Research Article

Azemeraw Wubalem*

Modeling of Landslide susceptibility in a part of Abay Basin, northwestern Ethiopia

https://doi.org/10.1515/geo-2020-0206
received July 25, 2020; accepted October 16, 2020

Abstract: The study area in northwestern Ethiopia is one of the most landslide-prone regions, which is characterized by frequent high landslide occurrences. To predict future landslide occurrence, preparing a landslide susceptibility mapping is imperative to manage the landslide hazard and reduce damages of properties and loss of lives. Geographic information system (GIS)-based frequency ratio (FR), information value (IV), certainty factor (CF), and logistic regression (LR) methods were applied. The landslide inventory map is prepared from historical records and Google Earth imagery interpretation. Thus, 717 landslides were mapped, of which 502 (70%) landslides were used to build landslide susceptibility models, and the remaining 215 (30%) landslides were used to model validation. Eleven factors such as lithology, land use/cover, distance to drainage, distance to lineament, normalized difference vegetation index, drainage density, rainfall, soil type, slope, aspect, and curvature were evaluated and their relationship with landslide occurrence was analyzed using the GIS tool. Then, landslide susceptibility maps of the study area are categorized into very low, low, moderate, high, and very high susceptibility classes. The four models were validated by the area under the curve (AUC) and landslide density. The results for the AUC are 93.9% for the CF model, which is better than 93.2% using IV, 92.7% using the FR model, and 87.9% using the LR model. Moreover, the statistical significance test between the models was performed using LR analysis by SPSS software. The result showed that the LR and CF models have higher statistical significance than the FR and IV methods. Although all statistical models indicated higher prediction accuracy, based on their statistical significance analysis result (Table 5), the LR model is relatively better followed by the CF model for regional land use planning, landslide hazard mitigation, and prevention purposes.

Keywords: landslide, susceptibility, geographic information system, certainty factor, frequency ratio, information value, logistic regression, Ethiopia

1 Introduction

As defined by Brunsden and Cruden [1,2], landslides are the downslope movements of debris, rocks, or earth material under the influence of the force of gravity [3]. Landslides are also defined as a large range of geotechnical phenomena under the influence of gravity. Although the cause of landslide incidence and its mechanisms are so complex, it is triggered by heavy rainfall, earthquake, and human interventions. It can occur when the driving force exceeds the resistance force because of the destabilization of natural soil or rock slopes [4,5]. Landslides can bury animals and humans in a short period as well as can destroy houses, farms, and infrastructures [6]. It is one of the most destructive and dangerous natural hazards that cause numerous fatality and economic losses worldwide [6–9]. Ethiopia is one of the countries affected by heavy rainfall, human intervention, and earthquake-triggered frequent landslide impacts yearly, resulting in loss of human and animal lives and damage of infrastructures and properties [5,10–14]. In the last 2 years, from 2018 to 2019, because of rainfall-triggered landslides 60 people died, 30 people injured, 5,091 households displaced, houses damaged, and a widely cultivated and non-cultivated land destructed in different parts of the country [5]. Besides, on 16 October 2019, 23 people died, and on 28 May, 22 people died and also there was a destruction of wide areas of farmlands by heavy rainfall-triggered landslide incidence. This shows that huge economic loss caused by landslide incidences retards the sustainable development of the economy of the country. Despite the landslide problem is critical in Ethiopia, still there is no adequate slope stability assessment applied in the different parts of the country [5]. The study area is one of the areas that were frequently affected

* **Corresponding author: Azemeraw Wubalem,** Engineering Geologist, College of Natural and Computational Sciences, Department of Geology, University of Gondar, Posta 196, Gondar, Ethiopia, e-mail: alubelw@gmail.com

by the rainfall-triggered landslide incidences, and so far, the area has not yet been studied. Landslide in this area resulted in the damage of three houses, farmlands, and loss of animal lives. From local people's witness, rainfall and stream cut triggered deep-seated rotational landslides that occurred in 2018 and reactivated in 2019 in the Desa Enese village, which destroyed wide ranges of farmlands that covered by crops. Therefore, landslide susceptibility mapping and assessment in this area are very important to disaster loss reduction and serve as a guideline for sustainable land use planning in the study area.

The mitigation measures of landslide incidence in the area, which is already failing or susceptible to fall, require identification of existing landslide, determination of the contribution of prevailing causal factors, and generation of landslide susceptibility map [15]. Landslide susceptibility is the result of the relationship between past landslide and environmental factors [16]. It is an important basis and scientific support for the government's response to landslide hazards management and land use practice [17,16]. The landslide susceptibility mapping is not only to ascertain the factors that are most influential to the landslides that occurred in the region but also to estimate the relative contribution of each factor for slope failures [16]. It is also important to inaugurate an association between the factors and landslides to foresee the landslide hazard in the future [16].

Several approaches are developed for landslide susceptibility mapping, which can be categorized into qualitative, semi-quantitative, and quantitative methods [19–23]. Qualitative methods are an expert-driven approach, which required field experience specialists [19,20,22,24–27]. It mainly includes field geomorphological analysis, parameter assignment superimposition, and landslide inventory analysis. Relying on the experience and professional background knowledge of experts and subjectivity is the drawback of these methods [19,20,22–26]. Semi-quantitative methods are the combination of qualitative and quantitative methods, which introduce grading and weighting of the effects of landslide factors on landslide incidence [23,28,29]. In this method, both qualitative and quantitative methods can be applied to evaluate the effects of landslide governing factors on landslide occurrence [30]. Analytical hierarchy process, weighted linear combination, and expert knowledge/heuristic [23,27,31,32] are examples of semi-quantitative methods. Although some statistical concepts are introduced in this method, it depends on the expert's experience and the background of professional knowledge and some subjectivity remains [23,33]. The quantitative methods can be divided into three categories such as statistical, machine learning/data mining, and deterministic methods [34–36]. Statistical methods are indirect susceptibility mapping methods widely or routinely used to evaluate the correlation between landslide governing factors and landslides based on mathematical expression [16,19–22]. The statistical methods can be further divided into multivariate and bivariate statistical methods, which are widely used throughout the world and provide reliable results [5,14,23,37–43]. The bivariate statistical methods are used to evaluate the relationship between landslide governing factors and past landslides. Certainty factor (CF) [44–48], frequency ratio (FR) [14,49–52], information value (IV) [5,53–55], and weight of evidence [46,49,55–58] are examples of bivariate statistical methods, which are simple and easy and produce reliable models. It also helps to evaluate the effects of a landslide at a factor class level that is impossible in data mining or multivariate methods. However, it requires quality input data, past landslide data, and lacking to evaluate the relationship among landslide governing factors. Multivariate statistical methods are used to examine the relationship between three and above dependent and independent variables [39,40,43,59,60]. Although logistic regression (LR) and discriminant analysis are examples of multivariate statistical methods used frequently in landslide susceptibility modeling and provide reliable results [14,16,39,40,43,60], it is incapable to examine the contribution of each factor class for landslide probability like data mining, unlike bivariate methods. Deterministic methods are used to calculate the quantitative value of the inherent slope materials of the factor of safety over a defined area [61]. These methods can be applied when landslide types are simple (shallow landslides) and the intrinsic properties of slope material are homogeneous [61]. It requires detailed ground data such as soil strength, unit weight of soil, pore water pressure, soil layer thickness, slope angle, slope height, and depth below the terrain surface. Deterministic method has been employed over a small area, and oversimplification, data availability to acquire frequently is impossible are the drawback of these methods [61]. In recent times, advanced data mining methods have been widely used in landslide susceptibility modeling, including random forest [62], boosted regression tree [63], classification and regression tree [63], Naïve Bayes [29,64], support vector machines [31,65,66], kernel LR [67], logistic model tree [67,62], index of entropy [68], and artificial neural networks [69–71]. However, data mining methods are time-consuming, are incapable to determine the effects of each

landslide factor class, and require high computing capacity, and the internal calculation process of these methods is intensive and cannot easily be understood. Although there is a bit little difference in the degree of predictive accuracy, both statistical and data mining methods provide reliable predictive accuracy [18,71,72]. In general, each of the methods has its merit and demerits. For example, the results of the qualitative methods highly depend on the expert of the researcher. In addition, the bivariate statistical method helps to evaluate the impact of each landslide governing factor classes on landslide occurrence, but excluded CF, and it does not consider the relationship between these factors and landslide occurrence [10], whereas multivariate statistical analysis (LR) can consider it.

Although CF, IV, FR, and LR methods are routinely used methods for landslide susceptibility mapping, they have some foreseeable limitations. For example, the LR model can perform multivariate statistical analysis between a dependent variable and a set of independent variables, but it is incapable to analyze the impacts of internal classes of landslide governing factors individually on landslide occurrence [61,73]. However, the CF can analyze both the impacts of individual factors and factor class on landslide occurrence. FR and IV are simple and effective statistical methods that can extract the influence of each landslide governing factor classes on landslide occurrence, but it cannot consider the relationship between these landslide governing factors and landslide occurrence [74]. Therefore, it is important to use both bivariate and multivariate statistical analyses to evaluate the impacts of landslide factors and landslide factor classes in prediction modeling of landslide susceptibility in the study area (in a part of Abay Basin).

The primary objective of this study is to determine landslide-prone areas using LR, FR, CF, and IV statistical methods in the study area. The second objective is to compare the application of FR, IV, LR, and CF methods for predictive modeling of landslide susceptibility in the study area. The study area is one of the areas characterized by populating settlements, intensive farming, and frequent landslide incidence, which destroyed widely cultivated land, and it is important to evaluate the factors that have more roles in causing slope failure and to minimize their socioeconomic impacts by generating a landslide susceptibility map. For this purpose, statistical methods including FR, LR, IV, and CF methods were applied. These methods are easy to apply and it gives a very well-meaning result. In the literature, various bivariate and multivariate approaches for landslide susceptibility mapping are available; however, a comparison among LR, CF, FR, and IV models has yet not been encountered. A comparison of the four models has been discussed in this study. The accuracy of the results of landslide susceptibility maps, which were generated using the statistical methods, was evaluated using the receiver operating characteristic curve (ROC) and landslide density. The resulted maps will be used for landslide mitigation purposes and regional land use planning.

2 Study area and geological setting

The study area is one part of the Abay basin (Gozamn, Baso Liben, and Awabel Districts), which is located in the northwestern highlands of Ethiopia. It lies within the latitude 10,90,000 mN to 11,40,000 mN and the longitude 3,30,000 mE to 4,00,000 mE. The minimum and maximum altitudes of the area are 902 m of the river gorge and 2,563 m in hills and plateau lands (Figure 1). Many tributaries are available in the entire study area, which drains into the Abay River. The various streams in the study area caused the removal of soil through stream bank erosion. The study area is characterized by variable topographic conditions, including ridge, cliff, hill, plateau, deep River gorge, and gentle slope. The fragile nature of topography has been in facilities the rate of soil erosion. Of this region, 50% is covered by agricultural lands, and rocky lands/bar lands, residential lands, and grazing lands cover the remaining lands. Tropical to subtropical climatic condition prevails in the study area. The main characteristic of the climates in the study area is the monsoon rainfall, which occurs between June and September and delivers an average of 90% of the total rainfall of the year. This resulted in landslide incidence in the study area. For example, reactivated landslides in the Desa Enese village occurred after heavy and prolonged rainfall in August 2019. The maximum annual rainfall is 1,394 mm, whereas 902 mm is the minimum annual rainfall.

Geologically, Ethiopia comprised the Precambrian basement rock, Paleozoic sedimentary rock, Mesozoic marine sedimentary rock, and Cenozoic volcanic rocks. However, the study area comprised mainly two geological units besides recent soil sediments at the slope toe of the study area, which is grouped into early Mesozoic and Cenozoic Era of sedimentary (red sandstone, gypsum, and limestone) and volcanic rocks (flood basalt), respectively. The flood basalt rock units are grouped into four geological units, such as Ashengie formation (lower basalt),

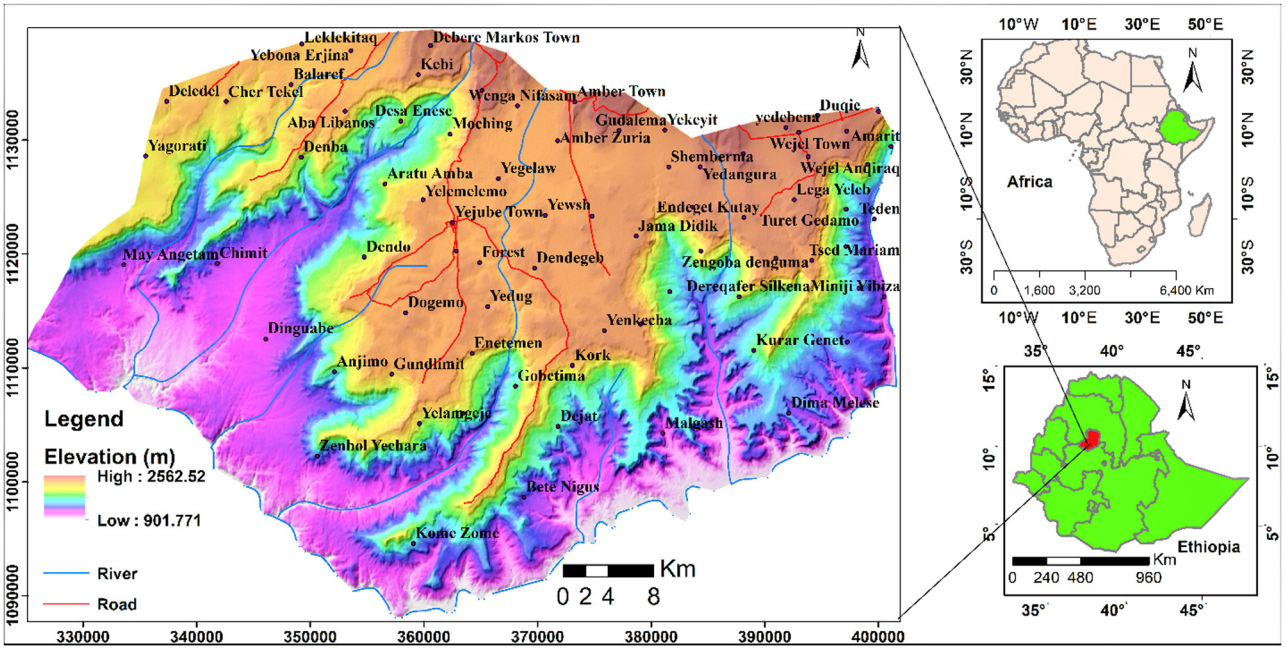

Figure 1: Location map of the study area.

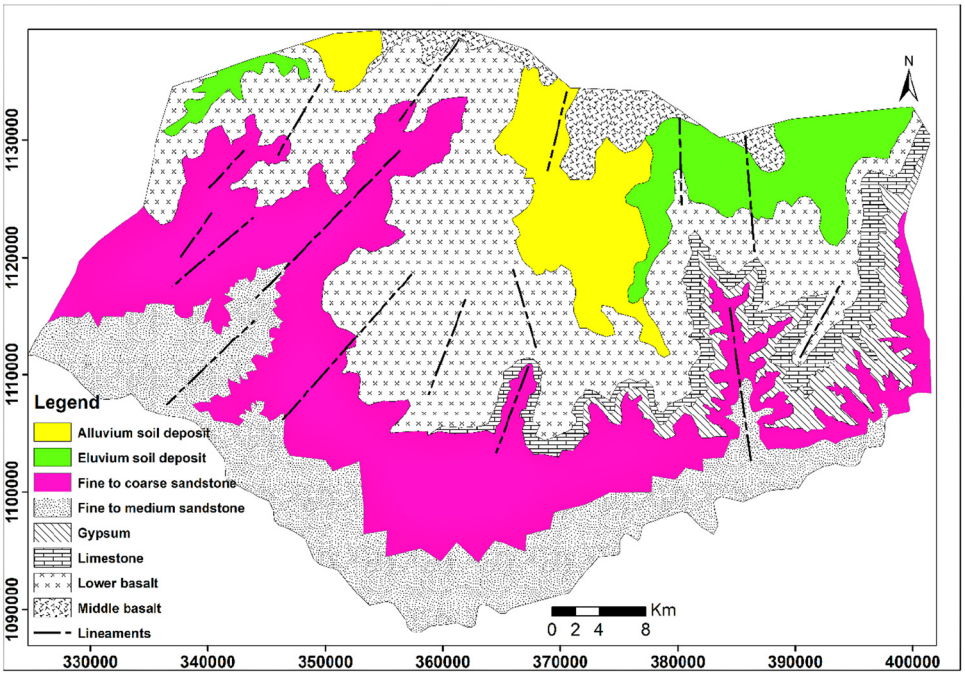

Figure 2: Geological map of the study area.

Aiba formation (middle basalt), Alaje formation (upper basalt), and Termaber formation. Nevertheless, the study area is comprised of only the lower basalt, middle basalt, the Mesozoic sedimentary rock (lower red sandstone, gypsum, and limestone), and the quaternary/recent soil deposit. This lithology is digitized from the existing 1:2,50,000 geological map of the Debre Maroks sheets. Figure 2 shows that the southern, southwestern, and southeastern parts of the study area are covered by the sedimentary rock (lower red sandstone, gypsum, and limestone, respectively). Lower sandstone is characterized by medium to thickly bedded thickness, fine to

coarse texture, red-to-red-brown color, and strongly cross-bedded. The northern, northwestern, southeastern, and central parts of the study area are dominantly covered by the volcanic rock (lower basalt), which is underlined early to middle Mesozoic sedimentary rock (red sandstone unit and gypsum unit) and covered by thin dark color soil deposit (Figure 2). This unit is characterized by a high degree of weathering and fracturing. The northwestern, northeastern, and eastern parts of the study area are covered by a very loss/unconsolidated soil deposit. In this soil deposit, unplanned intensive agricultural activities are common.

3 Methodology

For this study, data collection, landslide inventory mapping, Google Earth imagery analysis, landslide factor evaluation, and mapping, geographic information system (GIS)-based FR, IV, LR, and CF landslide susceptibility modeling and model validation was applied. Furthermore, relevant data, including digital elevation model (DEM) with 30 m resolution, topographic map, borehole data, historical landslide events, geological map, and meteorological data were collected (Table 1). These data were collected from the geological survey of Ethiopia (GSE), the United States Geological Survey (USGS), Amhara water well drilling enterprise (AWWDE), field survey, Google Earth imagery from the NASA, and Ethiopian National Meteorological Agency (Table 1). The landslide location of the study area was identified using historical records and Google Earth imagery analysis. These are classified into training and testing landslide data sets. The training landslide data sets are used for model preparation, whereas the testing landslide data sets are used for model prediction accuracy evaluation. Based on the data availability, local environmental conditions, literature, data evaluation, and local people interview, 11 landslide-driving factors were determined. The landslide driving factor maps and landslide inventory maps were prepared using ArcGIS 10.1. Distance to stream, drainage density, slope angle, slope aspect, and curvature were extracted from 30 m resolution of DEM, which was downloaded from the USGS website. The lithological layer and lineament map are digitized from the existing geological map of the Debre Markos sheet at a scale of 1:2,50,000. The soil map was prepared from the FAO soil map and the rainfall map was prepared using 30 year rainfall data, which was gathered from the Ethiopian Meteorological Agency. The land use map was prepared using ArcGIS and Google Earth imagery analysis. Land use map digitized from Google Earth imagery interpretation, which can export to a GIS layer format (Kml) and verified by ground truth as well as by the experience of the users in the local area for the final map due to high spatial resolution, easiness as well as user friendly. The land use map is also prepared using the supervised classification of satellite images downloaded from the USGS website. The general procedure flow chart that was followed in this research work is summarized in Figure 3.

Geodatabase building is one of the most fundamental elements in the landslide susceptibility mapping. Therefore, four databases were built for LR, FR, IV, and CF models. The FR, IV, and CF databases contain landslide inventory and landslide factors, whereas the LR database contains landslide and no landslide points with 11 weighted landslide factors. After the database was built, evaluation of the relationship between landslide and landslide factors and determination of the statistical significance of each landslide factor were the next steps in landslide susceptibility mapping. Therefore, 11 landslide factor maps were

Table 1: Information source for the various landslide factors used in the landslide susceptibility mapping

Data	Map	Format	Source
Landslide inventory	Landslide inventory map	Vector (shapefile)	Google Earth imagery, field survey, and historical record
Geology	Lithology and lineament map	Vector (shapefile)	Digitized from geological map of Debre Markos Sheet provided by the GSE at 1:50,000
DEM	Slope angle map Aspect map Curvature map	Raster (grid)	Derived from 30 m DEM, using ArcGIS 10.1, downloaded from USGS
Hydrology	Distance to stream and drainage density	Raster (grid)	Developed from DEM and buffering using distance to Euclidian
Topography	Topographic map	Vector (shapefile)	Ethiopian Mapping Agency at 1:50,000
Meteorological data	Rainfall map	Vector (shapefile)	Ethiopian National Meteorological Agency
Land use	Land use and soil map	Vector (shapefile)	Sentinel 2 images in the USGS, Field Survey, FAO, and Google Earth imagery
Borehole data	———	Vector (shapefile)	Amhara Water Well Drilling Enterprise (AWWAE)

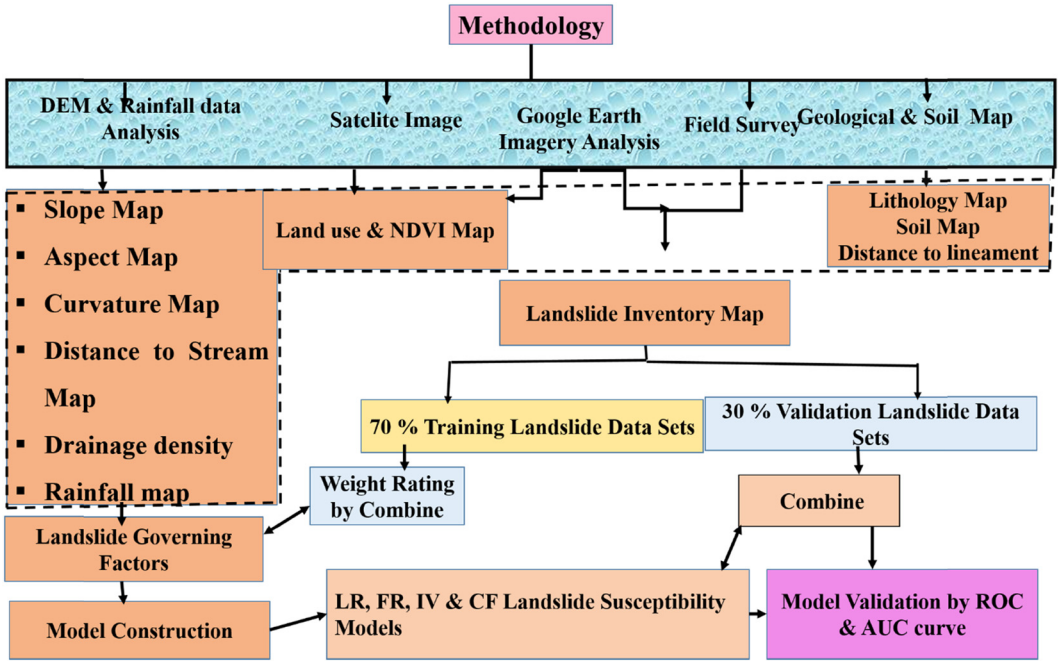

Figure 3: General flow charts for the methodology of the study area.

reclassified into subclass and overlaid with reclassified training landslide data sets. Weight ratings for all the landslide factor classes were assigned statistically using Excel as shown in equations (1), (3) and (5). These weighted maps were rasterized using lookup tool in spatial analysis. After rasterizing the factor maps, the landslide susceptibility index (LSI) maps were generated by the sum-up of all raster maps using a raster calculator in Map Algebra. These maps (LSI) are classified into a fivefold classification scheme: very low, low, moderate, high, and very high susceptibility classes using natural breaks (Figure 4b–d). In the case of the LR method, the study area was classified as training landslide and non-landslide points using GIS. Then, the weight of the 11 factors has been extracted to generate LR coefficients of each landslide factor in SPSS, and finally, the LSI of the area was generated using the logistic landslide probability equation (equation (8)) and GIS tools (Figure 4a). Finally, the accuracy of the four models was evaluated using the prediction rate curve and landslide density based on the observed testing landslide data sets (Figures 5 and 6).

4 Landslide inventory mapping

In landslide susceptibility mapping, landslide inventory mapping is one of the key elements, which can be prepared using various techniques such as the aerial photograph or Google Earth imagery interpretation, field investigation, and evaluation of archive data coupled with GIS tools [75]. Landslide inventory map is used as the base for future landslide prediction by evaluating the relationship between the existing landslide event and landslide driving factors [56,76,77]. A landslide inventory map was prepared using GIS from field investigation, historical landslide events, and satellite image analysis [75]. Landslide inventory maps can also be prepared using Google Earth imagery [5,14,54,78]. In the present research work, from the active and old landslide scarps, 717 landslides, which covered 7.8 km^2, are identified using historical landslide record and time series Google Earth imagery analysis (Figure 7). It digitized into polygons using a GIS tool with the help of Google Earth imagery, and finally, a landslide inventory map was produced (Figure 7). From local people witness and time series Google Earth imagery analysis, the study area is affected by landslide incidence frequently because of heavy and prolonged rainfall and the presence of unconsolidated soil deposit as well as highly weathered basalt rock unit. Soil slide, weathered rockslide, debris flow, earth flow, and earth fall types of landslides are dominant in the study area. In the literature, some researchers classified landslides into 80% for training landslide data sets and 20% for testing landslide data sets. However, most of the researchers classified landslides into 70% for training data sets and 30% for testing landslide data sets [14,54,58,68,78,79]. Using ArcGIS 10.1, these landslides were classified randomly into 70% (502) for training

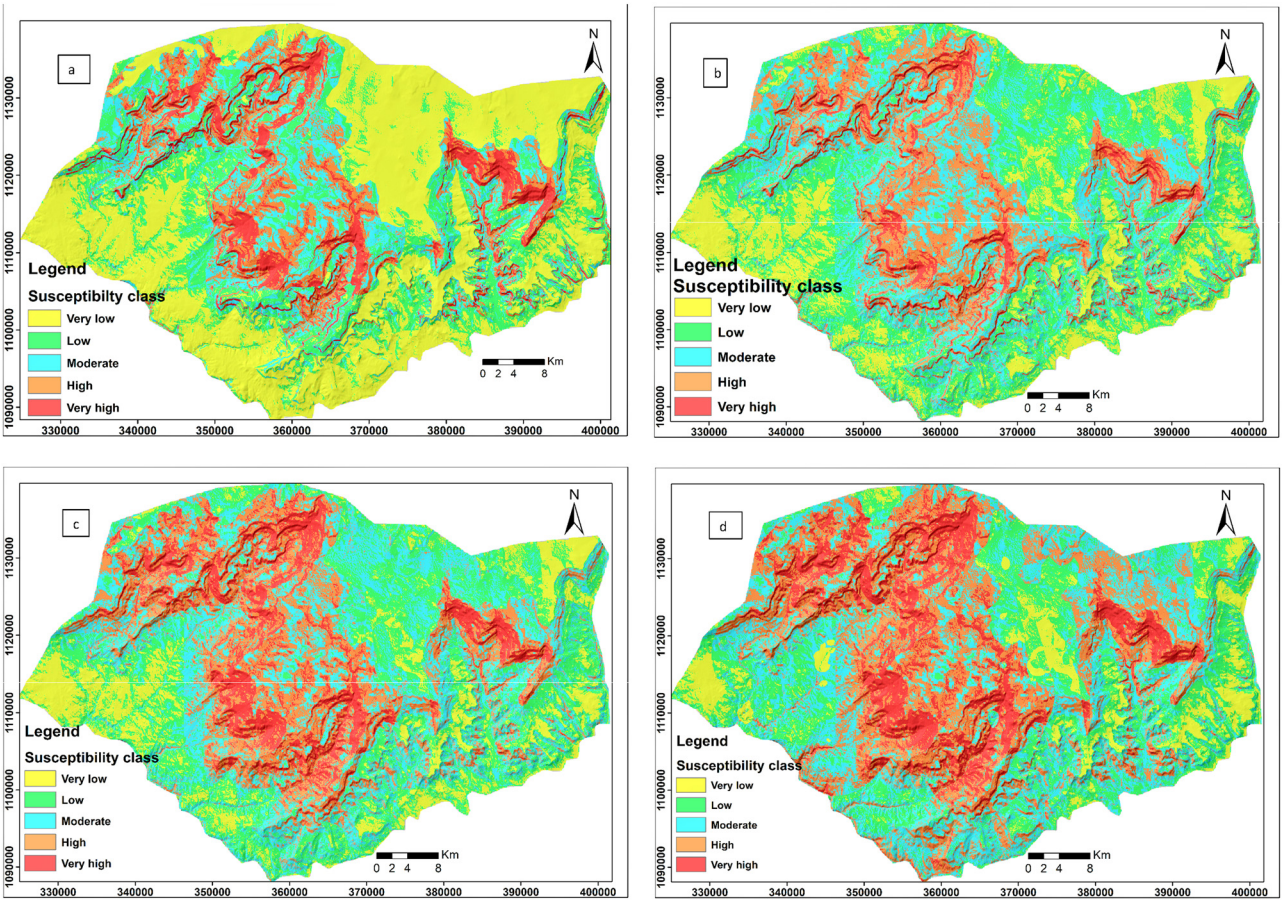

Figure 4: (a) Landslide susceptibility map using LR method. (b) Landslide susceptibility map using the FR method. (c) Landslide susceptibility map using the CF method. (d) Landslide susceptibility map of IV method.

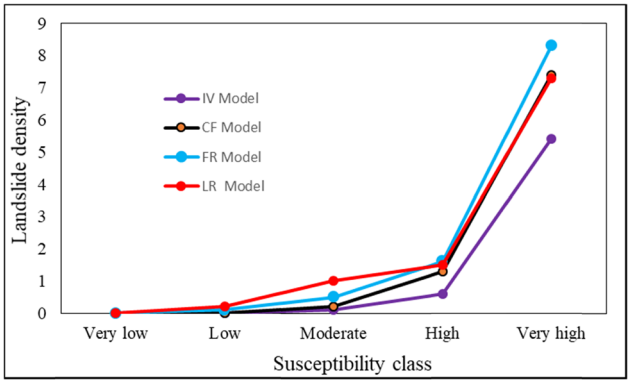

Figure 5: Landslide density in each susceptibility class.

landslide data sets and 30% (215) for validation data sets keeping their spatial distribution. The training landslide data sets observed landslides that were used to develop the model, whereas the validation landslide data sets observed landslides that were used to evaluate the performance and prediction accuracy of the model. As shown in Figures 8 and 9, the Desa Enese and Aba Libanos areas are affected by soil slides in 2019 because of heavy and prolonged rainfall. Rotational landslide in the Desa Enese area occurred because of the removal of the slope toe by a stream and resulted in damages in farmlands, which covered by crops, and two houses (Figure 8). As indicated in Figure 7, the spatial distribution of landslides concentrates dominantly on the ridge and along the stream bank.

5 Evaluation of landslide factors

In landslide susceptibility mapping, the selection of landslide factors is one of the most important elements. However, there are no well-defined standards to select the most significant landslide factors. The factors that initiate the landslide incidence in the study area are selected based on the data availability, local environmental conditions, literature review, local person

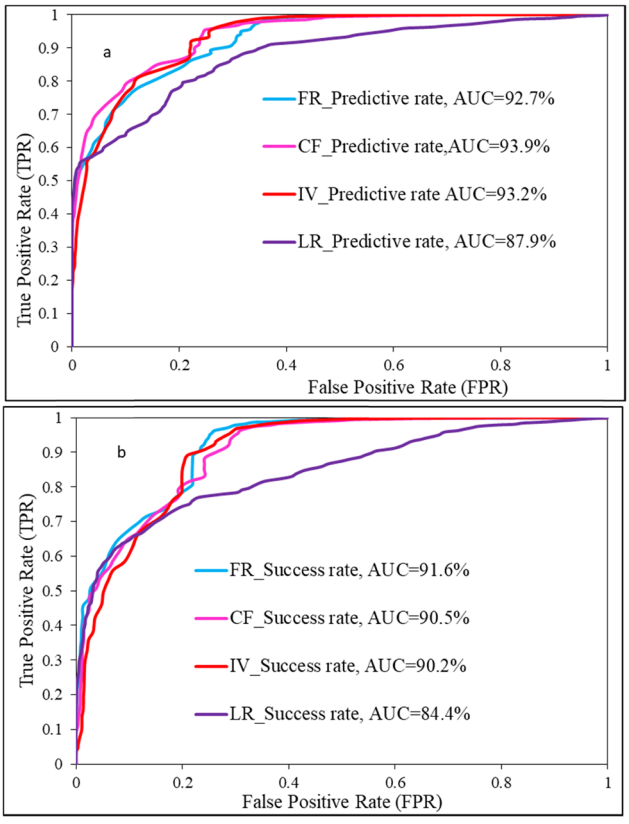

Figure 6: Receiver operating characteristics curve (ROC).

interview, and data evaluation using LR analysis. These are slope angle, aspect, slope curvature, normalized difference vegetation index (NDVI), land use, lithology, distance to lineament, drainage density, rainfall, soil, and distance to stream/river considered into account to examine the spatial relationship between them and landslide occurrence in the study area. Distance to stream (five classes), drainage density (five classes), slope angle (five classes), aspect (ten classes), and slope curvature (three classes) map constructed from 30 m resolution DEM was downloaded from the USGS website (Figure 10). The lithological map and lineament map of the study area were prepared through digitization from 1:2,50,000 existing geological maps of Debre Markos sheet from the GSE, which has eight classes (lower basalt, middle basalt, fine to coarse sandstone, fine to medium sandstone, gypsum, limestone, and unconsolidated sediments [alluvium and eluvium]). Google Earth imagery and the supervised classification of Sentinel 2 images were used to prepare land use maps of the study area. From the results, the land use map, which was prepared from Google Earth imagery, is more reliable compared to the supervised classification of Sentinel 2 images. Preparation of land use map using the supervised classification of satellite images could be best when the study area is so large and the users are not familiar to the region. Nevertheless, from the resulting point of view, using a manual land use classification of the Google

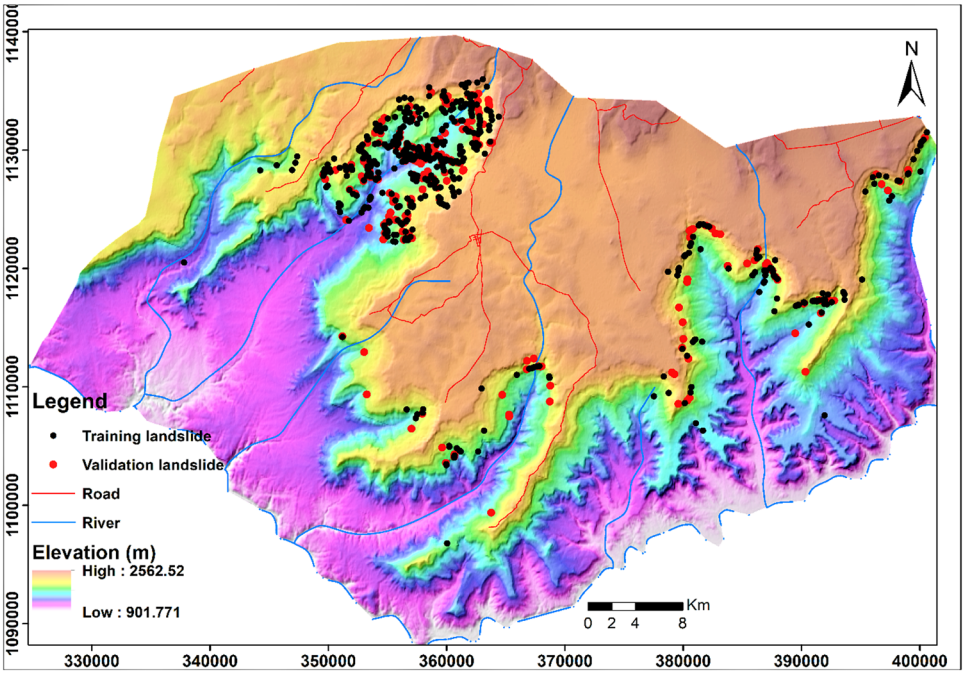

Figure 7: Landslide inventory map of the study area.

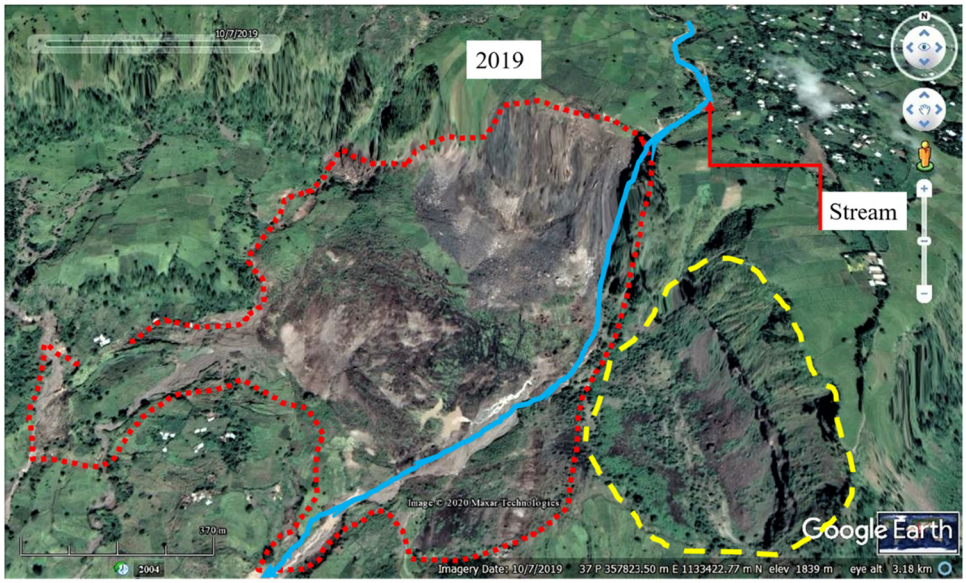

Figure 8: Time series Google Earth image showing failed slope at Desa Enese village in 2019 [80].

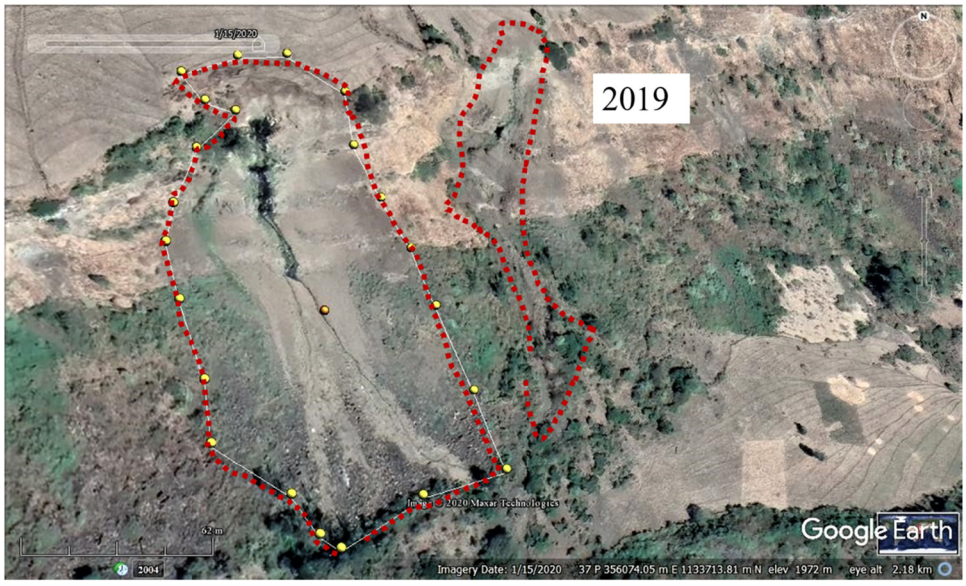

Figure 9: Time series Google Earth image showing failed slope at Aba Libanos village in 2019 [80].

Earth imagery found to be effective as it has a high spatial resolution, and the expert who classifies this image has direct control to identify what stands for what. However, Google Earth imagery requires an advanced Internet condition, and it is so effective when the area is well known by the user. The land use map has seven classes such as grassland, cultivated/cropland, scatter forest, woodland, bare land, residential, and scatters bush (Figure 10). Earthquake was not considered in the present work because the study area is so far from the active earthquake sites. The source of various landslide factors used in landslide susceptibility mapping is summarized in Table 1. To determine the effects of each landslide factor class on landslide occurrence, the weight rate for each factor classes is very important. For this purpose, all landslide factor maps converted into raster and reclassified with the same pixel size (30 m) and the same projection using GIS tools under the Arc toolbox in conversion as well as a spatial analysis tool. Then, the landslide inventory raster map overlaid through the combination in spatial analysis tool under the local toolbox with the landslide factor raster class to extract landslide pixels

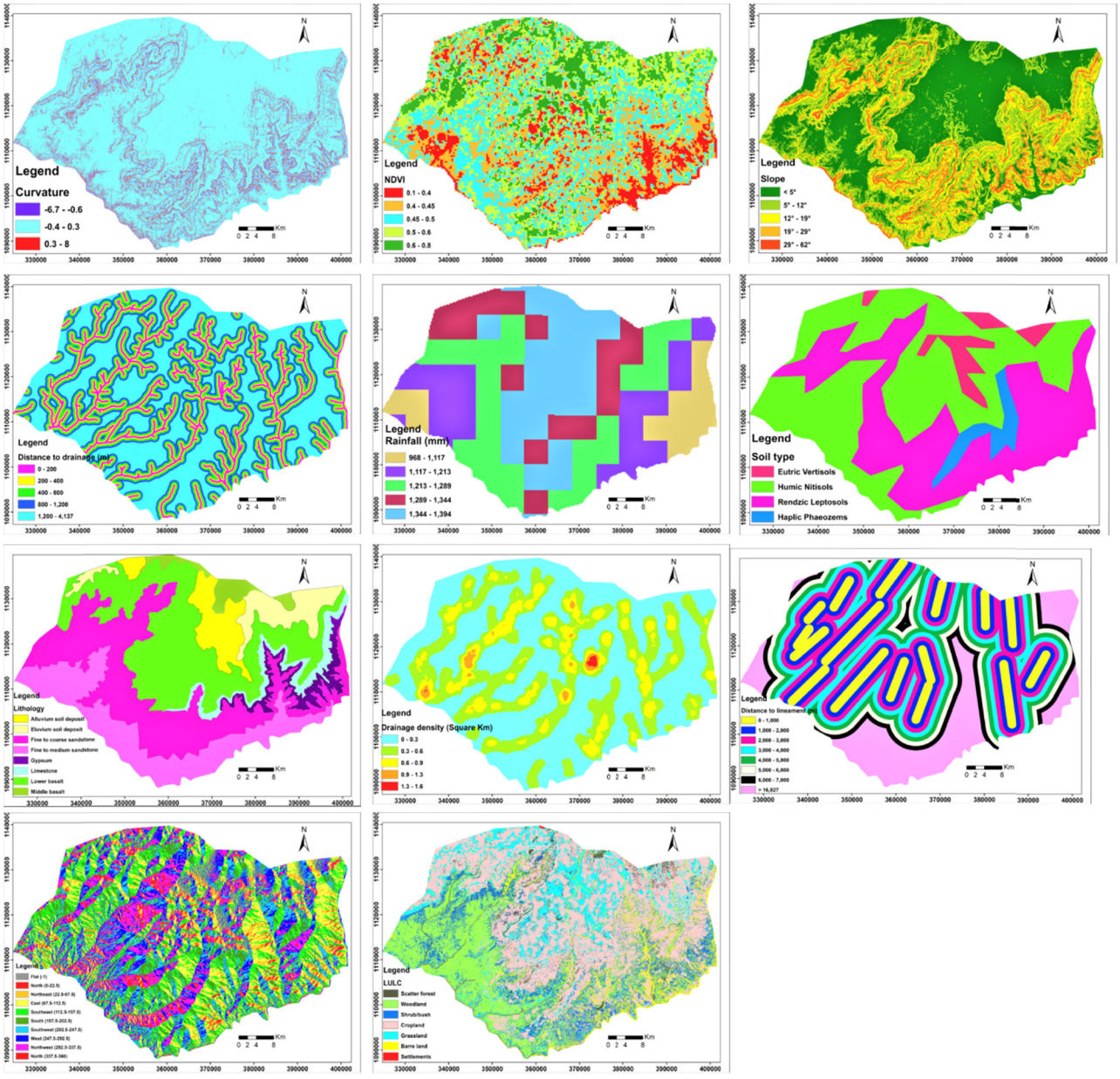

Figure 10: Landslide governing factor maps.

for each landslide factor class. Then the effects of each factor class were determined using the equation of FR (equation (1)), IV (equation (3)), and CF (equation (5)) methods, and the results are summarized in Table 3.

6 Modeling approaches

6.1 FR model

The FR model is one of the bivariate probability methods, which is applicable to determine the correlation between landslide occurrence and landslide causative factor classes. The FR is the ratio of areas in which the landslide occurred to areas in which landslides not occurred. When the ratio value is greater than one, it indicates the strong correlation between factor class and landslide occurrence in a given terrain; however, the ratio value less than one indicated that weak coloration between landslide occurrence and landslide factors, which means a low probability of landslide occurrence [82,83]. It can be calculated using equation (1).

$$\mathrm{FR} = \frac{a}{b} = \frac{\frac{\mathrm{Nslpix}}{\mathrm{Ntslpix}}}{\frac{\mathrm{Ncpix}}{\mathrm{Ntcpix}}}, \tag{1}$$

where FR is the frequency ratio, Nslpix is the landslide pixel/area in a landslide factor class, Ntslpix is the total area of a landslide in the entire study area (a), Ncpix is the area of the class in the study area, and Ntcpix is the total pixel area in the entire study area (b). In the present research work, the FR for each causative factor class is calculated using equation (1), and the results are summarized in Table 2.

After calculated the FR for each landslide factor class using Microsoft Excel and GIS, the FR value for each factor class assigned through the join in the ArcGIS tool. Then the weighted landslide factors were rasterized using the lookup tool in spatial analysis. The LSI indicated the degree of susceptibility of the area for landslide occurrence. The LSI of the study area was calculated by carefully summing up the weighted factor raster maps using equation (2) by the raster calculator in Map Algebra of the spatial analysis tool. To get the LSI, the FR of each factor type or class is summed as in equation (2).

$$\mathrm{LSI} = \sum_{n}^{i=1} \mathrm{FR}_i X_i, \tag{2}$$

$$\begin{aligned}\mathrm{LSI} = {} & \mathrm{FR} \times \text{slope raster} \\ & + \mathrm{FR} \times \text{slope aspect raster} \\ & + \mathrm{FR} \times \text{sope curvature raster} \\ & + \mathrm{FR} \times \text{lithology raster} \\ & + \mathrm{FR} \times \text{land use raster} \\ & + \mathrm{FR} \times \text{distance to stream raster},\end{aligned}$$

Table 2: Logistic regression coefficients and multicollinearity diagnosis of conditioning factors used in LR

	Collinearity statistics		
Independent variables	**Logistic coefficients (B)**	**Tolerance**	**VIF**
Aspect	0.648	0.937	1.068
Curvature	0.198	0.900	1.111
Distance to lineament	1.296	0.891	1.122
Distance to drainage	1.896	0.617	1.620
Lithology	1.012	0.782	1.279
Land use	0.184	0.964	1.038
NDVI	0.067	0.881	1.135
Rainfall	0.191	0.834	1.199
Drainage density	0.278	0.602	1.660
SLOPE	0.719	0.758	1.319
Soil type	1.235	0.762	1.312
Constant	−8.905		

where LSI is the landslide susceptibility index, n is the number of landslide factors, X_i is the landslide factor, and FR_i is the FR of each landslide factor type or classes. After the LSI was calculated, the index values were classified into a different level of landslide susceptibility zones using natural breaks in the ArcGIS tool. The higher the value of the LSI, the higher the probability of landslide occurrence, but the lower the LSI, the lower the probability of landslide occurrence.

Based on the natural break classification, the landslide susceptibility map of the study area has five classes such as very low, low, moderate, high, and very high landslide susceptibility class (Figure 4b).

6.2 IV model

The IV method is one of the probabilistic methods of a bivariate statistical method, which is used to envisage the correlation between landslides and landslide factor classes [55]. The IVs for each factor class were determined through the combination of reclassified landslide raster to reclassified landslide factor raster based on the presence of landslide in a given map unit (Figure 4c). These values are important to define the role of each causal factor in classes for landslide occurrence. This can be calculated as in equation (4).

$$\mathrm{IV} = \ln\left(\frac{\text{Conditional probability (CP)}}{\text{Prior probability (PP)}}\right) = \frac{\frac{\mathrm{Nslpix}}{\mathrm{Ncpix}}}{\frac{\mathrm{Ntslpix}}{\mathrm{Ntcpix}}}, \tag{3}$$

where conditional probability is the ratio of the pixel of a landslide in class to the pixel of a class and prior probability is the ratio of the total number of pixels of landslide to the total number of pixels of the study area. Nslpix is a landslide pixel/area in a landslide factor class. Ntslpix is the total area of a landslide in the entire study area. Ncpix is the area of the class in the study area and Ntcpix is the total pixel area in the entire study area. When IV > 0.1, the landslide occurrence with the factor classes has a high correlation, which means it will have a high probability of landslide occurrence; however, when the IV <0.1 or IV <0, it is a low correlation between landslide factors and landslide occurrence, which indicates a low probability of landslide occurrence. After calculated the IV for each landslide factor class using Microsoft excel and GIS, the IV for each factor class assigned through the join in the ArcGIS tool. Then, the weighted landslide factors were rasterized using the lookup tool in spatial analysis, and the LSI of the study area is calculated as in equation (3).

Table 3: Spatial relationship between each landslide factors and landslide using FR, CF, and IV methods

Factors	Class	Ncpix	Nspix	% Ncpix (b)	%Nspix (a)	FR = a/b	PP_a	PP_b	CF	IV	LR
LULC	Scatter forest	1,20,151	591	4.139	8.290	2.003	0.005	0.002	0.502	0.695	0.184
	Woodland	5,91,791	899	20.388	12.610	0.619	0.002	0.002	−0.382	−0.480	
	Bush	3,94,218	918	13.581	12.877	0.948	0.002	0.002	−0.052	−0.053	
	Cropland	11,46,131	2,412	39.485	33.834	0.857	0.002	0.002	−0.143	−0.154	
	Grassland	3,94,509	1,372	13.591	19.245	1.416	0.003	0.002	0.295	0.348	
	Barre land	2,54,507	921	8.768	12.919	1.473	0.004	0.002	0.322	0.388	
	Settlement	1,373	16	0.047	0.224	4.745	0.012	0.002	0.791	1.557	
NDVI	0.1–0.4	3,20,744	225	11.050	3.156	0.286	0.001	0.002	−0.715	−1.253	0.067
	0.4–0.45	7,32,428	1,483	25.232	20.802	0.824	0.002	0.002	−0.176	−0.193	
	0.45–0.5	8,59,086	2,541	29.596	35.643	1.204	0.003	0.002	0.170	0.186	
	0.5–0.6	6,27,764	2,051	21.627	28.770	1.330	0.003	0.002	0.249	0.285	
	0.6–0.8	3,62,729	829	12.496	11.629	0.931	0.002	0.002	−0.070	−0.072	
Lithology	Alluvium soil deposit	5,05,536	0	17.416	0.000	0.000	0.000	0.002	−1.000	0.000	1.012
	Eluvium soil deposit	8,25,764	1,758	28.448	24.660	0.867	0.002	0.002	−0.133	−0.143	
	Fine to coarse sandstone	1,20,674	52	4.157	0.729	0.175	0.000	0.002	−0.825	−1.740	
	Fine to medium sandstone	1,30,651	134	4.501	1.880	0.418	0.001	0.002	−0.583	−0.873	
	Gypsum	8,46,403	5,145	29.159	72.170	2.475	0.006	0.002	0.597	0.906	
	Limestone	2,01,392	40	6.938	0.561	0.081	0.000	0.002	−0.919	−2.515	
	Lower basalt	86,341	0	2.974	0.000	0.000	0.000	0.002	−1.000	0.000	
	Middle basalt	1,85,973	0	6.407	0.000	0.000	0.000	0.002	−1.000	0.000	
Distance to lineament	0–1,000	3,87,844	1,991	13.362	27.928	2.090	0.005	0.002	0.523	0.737	1.296
	1,000–2,000	4,37,943	972	15.089	13.634	0.904	0.002	0.002	−0.097	−0.101	
	2,000–3,000	4,59,530	1,181	15.832	16.566	1.046	0.003	0.002	0.044	0.045	
	3,000–4,000	3,94,585	1,072	13.595	15.037	1.106	0.003	0.002	0.096	0.101	
	4,000–5,000	2,88,500	869	9.940	12.190	1.226	0.003	0.002	0.185	0.204	
	5,000–6,000	2,28,501	609	7.873	8.543	1.085	0.003	0.002	0.079	0.082	
	6,000–7,000	1,74,276	212	6.004	2.974	0.495	0.001	0.002	−0.505	−0.703	
	7,000–16,827	5,31,312	223	18.305	3.128	0.171	0.000	0.002	−0.829	−1.767	
Drainage density	0–0.3	16,27,264	3,260	56.064	45.729	0.816	0.002	0.002	−0.226	−0.204	0.278
	0.3–0.6	9,22,131	2,701	31.770	37.888	1.193	0.003	0.002	0.162	0.176	
	0.6–0.9	3,22,233	1,166	11.102	16.356	1.473	0.004	0.002	0.322	0.387	
	0.9–1.3	25,892	2	0.892	0.028	0.031	0.000	0.002	−30.800	−3.459	
	1.3–1.6	4,986	0	0.172	0.000	0.000	0.000	0.002	0.000	0.000	
Distance to drainage											
Curvature	0–200	3,35,052	1,158	11.544	16.244	1.407	0.003	0.002	0.290	0.342	1.846
	200–400	3,15,809	980	10.881	13.747	1.263	0.003	0.002	0.209	0.234	
	400–800	5,93,467	1,607	20.447	22.542	1.102	0.003	0.002	0.093	0.098	
	800–1,200	5,36,717	1,459	18.492	20.466	1.107	0.003	0.002	0.097	0.101	
	1,200–4,137	11,21,461	1,925	38.638	27.002	0.699	0.002	0.002	−0.302	−0.358	
	−6.7 to −0.6	1,55,774	697	5.367	9.777	1.822	0.004	0.002	0.452	0.600	0.198
	−0.4 to 0.3	25,46,217	5,442	87.725	76.336	0.870	0.002	0.002	−0.130	−0.139	
	0.3 to 8	2,00,515	990	6.908	13.887	2.010	0.005	0.002	0.504	0.698	
Aspect	Flat	0	0	0.000	0.000	0.000	0.000	0.002	−1.000	0.000	0.645
	N	90,209	67	3.108	0.940	0.302	0.001	0.002	−0.698	−1.196	
	NE	2,15,295	110	7.418	1.543	0.208	0.001	0.002	−0.792	−1.570	
	E	3,99,044	385	13.748	5.400	0.393	0.001	0.002	−0.608	−0.934	
	SE	5,05,127	995	17.403	13.957	0.802	0.002	0.002	−0.198	−0.221	
	S	4,22,902	1,543	14.570	21.644	1.485	0.004	0.002	0.328	0.396	
	SW	4,02,422	1,353	13.865	18.979	1.369	0.003	0.002	0.270	0.314	
	W	4,23,429	1,952	14.588	27.381	1.877	0.005	0.002	0.468	0.630	
	NW	3,35,102	628	11.545	8.809	0.763	0.002	0.002	−0.237	−0.270	

Table 3: continued

Factors	Class	Ncpix	Nspix	% Ncpix (b)	%Nspix (a)	FR = a/b	PP_a	PP_b	CF	IV	LR
	N	1,08,976	96	3.755	1.347	0.359	0.001	0.002	−0.642	−1.025	
Slope	<5°	12,85,973	999	44.306	14.013	0.316	0.001	0.002	−0.684	−1.151	0.719
	5–12°	7,25,015	1,581	24.979	22.177	0.888	0.002	0.002	−0.112	−0.119	
	12–19°	4,81,190	1,776	16.578	24.912	1.503	0.004	0.002	0.335	0.407	
	19–29°	2,81,637	1,461	9.703	20.494	2.112	0.005	0.002	0.528	0.748	
	29–62°	1,28,691	1,312	4.434	18.404	4.151	0.010	0.002	0.761	1.423	
Soil type	Eutric vertisols	1,58,642	135	5.466	1.894	0.346	0.001	0.002	−0.653	−1.060	1.235
	Humic nitisols	14,61,188	2,805	50.344	39.346	0.782	0.002	0.002	−0.218	−0.246	
	Rendzic leptosols	11,57,550	4,082	39.882	57.259	1.436	0.004	0.002	0.303	0.362	
	Hapzic paeozems	1,25,040	107	4.308	1.501	0.348	0.001	0.002	−0.651	−1.054	
Rainfall	968–1,117	3,10,624	84	10.701	1.178	0.110	0.000	0.002	−0.890	−2.206	0.719
	1,117–1,213	5,38,921	413	18.566	5.793	0.312	0.001	0.002	−0.687	−1.165	
	1,213–1,289	7,70,880	1,836	26.557	25.754	0.970	0.002	0.002	−0.030	−0.031	
	1,289–1,344	5,15,393	1,851	17.755	25.964	1.462	0.004	0.002	0.315	0.380	
	1,344–1,394	7,66,946	2,945	26.421	41.310	1.564	0.004	0.002	0.360	0.447	

Note: Nspix is landslide pixel/area in a landslide factor class; Ncpix is the area of the class in the study area.

$$\mathrm{LSI} = \sum_{n}^{i=1} \mathrm{IV}_i X_i, \tag{4}$$

LSI = IV × slope raster + IV × slope aspect raster
+ IV × slope curvature raster + IV × lithology raster
+ IV × land use raster
+ IV × distance to stream raster,

where LSI is the landslide susceptibility index and IV is the information value of each factor class. The higher value of LSI has indicated the higher probability of landslide occurrence.

6.3 CF model

The CF is one of the probabilistic methods that is widely used for landslide susceptibility mapping for different data [44,45,47,48]. Shortliffe and Buchanan [83] proposed the CF (the probability function) for landslide susceptibility mapping, and later Heckeman [84] improved it and it is expressed mathematically as:

$$\mathrm{CF} = \begin{cases} \dfrac{PP_a - PP_b}{PP_a(1 - PP_b)} & \text{if } PP_a \geq PP_b \\ \dfrac{PP_b - PP_b}{PP_b(1 - PP_a)} & \text{if } PP_a \leq PP_b \end{cases}, \tag{5}$$

where PP_a is the conditional probability of landslide in the defined area a and PP_b is the prior probability of landslide in the defined entire study area b. The CF value ranges from –1 to 1, a positive value indicates increasing certainty of landslide occurrence, and a negative value indicates decreasing certainty of landslide occurrence. If the certainty value is close to zero, it means there is no adequate information about the relation between landslide factor classes and landslide occurrence; therefore, it is difficult to give any certainty of landslide occurrence [48,85].

The CF values were calculated for all landslide factor classes through overlaying landslide factors with landslides using equations (5) and (6). After the calculation of CF for each landslide factor class, the LSI is determined as in equation (7).

$$\mathrm{Z} = \begin{cases} X + Y - XYX, Y \geq 0 \\ \dfrac{X + y}{1 - \min(|X|, |Y|)} X \times Y < 0, \\ X + Y + XYX, Y < 0 \end{cases} \tag{6}$$

where Z is the calculated CF value and X and Y are the two different layers of information.

$$\mathrm{LSI} = \sum_{n}^{i=1} \mathrm{CF}_i X_i, \tag{7}$$

LSI = CF × slope raster + CF × slope aspect raster
+ CF × slope curvature raster + CF × lithology raster
+ CF × land use raster
+ CF × distance to stream raster,

where LSI is the landslide susceptibility index and CF_i is the certainty factor.

6.4 LR model

The LR is one of the popular multivariate statistical analysis methods, which can be used to establish a multivariate regression relationship between dependent and independent variables [86]. Among other statistical methods, the LR model has been proven as one of the most reliable approaches for landslide susceptibility mapping to determine the most landslide influencing factors [87–92]. This model is advantageous, as it does not require normal distribution and it uses continuous or discrete variables. The difficulty in using the LR model lies in the sample size selection of dependent and independent variables for landslide susceptibility analysis. There are three ways of sampling landslide and non-landslide points [93]. The first way is using all data from all the study areas. However, this leads to an uneven proportion of non-landslide and landslide pixels, which incorporate a large volume of data in the analysis [94,95]. Using all landslide pixels with equal non-landslide pixels is the second method, which also results in a less reliable output, but it can reduce sample size and sampling bias. The third method uses an unequal or equal proportion of landslide and non-landslide pixels by classifying landslide into training and testing data sets [96,97].

In the present work, the landslides of the study area were classified into training landslides (70% with 502 landslides) and testing landslides (30% with 2,015 landslides). In this study, the dependent data are a binary variable and are made up of 0 and 1, which represent the absence and presence of landslides, respectively. Consequently, an equal number of non-landslide sample points (502), whose dependent variable value is 0 were randomly selected from landslide free areas to represent the absence of landslides using GIS. The 502 landslides and 502 non-landslides were merged. Moreover, all the values of independent variables containing landslides and non-landslides were extracted from the maps of each landslide-governing factors using ArcGIS. Then, the LR was conducted and coefficients were calculated in the SPSS program. It can be expressed mathematically [89,98] as:

$$P = \frac{1}{1 + e^{-z}}, \tag{8}$$

where P is the probability of landslide occurrence that varies from zero to one and Z is the linear combination of the predictors and varies from $-1 < z < 0$ for higher odds of non-landslide occurrence to $0 < z < 1$ for odds of higher landslide occurrence. Z can be defined as:

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots \beta_n X_n \beta_n X_n, \tag{9}$$

where x_1, x_2, x_3, and x_n are independent variables, β_0 is the intercept of the slope of LR analysis, and β_1, β_2, β_3, and β_n are the coefficients of the LR analysis.

6.5 Model validation

Landslide susceptibility map without validation has no sense in the scientific world [5]. Therefore, validation of the landslide susceptibility model is very important to evaluate the degree of accuracy of modeling using different validation techniques [50,99]. For this purpose, the landslide area is classified based on time, space, and random partition [10,14,50]. In this case, the landslide in the study area was classified into 70% (502) training landslide data sets and 30% (215) validation landslide data sets randomly keeping their spatial distribution. As stated by ref. [100], the area under the curve (AUC) value is used to evaluate the performance of the model and its value range from 0.5 to 1. When the AUC value is in between the range of 0.9–1, the model has excellent performance; if in between the range of 0.8–0.9, the model has very good performance; if in between the range of 0.7–0.8, the model has good performance; and if in between the range of 0.6–0.7, the model has an average performance. However, if the AUC values are in between the range of 0.5–0.6 and equal to 0.5 or less than 0.5, the model has poor performance [100].

In the present work, the landslide area was randomly classified as 70% landslide for training and 30% landslide for model validation by keeping their spatial distribution using the random partition technique [14,50]. After the model is developed, its accuracy was validated by ROCs and landslide density.

7 Results and discussions

7.1 Results

In this section, the results of the four landslide susceptibility models generated using FR, LR, IV, and CF

statistical methods are presented and compared. Table 3 shows the correlation between landslide locations and landslide driving factor classes, which was determined using FR (equation (1)), IV (equation (3)), and CF (equation (5)). The higher value of the FR, IV, and CF indicated the strong correlation between the landslide and landslide factor classes.

7.2 FR

To understand the significance of landslide factor classes for landslide occurrence, weight value was computed using FR methods as shown in equation (1) (Table 3). The FR for all landslide factor classes was rating and showed the statistical significance of each factor class on slope instability (Table 3). As it can be observed from Table 3, the lithology class gypsum produced a higher value of the FR (2.475) which is >1, indicating a high landslide probability occurrence, but the remaining lithological classes indicated weak statistical correlation and its FR value is <1, indicating a low probability of landslide occurrence. As presented in Table 3, the slope classes <5° and 5–12° have low FR value (0.316 and 0.888, respectively) and the slope classes 12–19°, 19–29°, and 29–62° have high FR value (1.503, 3.112, and 4.151, respectively). This correlation indicated that landslide probability increased as the slope gradient increased [5,10]. This is because of the presence of shallow loose soil deposit, highly weathered rock, active soil erosion, and improper land use practice. However, it may not be always true when the steep slope comprises massive and strong slope material. In the case of the aspect factor class, the FR value is >1 for south facing (1.485), southwest facing (1.369), and west facing (1.877), indicating high landslide probability. However, the remaining slope aspect classes have FR value <1, indicating a low probability of landslide occurrence. The FR value of the slope curvature classes of −6.7 to −0.6 (1.822) and 0.3–8 (2.01) is >1, indicating high landslide probability occurrence. This is because of the effects of slope shape for rainwater impounding and gravity effect. However, the FR value of the slope curvature class of −0.4 to 0.3 (0.87) is <1, indicating a low probability of landslide occurrence. As presented in Table 3, as the distance to stream increased, the probability of landslide occurrence decreased. At a distance of 0–200, 200–400, 400–800, and 800–1,200 m, the value of the FR (1.407, 1.263, 1.102, and 1.107) is >1, indicating high landslide probability; however, at a distance >1,200 m, the value of the FR is <1, indicating low landslide probability. This is because of the effects of slope modification, gully erosion, riverbank erosion, and river undercutting. As noticed in Table 3, the value of the FR for land use/cover class of grassland (1.416), settlement (4.745), bar land (1.473), and scatter forest (2) is >1, indicating high landslide probability. This is because of the scatter forest, grassland can increase soil moisture. As the soil moisture increased in slope material, the weight of slope material as well as the pore water pressure in slope material increased in parallel [5]. This can result in a reduction in the normal force in the soil mass. This leads to slope failure when the driving force exceeds a resisting force. In the case of bare land class, FR value showed a higher correlation with the probability of landslide occurrence. Therefore, bare land in the study area was highly affected by gully erosion, which caused a reduction in shear strength of soil material. The remaining classes have FR value <1, indicating a low probability of landslide occurrence.

7.3 IV

The IV rating for different landslide factor classes was calculated by overlaid landslide raster with landslide factor, and it showed the significant effect of each factor class on slope instability (Table 3). When the IV value is >0.1, the given factor class will have a positive correlation for landslide occurrence, but the IV <0.1 indicates a low probability of landslide occurrence. As presented in Table 3, the IV >0.1 for lithology class such as gypsum (0.906) indicated high landslide probability, but the IV <0.1 for the remaining lithology class indicated a low probability of landslide occurrence. As observed in Table 3, the IV <0.1 for slope classes <5° and 5–12° (IV = −1.151 and −0.119, respectively) indicated low landslide probability, and the IV >0.1 for slope classes 12–19°, 19–29°, and 29–62°, respectively (IV = 0.407, 0.748, and 1.423), indicated high landslide probability. In the case of slope aspect factor class, the IV >0.1 for south facing (IV = 0.396), southwest facing (IV = 0.314), and west facing (IV = 0.630) indicated high landslide probability. However, IV <0.1 for the remaining slope aspect classes indicated a low probability of landslide occurrence. The IV >0.1 for the slope curvature class of −6.7 to −0.6 (IV = 0.6) and 0.3–8 (IV = 0.698) indicated high landslide probability. However, the IV <0.1 for the slope curvature class −4 to 0.3 (IV = −0.139) indicated low probability of landslide occurrence. At a distance of 0–200, 200–400,

400–800, and 800–1,200 m, the value of the IV >0.1, which is 0.342, 0.234, 0.095, and 0.101, indicated high landslide probability; however, at a distance of 1,200 m, the IV <0.1 indicated low landslide probability. As noticed in Table 3, the value of IV for land use/cover class of grassland (0.348), settlement (1.557), bar land (0.388), and scatter forest (0.6) is >0.1, indicating high landslide probability. The IV for the remaining factor classes is <0.1, indicating a low probability of landslide occurrence.

7.4 CF

The CF rating for different landslide factor classes was calculated by overlaid landslide raster with landslide factor using equations (5) and (6), and it showed a significant effect of each factor class on slope instability. As presented in Table 3, the lithology class such as gypsum has a positive and high value of CF (0.597), indicating high landslide probability, but the remaining lithology class has a negative CF value and indicated low probability of landslide occurrence. As observed in Table 3, the slope classes <5° and 5–12° produced negative CF value (CF = −0.684 and −0.112, respectively), indicating low landslide probability. For slope classes 12–19°, 19–29°, and 29–62°, the CF value is positive and relatively higher (CF = 0.335, 0.528, and 0.761, respectively), indicating high landslide occurrence probability. In the case of aspect factor class, the CF value is positive for south facing (0.328), southwest facing (0.27), and west facing (0.468), indicating high landslide occurrence probability. However, the remaining slope aspect classes have negative CF value, indicating a low probability of landslide occurrence. The CF value of the slope curvature class of −6.7 to −0.6 (0.452) and 0.3–8 (0.504) is positive, indicating high landslide occurrence probability. However, the slope curvature class −0.4 to 0.3 has a negative CF value (−0.130), indicating a low probability of landslide occurrence. At a distance of 0–200, 200–400, 400–800, and 800–1,200, the value of CF (0.29, 0.209, 0.093, and 0.097) is positive, indicating high landslide occurrence probability; however, a distance >1,200 m produced negative CF value, indicating low landslide probability. As noticed in Table 3, the value of CF for land use/cover class of grassland (0.295), settlement (0.791), bar land (0.322), and scatter forest (0.5) is positive, indicating high landslide probability. The remaining factor classes produced negative CF value, indicating a low probability of landslide occurrence.

7.5 LR

The LR was used for the modeling of landslide susceptibility in the study area. Before the LR analysis, it is essential to perform collinearity analysis to evade possible multicollinearity among the independent variables. For these purposes, two main indexes of Tolerance (TOL) and the variance inflation factor (VIF) were applied to measure the multicollinearity [67]. When the TOL is less than 0.1 or the VIF greater than 10, it indicates that a multicollinearity problem exists. As shown in Table 2, the TOL and VIF values of all independent variables have no multicollinearity problem among all independent variables. After collinearity analysis, the relationship between the dependent and independent variables was evaluated using SPSS software to obtain the logistic coefficients of landslide governing factors under forwarding stepwise. Then, the value of Z was calculated by entering the LR coefficients of landslide governing factors in equation (10).

$$\begin{aligned} Z = {} & -8.905 + 0.648 \times \text{aspect raster map} \\ & + 0.198 \times \text{curvature raster map} \\ & + 1.296 \times \text{distance to lineament raster map} \\ & + 1.896 \times \text{distance to drainage raster map} \\ & + 1.012 \times \text{lithology raster map} \\ & + 0.184 \times \text{LULC raster map} \\ & + 0.067 \times \text{NDVI raster map} \\ & + 0.191 \times \text{rainfall raster map} \\ & + 0.278 \times \text{drainage density raster map} \\ & + 0.719 \times \text{slope raster map} \\ & + 1.235 \times \text{soil raster map}. \end{aligned} \tag{10}$$

The calculated Z values of the study area range from −5.8 to 5.4. After calculating the values of Z, the probability landslide occurrence P of the study area was calculated by entering the values of Z to equation (8), which range from 0.0021 to 0.996. Finally, the landslide susceptibility map of the study area was classified into five classes by classifying the P values using the natural break classification method in ArcGIS: very low, low, moderate, high, and very high landslide susceptibility (Figure 4a). As shown in Table 2, all landslide factors indicated a positive relationship for landslide occurrence. For example, distance to drainage, distance to lineament, soil type, and lithology have higher LR coefficients compared to others, which indicates a strong statistical correlation between landslide factors and landslide occurrence. Based on the sample data, the goodness of fit model was evaluated through Nagelkerke's R-square, Homer–Lemeshow test, and the percentage

rate of classifications with a cutoff of 0.5. From the analysis, the Nagelkerke's R-square value is 0.487 and the model correctly classifies 76.7% of the landslides and 76.1% of the non-landslides when a cutoff of 0.5 is considered. All these results indicate that the present study model performed reasonably well.

7.6 Landslide susceptibility mapping

All weighted landslide-governing factors were summed using a raster calculator in ArcGIS in terms of equations (2), (4), (7) and (8) to obtain LSI. After calculating the LSI, it is important to classify the LSI into different susceptibility classes based on the LSI value. The LSI map of the study area, which was generated using LR method, IV method, CF method, and FR method, was classified into five levels of susceptibility classes (very low, low, moderate, high, and very high) using the natural break method (Figure 4a–d). From the results of the IV analysis (Table 4), 6.6 and 17.7% of the study area fell in very low and low susceptibility classes, respectively. Moderate, high, and very high landslide susceptibility classes comprised 31.7, 29.4, and 14.6% of the study area, respectively. As presented in Table 4, 0.1 and 0.6% of the validation landslides fell in very low and low susceptibility classes of the study area, respectively. The remaining 2.9, 17.2, and 79.3% of validation landslides fell into moderate, high, and very high landslide susceptibility classes, respectively. From the results of landslide susceptibility map produced using CF model (Table 4), very low and low susceptibility classes cover 13.4 and 26.7% of the total study area, however, 29, 22, and 9% of the total area fell into moderate, high and very high landslide susceptibility classes, respectively. As indicated in Table 4, 0 and 0.7% of the validation landslides fell in very low and low susceptibility classes of the study area, respectively. The remaining 5.8, 27.5, and 66% of validation landslides fell into moderate, high, and very high landslide susceptibility classes, respectively. As it observed from Table 4, the landslide susceptibility map produced using the FR model, very low and low landslide susceptibility classes covered 18.9 and 28.9% of the total area; however, 25.6, 20, and 6.5% of the total area fell into moderate, high, and very high landslide susceptibility classes, respectively. As presented in Table 4, 0 and 1.5% of the landslide fell in very low and low susceptibility classes of the study area, respectively. The remaining 11.7, 32.2, and 54.5% of validation landslides fell into moderate, high, and very high landslide susceptibility classes, respectively. The landslide susceptibility map, which produced using the LR method, covered 41.9% of a region by very low susceptibility class, 21.1% by low susceptibility class, 16% by moderate susceptibility class, 12.6% by high susceptibility class, and 8.4% by very high susceptibility class (Table 4). Besides, 0.6 and 3.4% of the landslide fell in very low and low susceptibility classes in the study area, respectively. The remaining 15.8, 19, and 61.2% of validation landslides fell into moderate, high, and very high landslide susceptibility classes, respectively (Table 4).

7.7 Model validation and sensitivity analysis

In this research, the ROC, the AUC, and landslide density were used to evaluate the accuracy of the landslide susceptibility model generated by LR, FR, IV, and CF methods. The four models were validated by the researcher experience in the area and comparing the existing training and validation landslide data sets with the produced landslide susceptibility maps. Both the success rate and prediction rate curves were calculated using training landslide data sets and validation/testing landslide data sets, respectively. The success rate curve can show how well the models classified the region based on the existing landslide events [14,102]. The prediction rate curve can show how well the models can predict the unknown forthcoming landslide events [101,102]. In this study, the success rate and prediction rate curves were calculated by reclassifying the LSI values into 100 for all cells and sorting in descending order and compared with both training and validation landslide data sets. Finally, the AUC and ROC curves for the four models were calculated using Real Statistics software. As the results of the analysis showed in Figure 6 and Table 4, the closer the ROC curve to the left of the top of the curve indicated the higher the accuracy of the model. As indicated in Table 4, the AUC value is closer to one, indicating the higher accuracy of the model. The training landslide data set was used to calculate the success rate (Figure 6b), which showed a success rate of 91.6% using the FR model that is better than the success rate of 90.5% using the CF model, 90.2% using the IV model, and 84.4% using the LR model. The validation landslide data set was used to measure or evaluate the prediction rate of the models (Figure 6a), which showed a prediction rate of 93.9% using the CF model that is better than the prediction rate of 93.2% using the IV model, 92.7% using FR model, and 87.9% using LR model. Although all methods showed little difference in

Table 4: Statistical results of landslide susceptibility map produced by FR, IV, and CF methods

	LSI value	LSI	Susceptibility class area (%)	Validation data set (a) (%)	Landslide density (a)	Training data set (%)	Landslide density (b)	AUC for validation landslide	AUC for training landslide
Information value method	−12 to −5	VLS	6.6	0.1	0.0	0.4	0.0	0.932	0.902
	−5 to −3	LS	17.7	0.6	0.0	2.5	0.1		
	−3 to −0.9	MS	31.7	2.9	0.1	14.7	0.6		
	−0.9 to 0.9	HS	29.4	17.2	0.6	34.7	1.7		
	0.9 to 6.2	VHS	14.6	79.3	5.4	47.7	7.3		
CF	−6 to −3	VLS	13.4	0.0	0.0	0.3	0.0	0.939	0.905
	−3 to −1.4	LS	26.7	0.7	0.0	1.2	0.0		
	−1.4 to −0.3	MS	29.0	5.8	0.2	12.6	0.4		
	−0.3 to 0.9	HS	22.0	27.5	1.3	28.1	1.3		
	0.9 to 4.3	VHS	9.0	66.0	7.4	57.8	6.4		
FR	5 to 9	VLS	18.9	0.0	0.0	0.41	0.1	0.927	0.916
	9 to 11	LS	28.9	1.5	0.1	1.05	0.1		
	11 to 12.3	MS	25.6	11.7	0.5	6.54	0.2		
	12.3 to 14.3	HS	20.0	32.2	1.6	21.08	0.7		
	14.3 to 20.7	VHS	6.5	54.5	8.3	70.92	4.9		
LR	0 to 0.2	VLS	41.9	0.6	0.0	1.89	0.0	0.879	0.844
	0.2 to 0.3	LS	21.1	3.4	0.2	7.01	0.3		
	0.3 to 0.5	MS	16.0	15.8	1.0	18.01	1.1		
	0.5 to 0.7	HS	12.6	19.0	1.5	21.60	1.7		
	0.7 to 0.996	VHS	8.4	61.2	7.3	51.48	6.1		

LSI is landslide susceptibility index, VLS is very low susceptibility, LS is low susceptibility, MS is moderate susceptibility, HS is high susceptibility, VHS is very high susceptibility, and AUC is the area under the curve

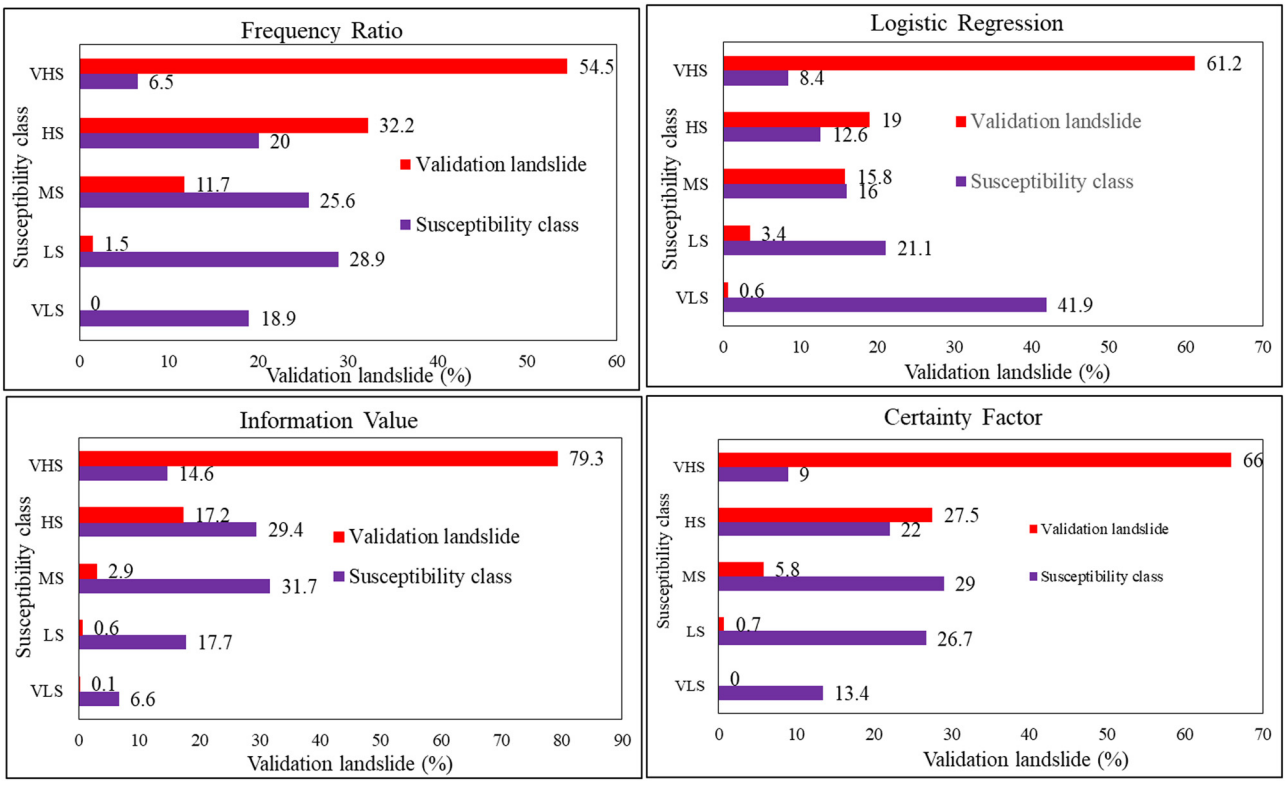

Figure 11: Spatial distribution of predicted and observed landslides.

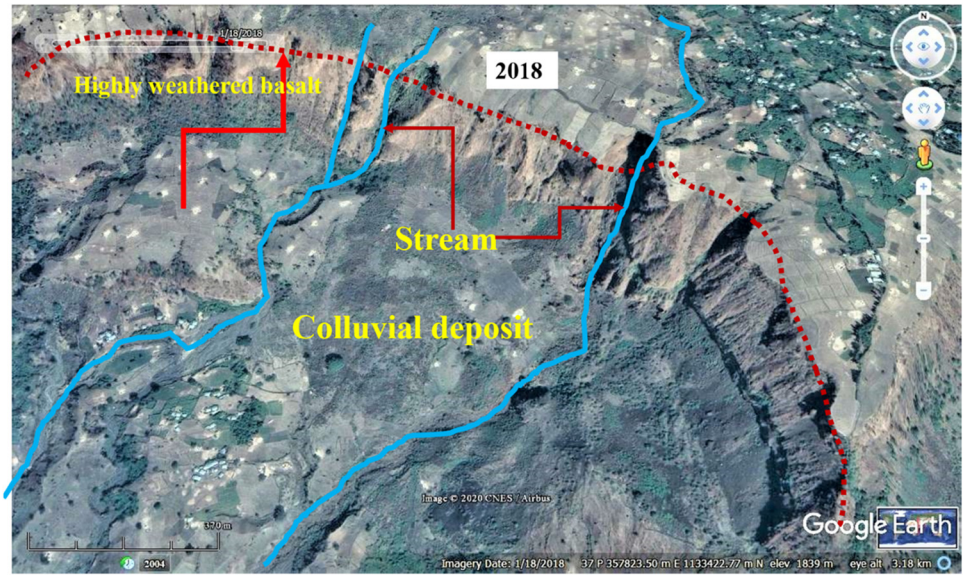

Figure 12: Time series Google Earth image showing unfailed slope at Desa Enese village in 2018 [80].

AUC values, it is evident that the four methods can serve as an effective method to perform landslide susceptibility assessment. Moreover, the AUC values of all methods except LR (very good) fell in the same range of excellent performance. These results indicated that the FR, CF, IV, and LR models have successfully estimated the landslide susceptibility classes of the region, and these models, which were used in this study, have reasonable accuracy in predicting the landslide susceptibility classes of the study area. For the four models,

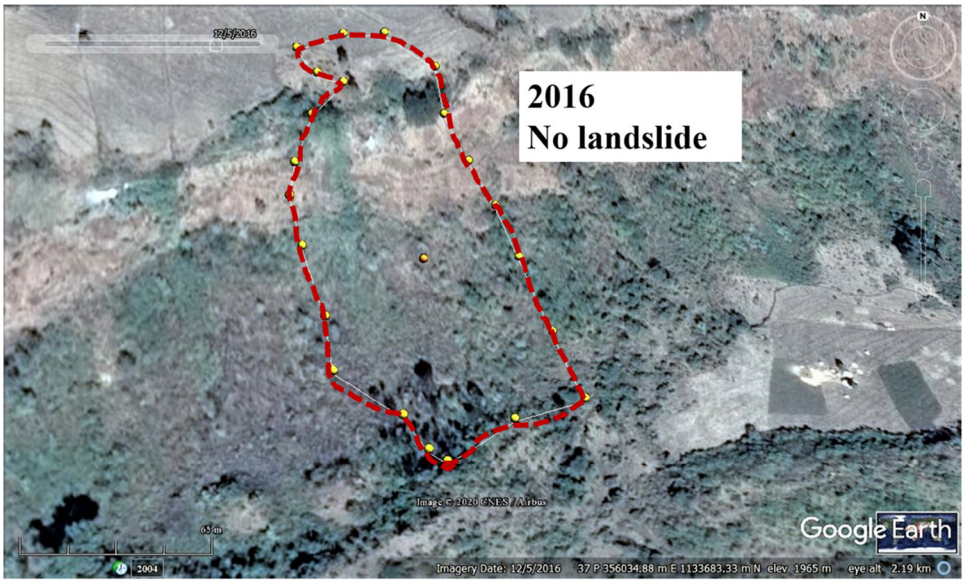

Figure 13: Time series Google Earth image showing unfailed slope at Aba Libanos village in 2018 [80].

greater than 80% of validation landslides fell in high and very high susceptibility classes (Figure 11), and the landslide density increased as the landslide susceptibility increased (Figure 5) which again confirms that the models have higher accuracy. However, based on AUC values, CF, IV, and FR models revealed a little better result than the LR model for landslide susceptibility mapping in the study area (Figures 6 and 11). As shown in Table 4 and Figure 11, the areas with high and very high landslide susceptibility cover 21, 26.5, 31, and 44% of the study area by LR, FR, CF, and IV models, respectively, and 96.5% for IV, 93.5% for CF, 86.7% for FR, and 80.2% of LR models of validation landslides occurred in these areas, whereas <5% of the validation landslides falls in the areas with very low and low landslide susceptibility class. From these, the IV model has overestimated the high and very high landslide susceptibility of a region followed by CF models. For the training landslide data set, a higher percentage that ranges from 73 to 93.9% is correctly distributed in the areas with high and very high landslide susceptibility, and a lower percentage of <14% of the training landslide data set falls in the areas with very low and low landslide susceptibility. Thus, it can be concluded that the proposed models produced an accurate prediction of the spatial probabilities of landslide occurrence in the study area.

To determine the statistical significance of the models, a statistical significance test between the models was performed (Table 5) using LR analysis by SPSS software, for both training and testing landslide data sets with an equal number of pixels of non-landslides. The results presented in Table 5 show that the LR and CF models have indicated higher statistical significance than the FR and IV methods. In this study, a factor sensitivity analysis was performed to determine the sensitivity of the landslide governing factors for landslide susceptibility mapping. It was conducted by (1) removing each landslide factor during the summation stage using FR under raster calculator and (2) validating the resulted map using success rate curves, and the success rate value was obtained by subtracting from the success rate value of the map, which produced using all landslide-governing factors. From the results of sensitivity analysis (Table 5), soil type, slope, land use land cover (LULC), curvature, distance to drainage, drainage density, aspect, and distance to lineament indicated a positive effect on landslide susceptibility mapping with the decrease of 1.3, 1.26, 1.1, 0.9, 0.7, 0.3, 0.1, and 0.1%. However, lithology, NDVI, and rainfall with −0.2, −0.5, and −0.85% of increments of success rate value from the model, which produced using all landslide-governing factors when they remove from modeling so they have a negative effect on landslide susceptibility mapping in the study area. The sensitivity analyses showed how a solution changes as the input landslide governing factors are changed (Table 5). The higher the values of decrease of success rate, the higher sensitivity of landslide factors to landslide susceptibility mapping. From the results, it can be concluded that the performance of landslide susceptibility

mapping is affected not only by landslide susceptibility mapping methods but also by incorporated landslide governing factors.

7.8 Discussion

Landslide susceptibility maps can forecast/provide important information about the occurrence of landslides in a region. This is a function of the relationship between preexisting landslide and the environmental condition of the area. These maps also show the spatial distribution of predicted landslides where it will occur. However, the maps cannot forecast the volume of material to displace, time, and return period of the landslides. Nevertheless, the predictive models can be important for the regional land use planning of landslide hazard mitigation and prevention relief [60,61,101–105]. The landslide susceptibility maps of the study area are classified into fivefold classification schemes of very low, low, moderate, high, and very high susceptibility classes using the natural break method, which is applicable to classify unevenly distributed data, and it is capable of classifying LSI map into different categories considering the inherent data value similarity. The resulted map's accuracy was validated using training and testing/validation landslide data sets through the success rate curve

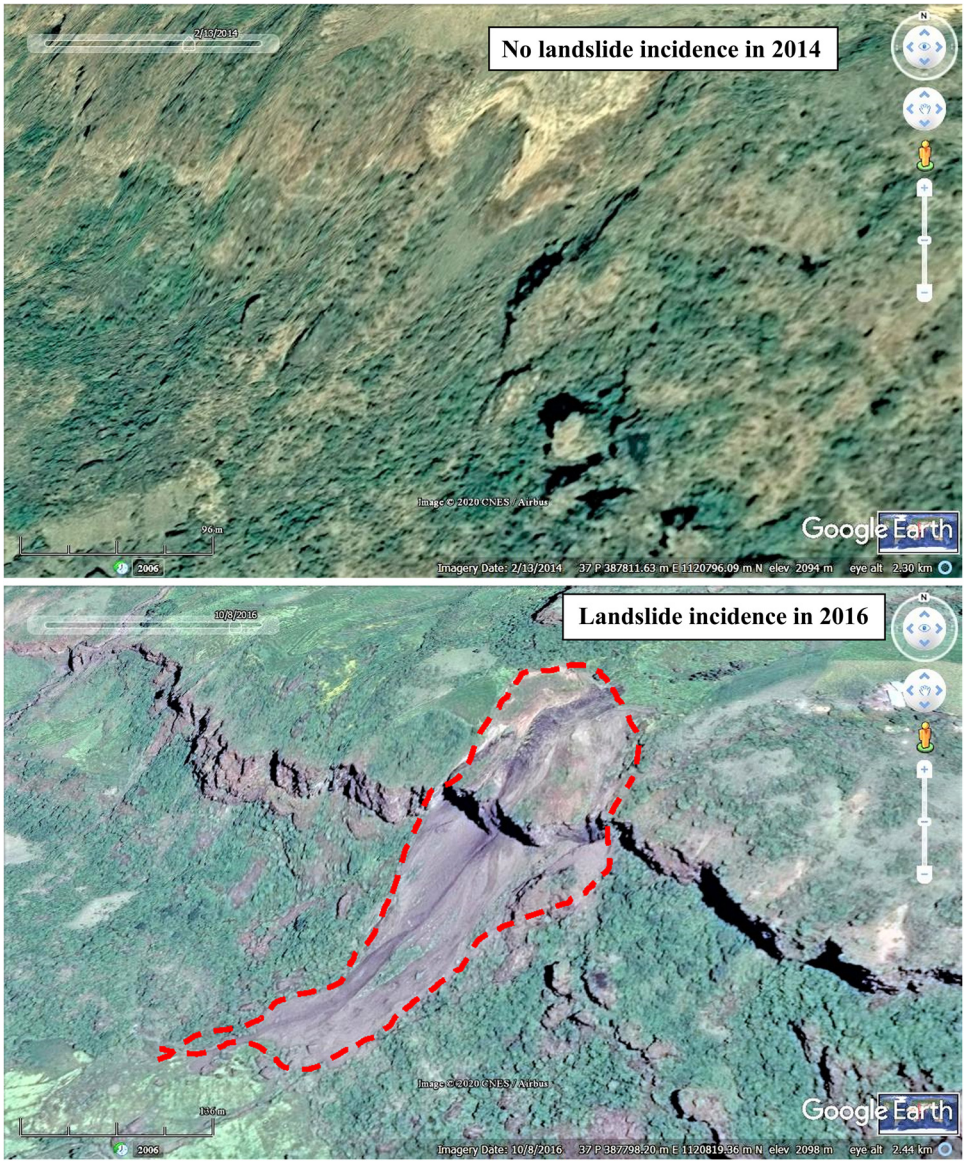

Figure 14: Time series Google Earth image showing unfailed slope at Gobetima village in 2014 and failed slope in 2016.

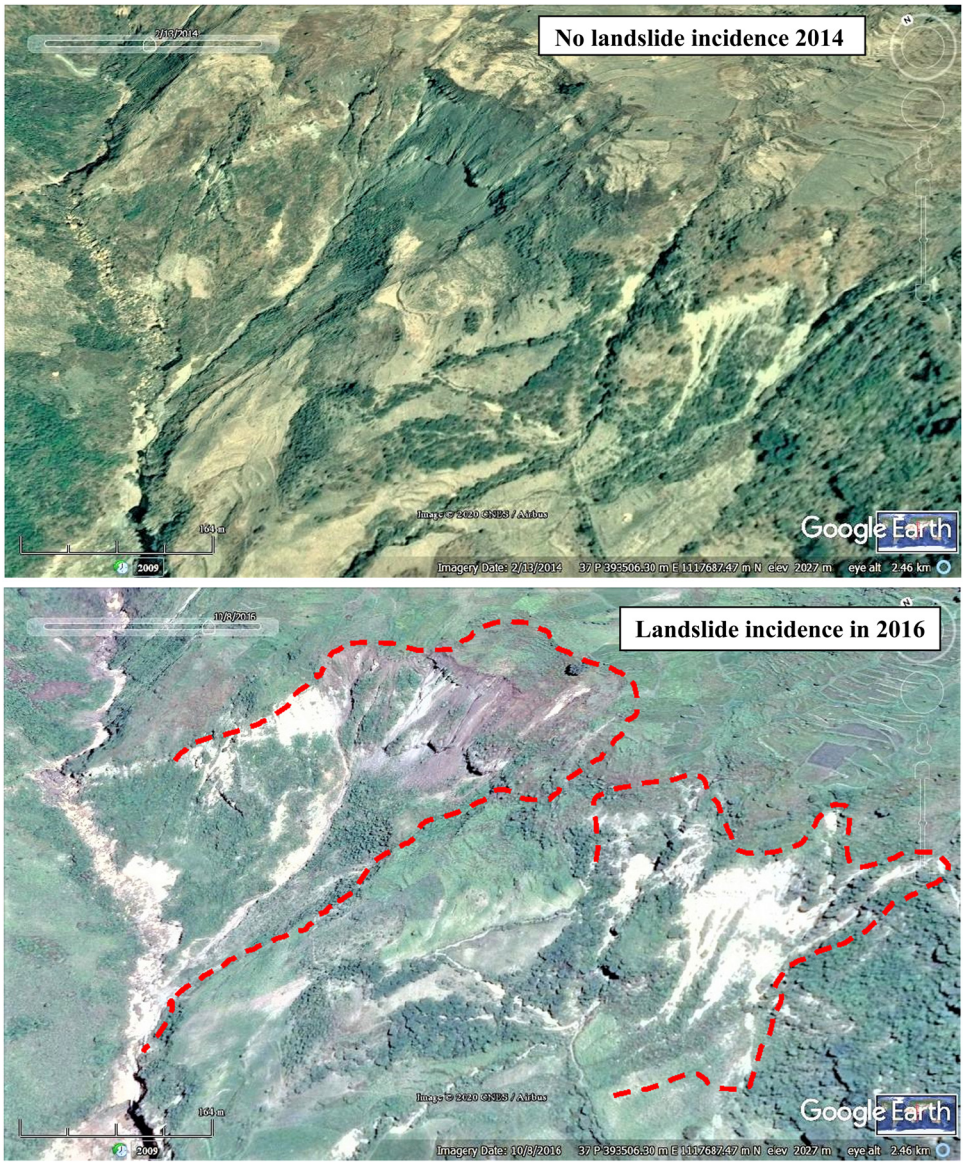

Figure 15: Time series Google Earth image showing unfailed slope at Gobetima village in 2014 and failed slope in 2016.

and predictive rate curve. The success rate curves for the three models were generated from the training landslide data sets through combining tools with landslide susceptibility classes, which was used to evaluate how well the models classified the region based on the existing landslide events [14,105]. The prediction rate curve for the four models was generated from the validation landslide data sets through combining tools with landslide susceptibility classes used to evaluate how well the models can predict the unknown forthcoming landslide events [101,105]. High and very high susceptibility classes in the region are fell in a steep slope, which covered with very lose shallow soil deposit, closer to the stream, agricultural land on a steep slope, active gully erosion, and concave slope shapes while the moderate susceptibility class fell in the area of highland landscapes. Low and very low susceptibility of a region fell in the area of low plain landscapes and areas, which are covered by massive weathering resistant rock masses.

Although FR and IV methods can quantitatively reflect the influence of numerous landslide governing factor classes to landslide occurrence probability, it does not consider the important degree of the factors to landslide occurrence [112]. However, LR and CF methods help to describe the relationship between a landslide occurrence and the landslide-governing factors [59]. Nevertheless, except for the CF, it does not analyze the influence of

Table 5: Statistics of the sensitivity analysis of landslide governing factors and statistical test results between landslide susceptibility models

Expected factors	Success rate value (%)	Decreased success rate value (%)	Landslide susceptibility models	Significant coefficients for training landslide data sets	Significant coefficients for testing landslide data sets
Aspect	91.5	0.1	Certainty factor	1.003	0.601
Curvature	90.7	0.9	Frequency ratio	−0.188	0.126
Distance to lineament	91.5	0.1	Information value	−0.178	−0.06
Distance to drainage	90.9	0.7	Logistic regression	3.75	0.698
Lithology	91.8	−0.2			
Land use	90.5	1.1			
NDVI	92.1	−0.5			
Rainfall	92.45	−0.85			
Drainage density	91.3	0.3			
SLOPE	90.34	1.26			
Soil type	90.3	1.3			

classes of each landslide-governing factor on landslide occurrence [61,73,81,108], and the process of input data, calculation, and output is time consuming [111,112]. Therefore, in landslide susceptibility mapping, using integrated methods are very important to solve the limitation among the statistical methods. For example, the FR and IV methods are incapable to determine the effects of landslide factor on landslide occurrence probability rather than evaluating the effects of landslide factor class, which can be solved using LR methods that are incapable to determine the effects of landslide factor classes. There is some literature regarding the comparison of the FR method with the IV method, LR method with IV and frequency methods, the CF method with the IV method, and the CF method with the FR method. Reference [78] states that the prediction rate of 89.05% using the IV is better than the prediction rate of 85.57% using the FR method. However, in this study, the FR method showed better performance for both success rates (AUC = 91.6%) and predictive rate curve (AUC = 92.7%) than the IV method with success rate curve (AUC = 90.2%) and predictive rate curve (AUC = 93.2%). Even though the FR model showed a little bit difference in AUC value in general, the accuracy of the two models fell in the same ranges, which is a very good performance. As shown from the work of ref. [58], the CF model showed a high predictive accuracy of AUC value of 75% compared to the IV model with prediction rate curve value (AUC = 64.08%), but their accuracy value fell in the same ranges, which is an excellent performance. Nevertheless, in the present model, the CF model showed a relatively few difference in prediction rate value (AUC = 93.9%) than the IV model with prediction rate value (AUC = 93.2%), but they have the same accuracy range, which is an excellent performance. From the work of ref. [68], based on the predictive rate value of the area under the receiver operating characteristic curve (AUC), the FR and CF models showed more or less similar predictive capacity, which is 81.18% for the CF model and 80.14% for the FR model. However, CF showed a bit of little performance than the FR model. In the present work, the prediction rate of 92.7% using the FR model is better performance than the prediction rate curve (93.9% for the CF model). Refs. [110,111] found that the IV method gave a more realistic landslide susceptibility map than the LR model. This result is the same as the work of ref. [5,106], which stated that the IV method has higher prediction accuracy than the LR. Similarly, in this study, the IV method has higher prediction accuracy (Figure 6. From the work of ref. [14], the prediction accuracy is a little better in the LR method (AUC = 74.5%) than the FR method (AUC = 73.7%). However, ref. [109] found that

the FR (AUC = 79.48%) has a higher prediction accuracy than the LR (AUC = 77.4%). This result is similar to the work of ref. [107], which stated that the prediction accuracy of FR (AUC = 81.4%) is better than the LR (75.1%). This result is the same, as the prediction rate of FR is better than the LR in the present study. Generally, the three bivariate statistical and multivariate statistical methods in the literature and this study showed the closer prediction capacity with AUC > 64% and AUC > 80%, respectively, which fell in the range of good and very good/excellent performance [100]. The landslide validation results for three models (FR, IV, and CF) are closer to each other, it fell in the same range of excellent performance, and LR produced a very good performance. Besides this, the percent of landslides that fell in the high and very susceptibility classes are more than 80% validation landslides. Therefore, from these results, the research work finds out that in landslide susceptibility mapping, the four models have equal potential to generate landslide-prone areas but factor selection should be playing a more important role than the methods. Nevertheless, in a specific case, the high and very high susceptibility area coverage of the IV model showed high differences compared to the CF, FR, and LR methods. This is because of the problems ascertained in IV during weight rating for each factor class, i.e., when no landslide exists in a certain factor class, the results of IV becoming zero. This brings an impact on the overall accuracy of the model. Although all statistical models indicated higher prediction accuracy, based on their statistical significance analysis result (Table 5), the LR model is relatively better followed by the CF model for regional land use planning, landslide hazard mitigation, and prevention purposes.

8 Conclusion

The study area is characterized by recent unconsolidated soil deposits, rugged topography, active gulley, and riverbank erosion, and improper land use practice that makes it very prone to different landslides, including soil slide, weathered rockslide, debris flow, earth flow, earthfall, and soil creep. Based on the LR coefficients, the impacts of 11 landslide factors (distance to drainage, distance to lineament, soil type, lithology, slope, aspect, drainage density, curvature, rainfall, land use, and NDVI with decreased order of coefficients) on landslide occurrence are determined. A landslide is the most serious natural hazard in the study area. To determine the landslide susceptibility–prone areas, FR (FR), CF, LR, and IV models were applied. The landslide susceptibility maps of the study area are categorized into very low, low, moderate, high, and very high susceptibility classes. The high and very high susceptibility classes are high in the villages including Tedem, Dendo, Anjimo, Dinguabe, Denba, Kok, Desa Enese, Moching, Yewebi Enefoch, Aratu Amba, Aba Libanos, Denba, and Kebi in order of decreasing the risk of landslide incidence because of the presence of active riverbank erosion, lose soil deposit, high concentration of stream density, and undulating topography. Therefore, these areas need to slope vegetation and water management tasks. The accuracy of the landslide susceptibility models was evaluated using the ROCs curve and landslide density through comparison of training and validation landslide raster with the models. The prediction rate curve value of AUC for the four models is closing in 1, indicating very good accuracy of the models. Based on the AUC value of the results and >80% of observed validation landslides which fell in high and very high susceptibility classes, the statistical methods can prove as the most economical and effective methods in landslide susceptibility mapping in the similar regions to the study area. The models, which were generated using the four statistical models, can help to understand the landslide hazard problems in the study area. Although the resulting maps cannot forecast the time, and how often it can occur, it has provided the spatial distribution of landslide probability. These models can also provide important information to the researchers, local people, government, and planners to reduce the landslide hazard problems in the study area. Therefore, the concerned bodies may at the Wereda/District, Zone, Region, and Federal levels take tangible activities to mitigate the landslide problem by afforestation of the high and very high regions with the integration of terracing and construction of check dams for streams, gabion, and retaining walls along the riverbanks.

Acknowledgments: First, the author would like to thank the almighty God who allowed him to accomplish this research work. Next, the author would like to give his gratitude to his lovely wife Ms. Seble Asmare and his brother Mr. Zemene Wubalem for their valuable support and advice during this research work. The author would like to give his special thanks to Dr. Veera and his friends for their continuous support during the research work. Finally, the author would like to thank the University of Gondar for geological equipment, the National Meteorological Agency staffs, the GSE, the disaster head office, natural resource management head office, and the rural community, for their valuable data.

Author contributions: Starting from the conception and design of the work, modeling, as well as the statistical analysis and interpretations of the results were done by the author.

References

[1] Brunsden D. Mass movement. In: Embleton C, Thornes J, editors, Processes in Geomorphology. London, UK: Edward Arnold Ltd; 1979. p. 436.

[2] Cruden DM. A simple definition of a landslide. Bull Eng Geol Environ. 1991;43(1):27–9.

[3] Ambrosi C, Strozzi T, Scapozza C, Wegmuller U. Landslide hazard assessment in the Himalayas (Nepal and Bhutan) based on Earth-Observation data. Eng Geol. 2018;237:217–28.

[4] Woldearegay K. Review of the occurrences and influencing factors of landslides in the highlands of Ethiopia with implications for infrastructural development; 2013.

[5] Azemeraw W, Meten M. Landslide susceptibility mapping using information value and logistic regression models in Goncha Siso Eneses area, northwestern Ethiopia. SN Appl Sciences, Switz AG. 2020;2:807. doi: 10.1007/s42452-020-2563-0.

[6] Hong H, Miao Y, Liu J, Zhu AX. Exploring the effects of the design and quantity of absence data on the performance of random forest-based landslide susceptibility mapping. Catena. 2019;176:45–64.

[7] Aleotti P, Chowdhury R. Landslide hazard assessment: summary review and new perspectives. Bull Eng Geol Environ. 1999;58:21–44.

[8] Gutiérrez F, Linares R, Roqué C, Zarroca M, Carbonel D, Rosell J, et al. Large landslides associated with a diapiric fold in Canelles reservoir (Spanish Pyrenees): Detailed geological–geomorphological mapping, trenching, and electrical resistivity imaging. Geomorphology. 2015;241:224–42.

[9] Jazouli E, Barakat A, Khellouk R. GIS-multicriteria evaluation using AHP for landslide susceptibility mapping in Oum Er Rbia high basin (Morocco); 2019.

[10] Lee S, Pradhan B. Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models, Landslides. 2007;4:33–41.

[11] Temesgen B, Mohammed U, Korme T. Natural hazard assessment using GIS and remote sensing methods, with particular reference to the Landslides in the Wondogenet Area, Ethiopia. Phys Chem Earth Part C: Sol Terrestrials & Planet Sci (C). 2001;26:615–65.

[12] Woldearegay K. Characteristics of a large-scale landslide triggered by heavy rainfall in Tarmaber area, central highlands of Ethiopia. Geophys Res. 2008;10.

[13] Ibrahim J. Landslide assessment and hazard zonation in Mersa and Wurgessa, North Wollo, Ethiopia. Master thesis, School of Graduate Studies. Addis Ababa, Ethiopia: Addis Ababa University; 2011.

[14] Meten M, Bhandary NP, Yatabe R. GIS-based frequency ratio and logistic regression modeling for landslide susceptibility mapping of Debre Sina area in central Ethiopia. J Mater Sci. 2015;12(6):1355–72.

[15] Rai PK, Mohan K, Kumra VK. Landslide hazard and its mapping using remote sensing and GIS. J Sci Res. 2014;58:1–13.

[16] Chen Z, Wang J. Landslide hazard mapping using a logistic regression model in Mackenzie Valley. Can Nat Hazard. 2007;42(1):75–89.

[17] Brabb EE. Innovative approaches to landslide hazard mapping. In: Proceedings 4th international symposium on landslides, Toronto. vol. 1; 1984. p. 307–24.

[18] Goetz JN, Brenning A, Petschko H, Leopold P. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. Comput Geosci. 2015;81:1–11.

[19] Bednarik M, Yilmaz I, Marschalko M. Landslide hazard and risk assessment: a case study from the Hlohovec–Sered' landslide area in south-west Slovakia. Nat Hazards. 2012;64:547–75. doi: 10.1007/s11069-012-0257-7.

[20] Pradhan B, Mansor S, Pirasteh S, Buchroithner M. Landslide hazard and risk analyses at a landslide-prone catchment area using the statistical-based geospatial model. Int J Remote Sens. 2011;32(14):4075–87. doi: 10.1080/01431161.2010.484433.

[21] Regmi AD, Yoshida K, Pourghasemi HR, Dhital MR, Pradhan B. Landslide susceptibility mapping along Bhalubang-Shiwapur area of mid-western Nepal using frequency ratio and conditional probability models. J Mt Sci. 2014;11(5):1266–85.

[22] Wang HB, Wu SR, Shi JS, Li B. Qualitative hazard and risk assessment of landslides: a practical framework for a case study in China. Nat Hazards. 2011;69:1281–94. doi: 10.1007/s11069-011-0008-1.

[23] Hong H, Junzhi L, A-Xing Z. Modeling landslide susceptibility using logitBoost alternating decision trees and forest by penalizing attributes with the bagging ensemble. Sci Total Environ. 2020;718:3–15.

[24] Jia N, Xie M, Mitani Y, Ikemi H, Djamaluddin I. A GIS-based spatial data processing system for slope monitoring. Int Geoinf Res Dev J. 2010;1(4):1.

[25] Karimi Nasab S, Ranjbar H, Akbar S. Susceptibility assessment of the terrain for slope failure using remote sensing and GIS, a case study of Maskoon area, Iran. Int Geoinf Res Dev J. 2010;1(3):1–13.

[26] Varnes DJ. Landslide hazard zonation, a review of principles and practice, International Association of Engineering Geology Commission on Landslides and Other Mass Movements on Slopes, Paris: UNESCO; 1984. p. 63.

[27] Wang Y, Zhice F, Mao W, Ling P, Hong H. Comparative study of landslide susceptibility mapping with different recurrent neural networks. Comp Geosci. 2020;138:10445.

[28] Tie Bui D, Shahabi H, Geertsema M, Omidvar E, Clagu JJ, Thai Pham B, et al. New ensemble models for shallow landslide susceptibility modeling in a semi arid watershed. Forests. 2019a;10(9):743.

[29] Pham BT, Bui DT, Pourghasemi HR, Indra P, Dholakia MB. 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. 2017a;128:255–73.
[30] Tie Bui D, Shahabi H, Omidvar E, Shizardi A, Geertsema M, Clagu JJ, et al. Shallow landslide prediction using a novel hybrid functional machine learning algorithm. Remote Sens. 2019b;11(8):931.
[31] Chen W, Pourghasemi HR, Kornejady A, Zhang N. Landslide spatial modeling: introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. Geoderma 2017a;305:314–27.
[32] Zhu AX, Miao Y, Wang R, Zhu T, Deng Y, Liu J, et al. A comparative study of an expert knowledge-based model and two data-driven models for landslide susceptibility mapping. Catena. 2018a;166:317–27.
[33] Tsegaratos P, Ilia I, Hong H, Chen W, Xu C. Applying information theory and GIS based quantitative methods to produce landslide susceptibility maps in mancheang county, China Landslides. 2017;14:1091–111.
[34] Canoglu MC. Deterministic landslide susceptibility assessment with the use of a new index (factor of safety index) under dynamic soil saturation: an example from Demircikoy watershed (Sinop/Turky). Carpathian J Earth Environ Sci. 2017;12:423–36.
[35] Chen W, Yan X, Pradhan B. Spatial prediction of landslide susceptibility using data mining based kernel logistic regression, naive Bayes, and RBFNetwork for long county area (China). Bull Eng Geol Env. 2019c;78:247–66.
[36] Ciurleo M, Cascini L, Calvello M. A comparison of statistical and deterministic methods for shallow landslide susceptibility zoning in clayey soils. Eng Geol. 2017;2017:223.
[37] Dai FC, Lee CF. Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. Geomorphology. 2002;42:213–28.
[38] Donati L, Turrini MC. An objective method to rank the importance of the factors predisposing to landslides with the GIS methodology application to an area of the Apennines (Valnerina; Perugia, Italy). Engg Geol. 2002;63:277–89.
[39] Luelseged A, Yamagishi H. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains. Cent Jpn Geomorphology. 2005;65:15–31.
[40] Duman TY, Can T, Gokceoglu C, Nefesliogocu HA, Sonmez H. Application of logistic regression for landslide susceptibility zoning of Cekmee area, Istanbul, Turkey. Verlag. 2006;51:242–56.
[41] Sarkar S, Rjan Martha T, Roy A. Landslide susceptibility Assessment using information value method in parts of the Darjeeling Himalayas. Geol Soc India. 2013;82: 351–62.
[42] Chandak PG, Sayyed SS, Kulkarni YU, Devtale MK. Landslide hazard zonation mapping using information value method near Parphi village in Garhwal Himalaya. Ljemas. 2016;4:228–36.
[43] Kouhpeima A, Feizniab S, Ahmadib H, Moghadamniab AR. Landslide susceptibility mapping using logistic regression analysis in Latyan catchment. Desert. 2017;85–95.
[44] Kanungo DP, Sarkar S, Sharma S. Combining neural network with fuzzy, certainty factor south-facing, and likelihood ratio concepts for spatial prediction of landslides, Nat Hazards. 2011;59(3):1491–512.
[45] Liu M, Chen X, Yang S. Collapse landslide and mudslide hazard zonation. In: Landslide Science for a Safer Geoenvironmental. Switzerland: Springer International Publishing; 2014:2. p. 457–62.
[46] Pourghasemi HR, Pradhan B, Gokceoglu C, Mohammadi M, Moradi HR. Application of weights-of-evidence and certainty factor models and their comparison in landslide susceptibility mapping at Haraz watershed, Iran. Arab J Geosci. 2013c;6(7):2351–65.
[47] Pourghasemi HR, Pradhan B, Gokceoglu C, Mohammadi M, Moradi HR. Application of weights-of evidence and certainty factor models and their comparison in landslide susceptibility mapping at Haraz watershed, Iran. Arab J Geosci. 2012a;6:2351–65. doi: 10.1007/s12517-012-0532-7.
[48] Sujatha ER, Rajamanickam GV, Kumaravel P. Landslide susceptibility analysis using probabilistic certainty factor approach: A case study on Tevankarai stream watershed, India. J Earth Syst Sci. 2012;121(5):1337–50.
[49] Akgün A, Dag S, Bulut F. Landslide susceptibility mapping for a landslide-prone area (Findikli, NE of Turkey) by likelihood-frequency ratio and weighted linear combination models. Env Geol. 2008;54:1127–43.
[50] Chung CJF, Fabbri AG. Validation of spatial prediction models for landslide hazard mapping. Nat Hazards. 2003;30(3):451–72.
[51] Lee S, Pradhan B. Probabilistic landslide hazards and risk mapping on Penang Island, Malaysia. J Earth Sys Sci. 2006;115(6):661–7.
[52] Pradhan B, Lee S, Buchroithner MF. Remote sensing and GIS-based landslide susceptibility analysis and its cross-validation in three test areas using a frequency ratio model. Photogramm Fernerkun. 2010c;1:17–32. doi: 10.1127/14328364/2010/0037.
[53] Kanungo DP, Arora MK, Sarkar S, Gupta RP. Landslide susceptibility zonation mapping a review. J South Asia Disaster stud. 2009;2:81–105.
[54] Saha AK, Gupta RP, Sarkar I, Arora KM, Csaplovics E. An approach for GIS-based statistical landslide susceptibility zonation with a case study in the Himalayas. Landslides. 2005;2(1):61–9.
[55] Sarkar S, Kanungo D, Ptra A, Kumar P. Disaster mitigation of debris flow, slope failure, and landslides. GIS-based landslide susceptibility case study in Indian Himalaya. Tokyo, Japan; Universal Academy Press; 2006. p. 617–24.
[56] Mohammad M, Pourghasemi HR, Pradhan B. Landslide susceptibility mapping at Golestan Province, Iran: a comparison between frequency ratio, Dempster–Shafer, and weights-of evidence models. J Asian Earth Sci. 2012;61:22136.
[57] Park N-W. Application of Dempster-Shafer theory of evidence to GIS-based landslide susceptibility analysis. Environ Earth Sci. 2011;62:367–76.
[58] Wang Q, Guo Y, Li W, He J, Wu Z. Predictive modeling of landslide hazards in Wen County, northwestern China based on information value, weights-of-evidence, and certainty factor. Geomatics, Natural Hazards, and Risk. 2019;10(1):820–35. doi: 10.1080/19475705.2018.1549111.
[59] Pham BT, Prakashi I, Singh SK, Shizardi A, Shahabi H, Bui DT. Landslide susceptibility modeling using reduce error pruning trees and different ensemble techniques: hybrid machine learning approach. Catena. 2019b;175:203–18.

[60] Das G, Lepcha K. Application of logistic regression (LR) and frequency ratio (FR) models for landslide susceptibility mapping in Relli Khola river basin of Darjeeling Himalaya. India. SN Appl Sci. 2019;1:1453. doi: 10.1007/s4245 2-019-1499-8.
[61] Yilmaz I, Keskin I. GIS-based statistical and physical approaches to landslide susceptibility mapping (Sebinkarahisar, Turkey). Bull Eng Geol Env. 2009;68:459–71.
[62] Chen W, Xie X, Wang J, Pradhan B, Hong H, Bui DT. A comparative study of the logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. Catena. 2017c;151:147–60.
[63] Youssef AM, Pourghasemi HR, Pourtaghi ZS, Al-Katheeri MM. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. Landslides. 2016;13:839–56.
[64] Tsangaratos P, Ilia I. Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: the influence of models complexity and training dataset size. Catena. 2016;145:164–79.
[65] Hong HY, Pradhan B, Xu C, Tien Bui D. Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree, and support vector machines. Catena. 2015b;133:266–81.
[66] Lin GF, Chang MJ, Huang YC, Ho JY. Assessment of susceptibility to rainfall-induced landslides using improved self-organizing linear output map, support vector machine, and logistic regression. Eng Geol. 2017;224:62–74.
[67] Bui DT, Tuan TA, Klempe H, Pradhan B, Revhaug I. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides. 2016;13:361–78.
[68] Hong H, Chen W, Xu C, Youssef AM, Pradhan B, Bui DT. Rainfall-induced landslide susceptibility assessment at the Chongren area (China) using frequency ratio, certainty factor, and index of entropy. Geocarto Int. 2016;32:139–54. doi: 10.1080/10106049.2015.1130086.
[69] Chen W, Pourghasemi HR, Zhao Z. A GIS-based comparative study of Dempster-Shafer, logistic regression and artificial neural network models for landslide susceptibility mapping. Geocarto Int. 2017b;32:367–85.
[70] Yilmaz I. Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. Env Earth Sci. 2010;61: 821–36.
[71] Soyoung P, 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. Env Earth Sci. 2013;68:1443–64.
[72] Lee S, Pradhan B. Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models, Landslides. 2007;4:33–41.
[73] Umar Z, Pradhan B, Ahmad A, Neamah M, Tehrany MS. Earthquake-induced landslide susceptibility mapping using an integrated ensemble frequency ratio and logistic regression models in West Sumatera Province. Indonesia Catena. 2014;118:124–35. doi: 10.1016/j.catena.2014.02.005.
[74] Yang ZH, Lan HX, Gao X, Li LP, Meng YS, Wu YM. Urgent landslide susceptibility assessment in the 2013 Lushan earthquake impacted area, Sichuan Province, China. Nat Hazards. 2015;75(3):2467–87. doi: 10.1007/s11069-014-1441-8.
[75] Van Westen CJ, Castellanos E, Kuriakose SL. Spatial data for landslide susceptibility, hazard, and vulnerability assessment: an overview. Eng Geol. 2008;102:112–31.
[76] Yalcin A, Reis S, Aydinoglu A, Yomralioglu T. AGIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics, and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey. Catena. 2011;85:274–87.
[77] Corominas J, Van Westen C, Frattini P, Cascini L, Malet J-P, Fatopoulou S, et al. Recommendations or the quantitative analysis of landslide risk. Bull Eng Geol Environ 2014;73:209–63.
[78] Zine El Abidine R, Abdel Mansour N. Landslide susceptibility mapping using information value and frequency ratio for the Arzew sector (Northwestern of Algeria). Bull Miner Res Exploration. 2019;160:197–211. doi: 10.19111/bulletinofmre.502343
[79] Zorgati A, Wissem G, Vali V, Smida H, Essghaiem GM. GIS-based landslide susceptibility mapping using bivariate statistical methods. Open Geo Sci. 2019;11:708–26.
[80] Azemeraw W. Landslide susceptibility mapping using statistical methods in Uatzau catchment area, Northwestern Ethiopia. Geoenvironmental Disasters. Forthcoming; 2020. doi: 10.1186/s40677-020-00162-y.
[81] Bonham-Carter GF. Geographic information systems for geoscientists. Modeling with GIS. Oxford: Pergamon; 1994. p. 398.
[82] Lee S, Talib JA. Probabilistic landslide susceptibility and factor effect analysis. J Env Geol. 2005;47:982–90.
[83] Shortliffe EH, Buchanan BG. A model of inexact reasoning in medicine. Math Biosci. 1975;23(3):351–79.
[84] Heckeman D. Probabilistic interpretation of MYCIN's certainty factors. In: Uncertainty in artificial intelligence. Kanal LN, Lemmer JF, editors. New York: Elsevier; 1986:4. p. 167–196.
[85] Dou J, Oguchi T, Hayakawa YS, Uchiyama S, Saito H, Paudel U. GIS-based landslide susceptibility mapping using a certainty factor model and its validation in the Chuetsu area, central Japan. In: Landslide Science for a Safer Geoenvironment. Springer International Publishing; 2014. p. 419–24.
[86] Pradhan B, Lee S. Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modeling. Environ Model Softw. 2010;25:747–59.
[87] 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.
[88] Chau KT, Chan JE. The regional bias of landslide data in generating susceptibility maps using logistic regression: Case of Hong Kong Island. Landslide. 2005;2:280–90.
[89] Lee S, Sambath T. Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. Env Geol. 2006;50:847–55.

[90] Lee S, Pradhan B. Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. Landslides. 2007;4:33–41.
[91] Chen Z, Wang J. Landslide hazard mapping using a logistic regression model in Mackenzie Valley. Can Nat Hazards. 2007;42:75–89.
[92] Rickli C, Graf F. Effects of forests on shallow landslides – case studies in Switzerland. Forest Snow and Landscape Res. 2009;82:33–44.
[93] Zhang YS, Javed I, Yae Y. Landslide susceptibility mapping using an integrated model of information value and logistic regression methods in the Bailongjiang watershed, Gansu province, China. J Mt Sci. 2017;14:249–68.
[94] Guzzetti F, Carrara A, Cardinal M, Reichenbach P. Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, central Italy. Geomorphology. 1999;31(1–4):181–216.
[95] Ohlmacher GC, Davis JC. Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. Eng Geol. 2003;69:331–43.
[96] Atkinson PM, Massari R. Generalized linear modeling of landslide susceptibility in the Central Apennines. Computers Geosci. 1998;4:373–85.
[97] Dai FC, Lee CF. Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. Geomorphology. 2002;42:213–28.
[98] Schicker R, Moon V. Comparison of bivariate and multivariate statistical approaches in landslide susceptibility mapping at a regional scale. Geomorphology. 2012;161–162: 40–57.
[99] Gorsevski PV, Gessler P, Foltz RB. Spatial prediction of landslide hazard using discriminant analysis and GIS. In: GIS in the Rockies 2000 Conference and Workshop; 2000.
[100] Yesilnacar E, Topal T. Landslide susceptibility mapping: A comparison of logistic regression and neural networks method in a medium scale study, Hendek region (Turkey). Eng Geol. 2005;79:251–66.
[101] Bui DT, Tuan TA, Klempe H, Pradhan B, Revhaug I. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides. 2016;13:361–78.
[102] Mezughi TH, Akhir JM, Rafek AG, Abdullah I. Landslide susceptibility assessment using frequency ratio model applied to an area along the E-W Highway (Gerik-Jeli). Am J Env Sci. 2011;7:43–50.
[103] Silalahi FES, Pamela YA, Hidayat F. Landslide susceptibility assessment using frequency ratio model in Bogor, West Java, Indonesia. Geosci Lett. 2019;6:10.
[104] Fell R, Corominas J, Bonnard C, Cascini L, Leroi E, Savage WZ. Guidelines for landslide susceptibility, hazard, and risk zoning for land-use planning, joint technical committee (JTC-1) on landslides, and engineered slopes. Eng Geol. 2008;102:85–98.
[105] Oh HJ, Lee S, Wisut C, Kim CH, Kwon JH. Predictive landslide susceptibility mapping using spatial information in the Pechabun Area of Thailand. Env Geol. 2009;57: 641–51.
[106] Mandal S, Mondal S. Probabilistic approaches, and landslide susceptibility. Geoinformatics and modeling of landslide susceptibility and risk. Environmental science and engineering. Springer Book Ser (ESE). 2019;145–63.
[107] Pham BT, Prakashi I, Singh SK, Shizardi A, Shahabi H, Bui DT. Landslide susceptibility modeling using reduce error pruning trees and different ensemble techniques: hybrid machine learning approach. Catena. 2019b;175:203–18.
[108] Yilmaz I, Marschalko M, Bendarik M. An assessment on the use of bivariate, multivariate and soft computing techniques for collapse susceptibility in a GIS environment. J earth Syst Sci. 2013;122(2):371–88. doi: 10.1007/s12040-013-0281-3.
[109] Yalcin A, Reis S, Aydinoglu AC, et al. A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey. Catena. 2011;85(3):274–87. doi: 10.1016/ j.catena.2011.01.014.
[110] Chen T, Niu R, Jia X. A comparison of information value and logistic regression models in landslide susceptibility mapping by using GIS. Environ Earth Sci. 2016;75(10):1–16. doi: 10.1007/s12665-016-5317-y.
[111] Solaimani K, Seyedeh ZM, Ataollah K. Landslide susceptibility mapping based on frequency ratio and logistic regression models. Arab J Geosci. 2013;6:2557–69.
[112] Das G, Lepcha K. Application of logistic regression (LR) and frequency ratio (FR) models for landslide susceptibility mapping in the Relli Khola river basin of Darjeeling Himalaya, India. SN Applied Sciences. Switzerland AG: Springer Nature; 2019. p. 1453. doi: 10.1007/s42452-019-1499-8.