pham BT, Shirzadi A, Bui DT, Prakash I, Dholakia M. 2018. A hybrid machine learning ensemble approach based on a Radial Basis Function neural network and Rotation Forest for landslide susceptibility modeling: A case study in the Himalayan area, India. Int J Sediment Res. 33(2):157–170.

Pham BT, Nguyen V-T, Ngo V-L, Trinh PT, Ngo HTT, Bui DT. 2017. A novel hybrid model of rotation forest based functional trees for landslide susceptibility mapping: a case study at Kon Tum Province, Vietnam. Paper presented at the International Conference on Geo-Spatial Technologies and Earth Resources.

Pisano L, Zumpano V, Malek Ž, Rosskopf CM, Parise M. 2017. Variations in the susceptibility to landslides, as a consequence of land cover changes: A look to the past, and another towards the future. Sci Total Environ. 601-602:1147–1159.

Pourghasemi HR, Yansari ZT, Panagos P, Pradhan B. 2018. Analysis and evaluation of landslide susceptibility: a review on articles published during 2005–2016 (periods of 2005–2012 and 2013–2016). Arab J Geosci. 11(9):193.

Pradhan B, Mansor S, Pirasteh S, Buchroithner MF. 2011. Landslide hazard and risk analyses at a landslide prone catchment area using statistical based geospatial model. Int J Remote Sens. 32(14):4075–4087.

Pradhan B, Lee S. 2010. Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. Environ Model Soft. 25(6):747–759.

RAB. 2000. Rwanda agriculture board: Rwanda soils properties database. Kigali, Rwanda: MINAGRI.

Ramani SE, Pitchaimani K, Gnanamanickam VR. 2011. GIS based landslide susceptibility mapping of Tevankarai Ar sub-watershed, Kodaikkanal, India using binary logistic regression analysis. J Mt Sci. 8(4):505–517.

RNRA. 2015. Rwanda geology and mines maps; [Accessed 2017 July 6]. http://www.rnra.rw/index/php?id=15.

Tharwat A. 2018. Classification assessment methods. Appl Comput Informatics. doi: 10.1016/j.aci.2018.08.003

Tsangaratos P, Ilia I. 2016. Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size. CATENA. 145:164–179.

Tseng C, Lin C, Hsieh W. 2015. Landslide susceptibility analysis by means of event-based multi-temporal landslide inventories. Nat Hazards Earth System Sci Discussions. 3(2): 1137–1173.

UNISDR. 2016. Disaster in numbers *Prevention Web*: United National International Strategy for Disaster Risk Reduction. https://www.unisdr.org/files/52253_unisdr2016annualreport.pdf. Accessed 3th March, 2019

Valentino R, Meisina C, Montrasio L, Losi GL, Zizioli D. 2014. Predictive power evaluation of a physically based model for shallow landslides in the area of Oltrepò Pavese, Northern Italy. Geotech Geol Eng. 32(4):783–805.

Van Den Eeckhaut M, Marre A, Poesen J. 2010. Comparison of two landslide susceptibility assessments in the Champagne–Ardenne region (France). Geomorphology. 115(1–2):141–155.

Van Den Eeckhaut M, Poesen J, Verstraeten G, Vanacker V, Moeyersons J, Nyssen J, Van Beek L. 2005. The effectiveness of hillshade maps and expert knowledge in mapping old deep-seated landslides. Geomorphology. 67(3–4):351–363.

Van Tien P, Sassa K, Takara K, Fukuoka H, Dang K, Shibasaki T, … Luong LH. 2018. TXT-tool 4.081-1.1: Mechanism of large-scale deep-seated landslides induced by rainfall on gravitationally deformed slopes: A case study of the Kuridaira Landslide in the Kii Peninsula. In Japan landslide dynamics: ISDR-ICL landslide interactive teaching tools. Springer, Cham; p. 793–806.

Van Westen CJ, Ghosh S, Jaiswal P, Martha TR, Kuriakose SL. 2013. From landslide inventories to landslide risk assessment; an attempt to support methodological development in India. In Landslide science and practice. Springer, Berlin, Heidelberg; p. 3–20.

Yalcin A. 2008. GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): comparisons of results and confirmations. CATENA. 72(1):1–12.

Yilmaz I, Ercanoglu M. 2019. Landslide inventory, sampling and effect of sampling strategies on landslide susceptibility/hazard modelling at a glance. In Natural Hazards GIS-based spatial modeling using data mining techniques. Springer, Cham; p. 205–224.

Youssef AM, Pourghasemi HR, El-Haddad BA, Dhahry BK. 2016. Landslide susceptibility maps using different probabilistic and bivariate statistical models and comparison of their performance at Wadi Itwad Basin, Asir Region, Saudi Arabia. Bull Eng Geol Environ. 75(1): 63–87.

Zêzere J, Pereira S, Melo R, Oliveira S, Garcia R. 2017. Mapping landslide susceptibility using data-driven methods. Sci Total Environ. 589:250–267.

Zizioli D, Meisina C, Valentino R, Montrasio L. 2013. Comparison between different approaches to modeling shallow landslide susceptibility: a case history in Oltrepo Pavese, Northern Italy. Nat Hazards Earth Syst Sci. 13(3):559–573.