

RESEARCH

Open Access



Landslide susceptibility mapping using statistical methods in Uatzau catchment area, northwestern Ethiopia

Azemeraw Wubalem

Abstract

Uatzau basin in northwestern Ethiopia is one of the most landslide-prone regions, which characterized by frequent high landslide occurrences causing damages in farmlands, non-cultivated lands, properties, and loss of life. Preparing a Landslide susceptibility mapping is imperative to manage the landslide hazard and reduce damages of properties and loss of lives. GIS-based frequency ratio, information value, and certainty factor methods were applied. The landslide inventory map was prepared from detailed fieldwork and Google Earth imagery interpretation. Thus, 514 landslides were mapped, and out of which 359 (70%) of landslides were randomly selected keeping their spatial distribution to build landslide susceptibility models, while the remaining 155 (30%) of the landslides were used to model validation. In this study, six factors, including lithology, land use/cover, distance to stream, slope gradient, slope aspect, and slope curvature were evaluated. The effects of the landslide factor of slope instability were determined by comparing with landslide inventory raster using the GIS environment. The landslide susceptibility maps of the Uatzau area were categorized into very low, low, moderate, high and very high susceptibility classes. The landslide susceptibility maps of the three models validated by the ROC curve. The results for the area under the curve (AUC) are 88.83% for the frequency ratio model, 87.03% for certainty factor, and 84.83% of information value models, which are indicating very good accuracy in the identification of landslide susceptibility zones of a region. From these resulted maps, it is possible to recommend, the statistical methods (Frequency Ratio, Information Value, and Certainty Factor Methods) are adequate to landslide susceptibility mapping. The landslide susceptibility maps can be used for regional land use planning and landslide hazard mitigation purposes.

Keywords: Landslide, Susceptibility, Geographic information system (GIS), Certainty factor, Frequency ratio, Information value, Ethiopia

Introduction

As defined by Brunsden (1979) and Cruden (1991), landslides are the downslope movements of debris, rocks, or earth material under the influence of the force of gravity. It has occurred when the driving force exceeds the resistance force due to the destabilization of natural soil or rock slopes. The natural slope will be destabilized by the natural and anthropogenic factors including improper

land use practice, the presence of loss sediment, heavy and prolonged rainfall, highly weathered and fractured rocks, gully and riverbank erosion, earthquake, due to superficial soil-rock interfere and unplanned urban explanation (Woldearegay, 2013; Wubalem and Meten, 2020). The landslides activities in Ethiopia are mostly associated in northern, northwestern, central, southern, southwestern and the rift escarpments due to the presence of complex geomorphological setting, hydrological setting, geological setting, active geodynamic process and unplanned land use practice (Woldearegay, 2013).

Correspondence: alubelw@gmail.com
Engineering Geologist, College of Natural and Computational Sciences,
Department of Geology, University of Gondar, Gondar, Ethiopia

Globally, a landslide is causing thousands of victims and deaths, hundreds of billion dollars of damages, and environmental losses every year (Aleotti and Chowdhury 1999; Gutiérrez et al. 2015; Jazouli et al. 2019). Heavy rainfall and earthquakes mostly trigger landslides in Ethiopia (Woldearegay, 2013). It has resulted in a loss of human and animal lives, damages in infrastructures and properties (Ayalew 1999; Temesgen et al. 2001; Woldearegay 2008; Ibrahim 2011; Meten et al. 2015; Wubalem and Meten 2020). In the last 2 years, from 2018 to 2019, rainfall triggered landslides also caused 60 people to die, 30 people were injured, 5091 households were displaced, houses were damaged, and a widely cultivated and non-cultivated land was destructed in different parts of the country (Wubalem and Meten 2020). Despite the landslide problem is critical in Ethiopia, still there is no adequate slope stability assessment has applied in the different parts of the country (Wubalem and Meten 2020). Uatzau basin is one of the areas that frequently affected by the rainfall triggered landslide incidences and so far, the area not yet 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. This contribution provides the originality of this study. Therefore, landslide susceptibility mapping and assessment in this area can be provided with useful information that helps us to disaster loss reduction and serve as a guideline for sustainable land use planning.

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 generations of landslide susceptibility map (Rai et al. 2014). Landslide susceptibility is the likelihood of a landslide occurrence in an area depend on the terrain condition (Brabb 1984). It is an estimate of where landslides will have occurred. The landslide susceptibility mapping is not only to ascertain the factors that have most influential to the landslides occurred in the region but also to estimate the relative contribution of each factor for slope failures (Chen and Wang 2007). It is also important to inaugurate an association between the factors and landslides to foresee the landslide hazard in the future (Chen and Wang 2007). Before nowadays, because of the lack of remote sensing data and advancements of GIS tools, landslide susceptibility mapping has been difficult tasks. However, at the present day, the advancement of computers, remote sensing and GIS makes easy the preparation of landslide susceptibility map (Jia et al. 2010; Karimi et al. 2010; Wang et al. 2011; Pradhan et al.

2011; Bednarik et al., 2012). Although several approaches are developed for landslide susceptibility mapping, generally they can be categorized into (Akgun et al., 2008) deterministic (or engineering or Geotechnical), (Aleotti & Chowdhury, 1999) heuristic (or index), (Zorgati et al., 2019) the statistical methods (Varnes 1984, Regime et al. 2014) and (Bednarik et al., 2012) machine learning methods or data mining methods. The statistical approaches (multivariate and bivariate statistical techniques) are widely used throughout the world and provides reliable results (Dai and Lee 2002; Donati and Turrini 2002; Ayalew and Yamagishi 2005; Duman et al. 2006; Sakar et al. 2013; Meten et al. 2015; Chandak et al. 2016; Zhang K, et al (2017) The assessment of landslide susceptibility mapping using random forest and decision tree methods in the three Gorges reservoir area, China. Environ Earth Sci. 76:405. et al. 2017; Kouh-peima et al. 2017; Wubalem and Meten 2020). Certainty factor is one of the probability bivariate statistical methods, which can provide with reliable results and help to determine the correlation between landslide factor and landslide occurrence (Kanungo et al. 2011; Pourghasemi et al. 2012; Sujatha et al. 2012; Pourghasemi et al. 2013; Liu et al. 2014). The frequency ratio model is one of the bivariate statistical methods which is easy and provide reliable models (Chung and Fabbri 2003, 2005; Lee and Pradhan 2006, 2007; Akgun et al. 2008; Pradhan et al. 2010, 2011, 2012; Meten et al. 2015). Another commonly practiced method in landslide susceptibility mapping is an information value method, which easily operated and provided reliable results (Saha et al. 2005; Sarkar et al. 2006; Kanungo et al. 2009; Wubalem and Meten 2020).

The Uatzau basin is one of the areas characterized by populating settlements, intensive farming, and frequent landslide incidence, which destroyed widely cultivated land, is important to evaluate the factors that have more role in causing slope failure and to minimize their socioeconomic impacts by generating a landslide susceptibility map. For this purpose, statistical methods, including frequency ratio, information value, and certainty factor methods were applied. These methods are easy to apply and it gives a very well-meaning result. In literature, various bivariate approaches for landslide susceptibility mapping are available, however, a comparison among CF, FR, and IV models yet have not encountered. A comparison among the three models has discussed in this paper. The accuracy of results of landslide susceptibility maps, which generated using the statistical methods, evaluated using the receiver operating characteristic curve (ROC). The resulted maps will be used for landslide mitigation purposes and regional land use planning.

Study area and geological setting

The study area is located in the northwestern highlands of Ethiopia. It lies within the latitude, 117, 215 m N to 1, 138, 231 m N and the longitude 349, 253 m E to 364, 786 m E. The study area covers an area of about 138 km². The minimum and maximum altitudes of the area are 1332 m of the river gorge and 2, 498 m in hills and plateau lands (Fig. 1). Many tributaries are available in the entire study area and joined the Uatzau River, which drains into the Abay River. The various streams in the study area caused the removal of soil through stream bank erosion. From field observation evidence, the study area is highly affected by gully erosion, which also resulted in small-scale landslide incidences. The study area 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. 54% of this region covered by agricultural lands and rocky lands/bar lands, Residential, and Grazing land cover 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 in 2019. The maximum annual rainfall is 1762 mm while the 970 mm is the minimum annual rainfall, with a mean annual rainfall of 1, 346 mm.

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 (Adigrat sandstone) and volcanic rocks (flood basalt) respectively. The flood basalt rock units are grouped into four geological units, including, Ashengie formation (Lower basalt), Aiba formation (Middle basalt), Alaje formation (Upper basalt), and Termaber formation. Nevertheless, the study area is comprised of only the lower basalt, the early Mesozoic sedimentary rock (lower red sandstone), and

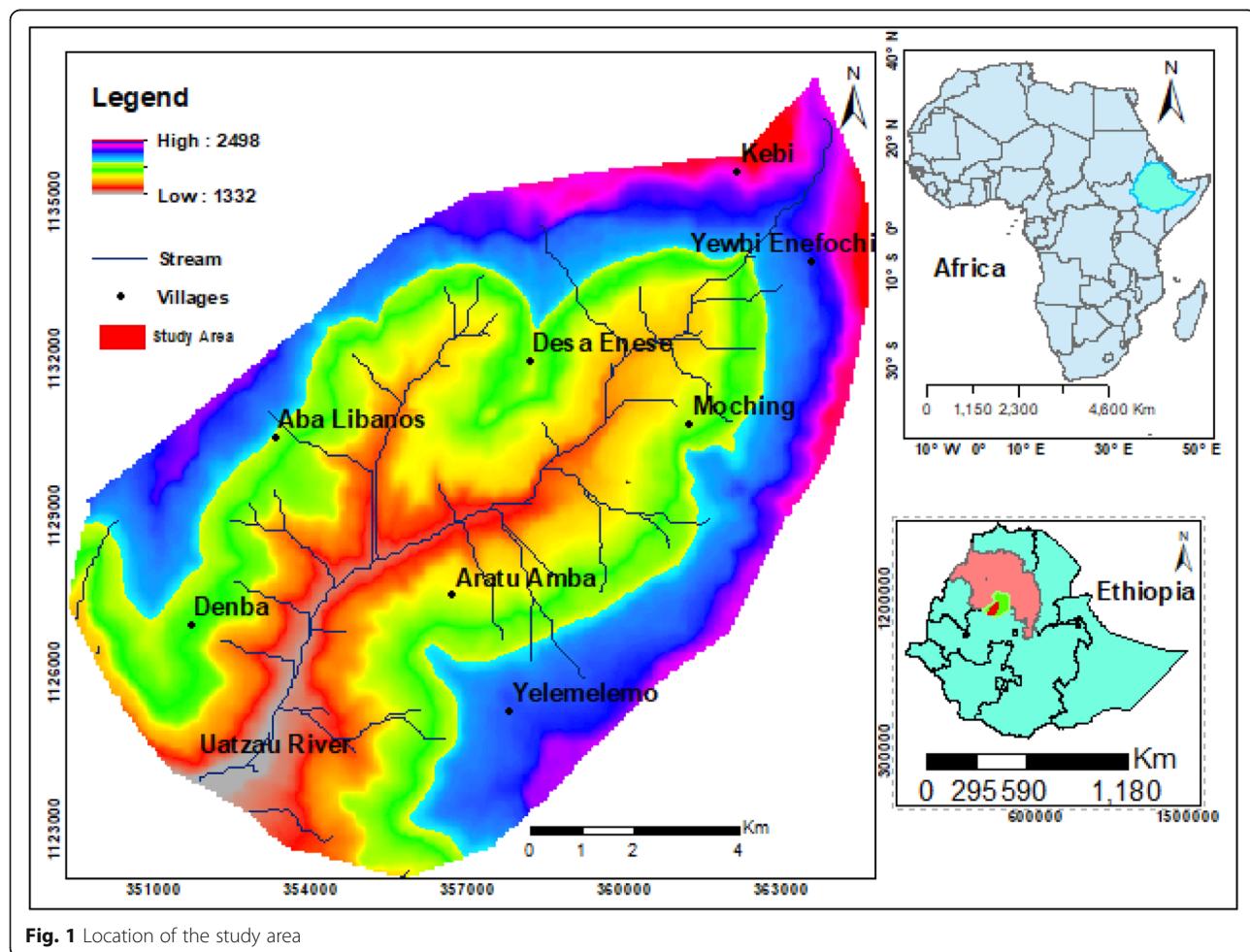


Fig. 1 Location of the study area

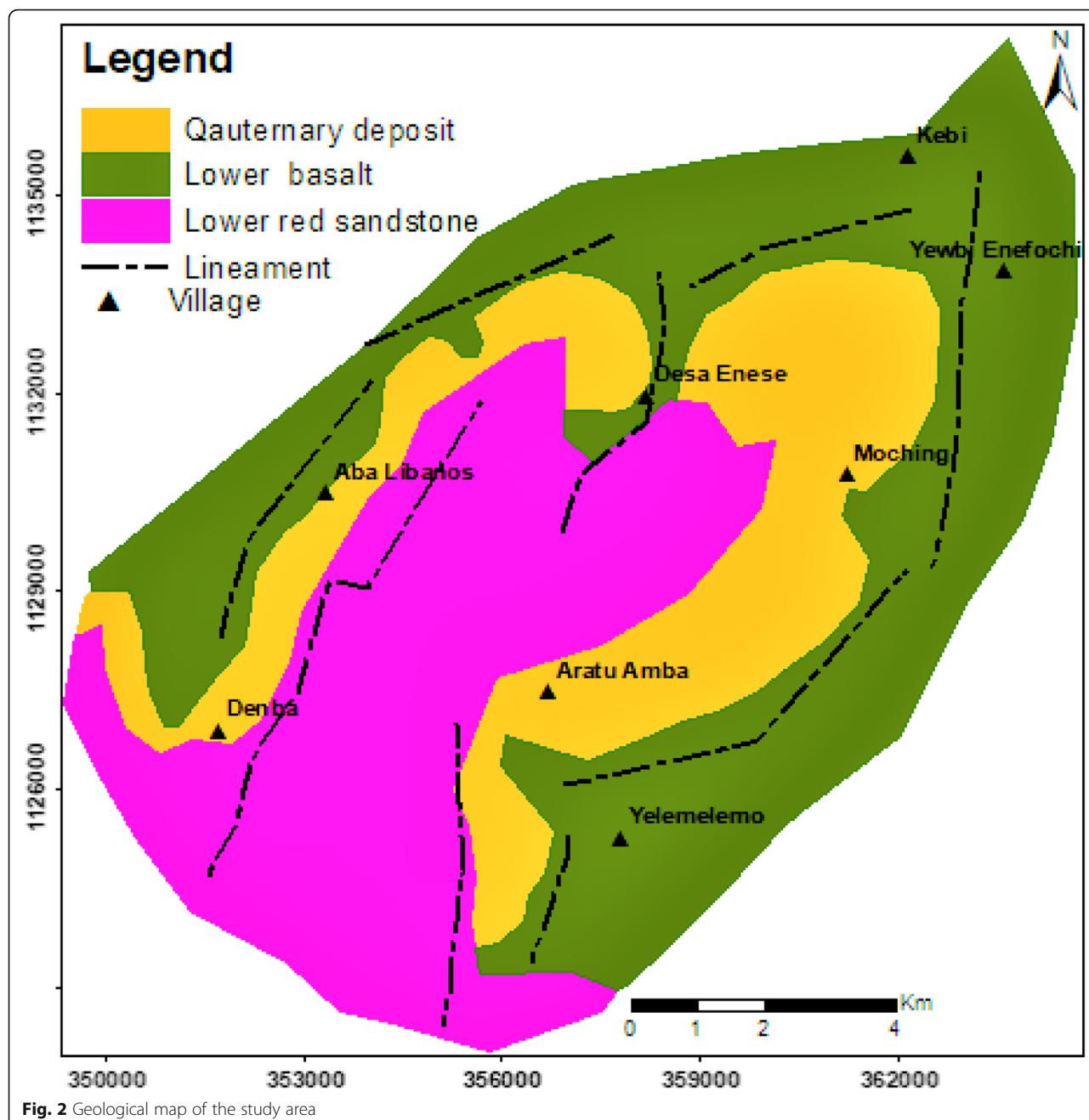


Fig. 2 Geological map of the study area

the Quaternary /recent soil deposit. This lithology digitized from the existing 1:250,000 geological map of the Debre Maroks sheets. As shown from Fig. 2, the southern parts of the study area covered by the sedimentary rock (Lower red sandstone), which is characterized by medium to thickly bedded thickness, fine to coarse texture, red to red-brown color, and strongly cross-bedded. The northern, northwestern, northeastern, and southwestern parts of the study area are dominantly covered by the volcanic rock (lower basalt), which is underlined early Mesozoic red sandstone unit and covered by thin

dark color soil deposit (Fig. 2). This unit characterized by a high degree of weathering and fracturing. The central parts of the study area covered by a very loss /unconsolidated soil deposit, which formed due to slope failure and gravity effects. In this soil deposit, unplanned intensive agricultural activities are common.

Methodology

For this research, data collection, Field investigation, landslide inventory mapping, Google Earth Imagery analysis, landslide factor evaluation, and mapping, GIS-

based frequency ratio, information value, and certainty factor landslide susceptibility modeling and validation were 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), United States Geological Survey (USGS), Amhara Water Well Drilling Enterprise (AWWD E), Field Survey, Google Earth Imagery from the USGS website and Ethiopian National Meteorological Agency (Table 1). The landslides location of the study area identified using field surveys, historical records, and Google Earth imagery analysis. These classified into training and testing landslide data sets. The training landslide data sets used for model preparation, whereas the testing landslide data sets used for model prediction accuracy evaluation. Based on the data availability, literature, field evaluation, and local people interview, six landslides-driving factors were determined. Using ArcGIS 10.1, the landslide driving factor maps and landslide inventory maps were prepared. Distance to stream, slope angle, slope aspect, and curvature extracted from 30 m resolution of Digital Elevation Model (DEM), which downloaded from the USGS website. The lithological layer digitized from the existing geological map of the Debre Markos sheet at a scale of 1:250,000. The land use map was prepared using ArcGIS and Google Earth Imagery analysis. It digitized from Google Earth Imagery interpretation, which can export to a GIS layer format (Kml) and verified in the field 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 also prepared using the supervised classification of

satellite images downloaded from the USGS website. Generally, the general procedure flow chart that followed in this research work summarized in Fig. 3.

Geodatabase building is one of the most fundamental elements in the landslide susceptibility mapping. Therefore, three databases built for frequency ratio, information value, and certainty factor models. These data contained landslide inventory and landslide factors with the same projection (UTM) and pixel size (30mx30m). After the database built, an evaluation of the relationship between landslide and landslide factors as well as the determination of the statistical significance of each landslide factor was the next step in landslide susceptibility mapping. Therefore, six landslide factor maps reclassified into subclass and overlaid with reclassified training landslide data sets raster. Weight ratings for all landslide factor classes assigned statistically using Excel as shown in Eq. (Akgun et al., 2008; Zorgati et al., 2019; Bonham-Carter, 1994). These weighted maps were rasterized-using lookup in spatial analyst. After rasterized the factor maps, the landslide susceptibility index maps generated by the sum-up of all raster maps using a raster calculator in Map Algebra. These maps (LSI) classified into a fivefold classification scheme: very low, low, moderate, high, and very high susceptibility classes using natural breaks (Fig. 4). Finally, the accuracy of the three models evaluated using the prediction rate curve and landslide density based on observed testing landslide data sets.

Landslide inventory mapping

In landslide susceptibility mapping, landslide inventory mapping is one of the key elements, which can be prepared using various techniques like the aerial photograph or Google Earth Imagery interpretation, field investigation, and

Table 1 Unknown switch argument. Information source for the various landslide factors used in the landslides susceptibility mapping

Data	Map	Format	Source
Landslide Inventory	Landslide Inventory Map	Vector (shapefile)	Google Earth Imagery, Field Survey and Historical record
Geology	Lithology Map	Vector (shapefile)	Digitized from Geological map of Debre Markos Sheet provided by the Geological Survey of Ethiopia at 1:250,000
Digital Elevation Model (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	Raster (grid)	Developed from DEM and buffering using distance to Euclidian
Topography	Topographic Map	Vector (shapefile)	Ethiopian Mapping Agency at 1:250,000
Meteorological data	–	Vector (shapefile)	Ethiopian National Meteorological Agency
Land use	Land use Map	Vector (shapefile)	Sentinel 2 images in the USGS, Field Survey and Google Earth Imagery
Borehole data	–	Vector (shapefile)	Amhara Water Well Drilling Enterprise (AWWE)

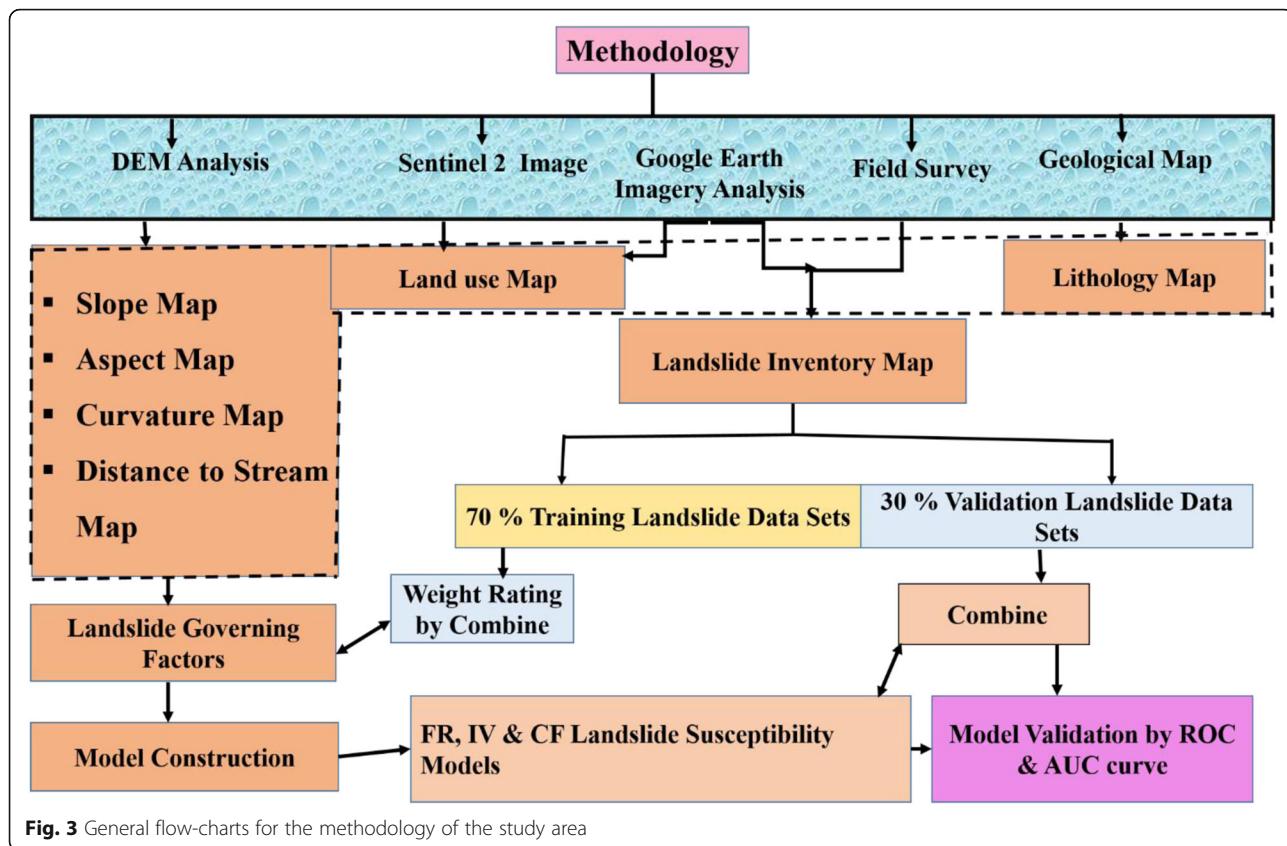


Fig. 3 General flow-charts for the methodology of the study area

evaluation of archived data coupled with GIS tools (Van Westen et al. 2008). Landslide inventory map used as the base for future landslide prediction by evaluating the relationship between the existing landslide event and landslide driving factors (Mohammad et al., 2012; Yalcin et al. 2011; Corominas et al. 2014). Van Westen et al. (2008) prepared a landslide inventory map using GIS from field investigation, historical landslide events, and satellite image analysis. Landslide inventory maps can be also prepared using field surveys and Google Earth Imagery (Meten et al. 2015; Roy and Saha 2019; Zine et al. 2019; Wubalem and Meten 2020). In the present research work, from active and old landslide scarps, 514 landslides, which covered 5.6 km², identified using detailed fieldwork, historical landslide record, and time series Google Earth Imagery analysis (Fig. 5). It was digitized into polygons using a GIS tool with the help of Google Earth Imagery, finally, a landslide inventory map was produced (Fig. 5). From local people witness and time series Google Earth Imagery analysis, the study area was frequently affected by landslide incidence due to 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 literature, some researchers classified landslides into 80% for training landslide and 20% for testing

landslide data sets. But most of the researchers' classified landslides into 70% for training data sets and 30% for testing landslide data sets (Meten et al. 2015; Haoyuan et al. 2016; Anis et al. 2019; Wang et al., 2019; Roy and Saha 2019; Zine et al. 2019). Using ArcGIS 10.1, theses landslides classified randomly into 70% (359) for training landslide data sets and 30% (155) for validation data sets keeping their spatial distribution. The training landslide data sets are observed landslides that used to develop model while the validation landslide data sets are observed landslides that used to evaluate the performance and prediction accuracy of the model. As shown in Figs. 6 & 7, the Desa Enese and Aba Libanos area affected by soil slides in 2019 due to heavy and prolonged rainfall. However, they are not affected by landslide incidence in 2018 and 2016 (Figs. 8 and 9). Rotational landslide in the Desa Enese area was occurring due to the removal of the slope toe by a stream and resulted in damages in farmlands, which covered by crops, and two houses (Fig. 6). As indicated in Fig. 5, the spatial distribution of landslides concentrate dominantly on the ridge, and along the stream bank.

Evaluation of landslide factors

In landslide susceptibility mapping, the selection of landslide factors is one of the most important elements. However, there is no well-defined standard to select the most

significant landslide factors. The factors that initiate the landslide incidence in the study area selected based on data availability, literature review, local person interview, and field evaluation. These are slope angle, slope, aspect, slope curvature, land use, lithology, and distance to stream/river were taken into account to examine the spatial relationship between them and landslide occurrence in the study area. Distance to stream (5 classes), slope angle (5 classes), slope, aspect (10 classes), and slope curvature (3 classes) maps were constructed from 30 m resolution Digital Elevation Model (DEM) which was downloaded from the USGS website. The lithological map of the study area was prepared through digitization from 1:250,000 existing geological maps of Debre Markos sheet from the Geological Survey of Ethiopia, which has three classes (weathered basalt, sandstone, and unconsolidated/colluvial sediments). The land use map of the study area was prepared by digitized from Google Earth Imagery, and the supervised classification of Sentinel 2 images. 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 not familiar to the region. Nevertheless, from the resulting point of view, using a manual land use classification of the Google 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 well known by the user. Land use map has five classes such as grazing land, cultivated land, bare land,

residential, and scatters bush. Even though rainfall is one of the factors that can be triggered landslide incidence, it is not included in this landslide susceptibility modeling because of the lack of rain gage station in the target area. The earthquake did not consider 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 summarized in Table 1. To determine the effects of each landslide factor class on landslide occurrence, weight rating through landslide factor raster combined with landslide raster map is important. For this purpose, all landslide factor maps converted into raster and reclassified with the same pixel size (30 m × 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 combine in spatial analysis tool under local toolbox with the landslide factor raster class to extract landslide pixels for each landslide factor class. Then the effects of each factor class were determined using the equation of frequency ratio (Eq.1), information value (Eq. 3), and certainty factor (Eq. 5) methods, and the results summarized in Table 2.

Modeling approaches

Frequency ratio model

It is one of the bivariate probability methods, which is applicable to determine the correlation between landslide occurrence and landslide causative factor classes. The frequency ratio is the ratio of areas where the landslide occurred in the areas of the landslide factor class. When the ratio value is greater than one, it indicates the strong correlation between factor class and landslide occurrence in a

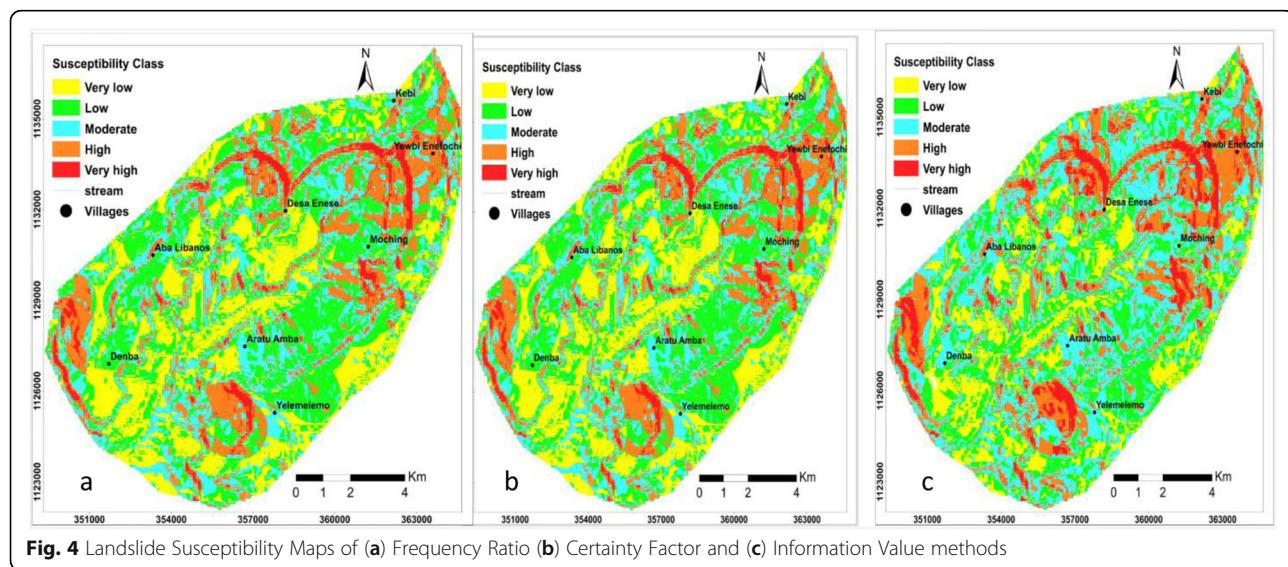


Table 2 Spatial relationship between each landslide factors and landslide using frequency ratio (FR), Certainty factor (CF) and information value (IV) methods

Factors	Class	Ncpix	% Class Area (b)	Nselpix	% landslide in class (a)	FR				IV	CF
						CP	PP	FR	CF		
Lithology	Colluvial Deposit	39,011	25.2	524	32.9	1.3	0.01	0.01	0.27	0.24	
	Sandstone	51,186	33.1	320	20.1	0.6	0.01	0.01	-0.50	-0.40	
	Weathered Basalt	64,419	41.7	748	47.0	1.1	0.01	0.01	0.12	0.11	
Land use/ cover	Agriculture	83,991	54.3	375	8.4	1.1	0.004	0.004	0.1	0.07	
	Settlement	43,115	27.9	133	4.3	0.7	0.003	0.004	-0.3	-0.26	
	Grazing Land	25,490	16.5	76	2.6	0.7	0.003	0.004	-0.3	-0.29	
	Bar Land	1364	0.9	61	0.1	10.7	0.045	0.004	2.4	0.91	
	Scatter Bush	616	0.4	0	0.1	0.0	0.000	0.004	0	-1.00	
Distance to stream	0–50	47,432	30.7	2115	37.5	1.2	0.04	0.04	0.20	0.19	
	50–100	42,442	27.4	1487	26.4	1.0	0.04	0.04	-0.04	-0.03	
	100–150	33,253	21.5	1416	25.1	1.2	0.04	0.04	0.16	0.15	
	150–200	22,225	14.4	469	8.3	0.6	0.02	0.04	-0.55	-0.43	
	200–2,355	9270	6.0	149	2.6	0.4	0.02	0.04	-0.82	-0.57	
Slope angle	0°–7°	48,562	31.4	1577	28.0	0.89	0.03	0.04	-0.12	-0.11	
	7°–14°	48,082	31.1	1329	23.6	0.76	0.03	0.04	-0.28	-0.25	
	14°–21°	29,704	19.2	1124	19.9	1.04	0.04	0.04	0.04	0.04	
	21°–28°	19,003	12.3	899	16.0	1.30	0.05	0.04	0.26	0.24	
	28°–68°	9271	6.0	707	12.5	2.09	0.08	0.04	0.74	0.54	
Slope Aspect	Flat (-1)	37,796	24.4	1143	20.3	0.83	0.03	0.04	-0.19	-0.18	
	North (0–22.5)	2094	1.4	23	0.4	0.30	0.01	0.04	-1.20	-0.71	
	NE (22.5–67.5)	17,652	11.4	388	6.9	0.60	0.02	0.04	-0.51	-0.41	
	E (67.5–112.5)	6855	4.4	146	2.6	0.58	0.02	0.04	-0.54	-0.42	
	SE (112.5–157.5)	6756	4.4	187	3.3	0.76	0.03	0.04	-0.28	-0.25	
	S (157.5–202.5)	26,889	17.4	1299	23.0	1.33	0.05	0.04	0.28	0.25	
	SW (202.5–247.5)	6981	4.5	428	7.6	1.68	0.06	0.04	0.52	0.42	
	W (247.5–292.5)	32,180	20.8	1656	29.4	1.41	0.05	0.04	0.34	0.30	
Slope	NW (292.5–337.5)	9839	6.4	231	4.1	0.64	0.02	0.04	-0.44	-0.36	
	N (337.5–360)	7580	4.9	135	2.4	0.49	0.02	0.04	-0.72	-0.52	
Slope	Concave Slope (-26–-2)	22,	14.8	1188	21.1	1.42	0.05	0.04	0.35	0.31	

Table 2 Spatial relationship between each landslide factors and landslide using frequency ratio (FR), Certainty factor (CF) and information value (IV) methods (*Continued*)

Factors	Class	Ncpix	% Class Area (b)	Nslpix	% landslide in class (a)	FR	CP	PP	IV	CF
Curvature		881								
Flat Slope (-2–2)		109, 717	71.0	3385	60.1	0.85	0.03	0.04	-0.17	-0.16
Convex Slope (Aleotti & Chowdhury, 1999; Zorgati et al., 2019; Bednarik et al., 2012; Bonham-Carter, 1994; Brabb, 1984; Chandak et al., 2016; Brunsden, 1979; Chen & Wang, 2007; Chung & Fabbri, 2005; Chung & Fabbri, 2003; Corominas et al., 2014; Cruden, 1991; Dai & Lee, 2002; Das & Lepcha, 2019; Donati & Turrini, 2002; Dou et al., 2014; Duman et al., 2006; Jazouli et al., 2019; Fell et al., 2008; Gorsevski et al., 2000; Gutiérrez et al., 2015; Hong et al., 2016)	22, 024	14.2	1063	18.9	1.32	0.05	0.04	0.28	0.25	

IV is information value, FR is frequency ratio, CF is certainty factor, CA is a class area, LA is a landslide area in a class, CP is the conditional probability of landslide in a class, and PP is the prior probability of landslide in the entire area

given terrain, however, the ratio value less than one is indicated that weak coloration between landslide occurrence and landslide factors, which means a low probability of landslide occurrence (Bonham-Carter 1994; Lee and Talib 2005). It can be calculated using Eq. 1.

$$FR = \frac{a}{b} = \frac{\text{Nslpix}}{\frac{\text{Ntslpix}}{\text{Ncpix}}} \quad (1)$$

Where FR is frequency ratio, Nslpix is a landslide pixel/area in a landslide factor class, Ntslpix is the total area of a landslide in the entire study area (**a**), Ncpix is an 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 frequency ratio for each causative factor class was calculated using the equation, one, and the results are summarized in Table 1.

After calculation of the frequency ratio for each landslide factor class using Microsoft Excel and GIS, the frequency ratio 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 landslide susceptibility index indicated the degree of susceptibility of the area for landslide occurrence. The landslide susceptibility index (LSI) of the study area calculated by carefully summing up the weighted rasterized factor raster maps using eq. 2 by the raster calculator in Map Algebra of the spatial analysis tool. To get the landslide susceptibility index, the frequency ratio of each factor type or class summed as in Eq. 2.

$$LSI = \sum_{i=1}^n FR_i X_i \quad (2)$$

$$\begin{aligned} LSI = & FR * \text{Slope raster} + FR * \text{Slope aspect raster} \\ & + FR * \text{Slope curvature raster} \\ & + FR * \text{Lithology raster} + FR * \text{Land use raster} \\ & + FR * \text{Distance to stream raster} \end{aligned}$$

Where LSI is the landslide susceptibility index, n is the number of landslide factors, X_i is landslide factor and FR_i is the frequency ratio of each landslide factor type or classes. After landslide susceptibility index calculation, 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 landslide susceptibility index (LSI), the higher the probability of landslide occurrence, but the lower the LSI is indicated, 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 (Fig. 4a).

Information value model

The information value method is one of the probabilistic methods of a bivariate statistical method, which used to envisage the correlation between landslides and landslide factor classes (Sarkar et al. 2006). The information values for each factor class have been determined through the combination of reclassified landslide raster to reclassified landslide factor raster based on the presence of landslide in a

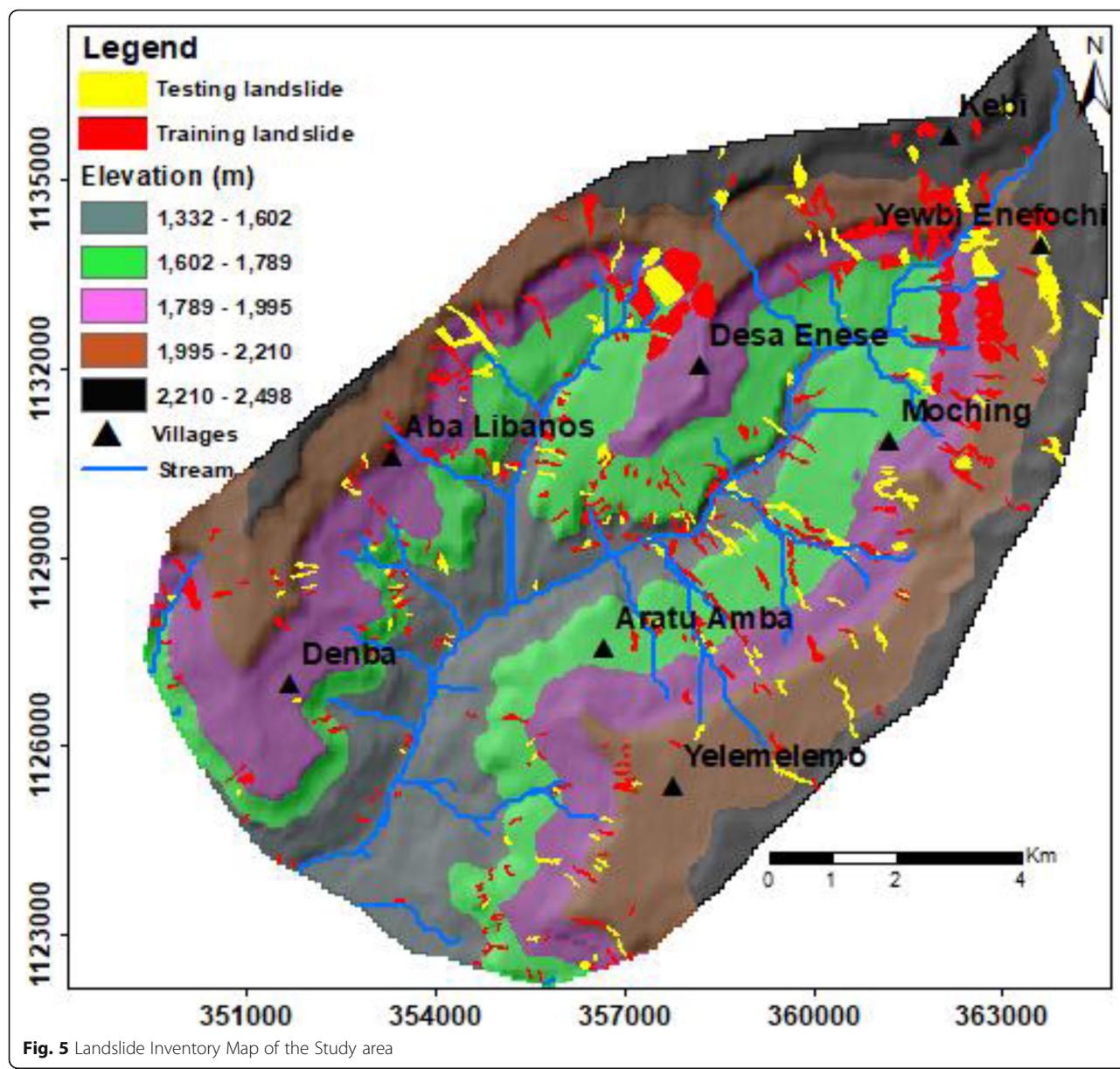


Fig. 5 Landslide Inventory Map of the Study area

given map unit (Fig. 4c). These values are important to define the role of each causal factor in classes for landslide occurrence (Kanungo et al., 2009). This can be calculated as in Eq.3.

$$\begin{aligned}
 IV &= \frac{\text{Conditional probability (CP)}}{\text{Prior probability (PP)}} \\
 &= \frac{\frac{Nsdpix}{Ncpix}}{\frac{Ntsdpix}{Ntcpix}} \quad (3)
 \end{aligned}$$

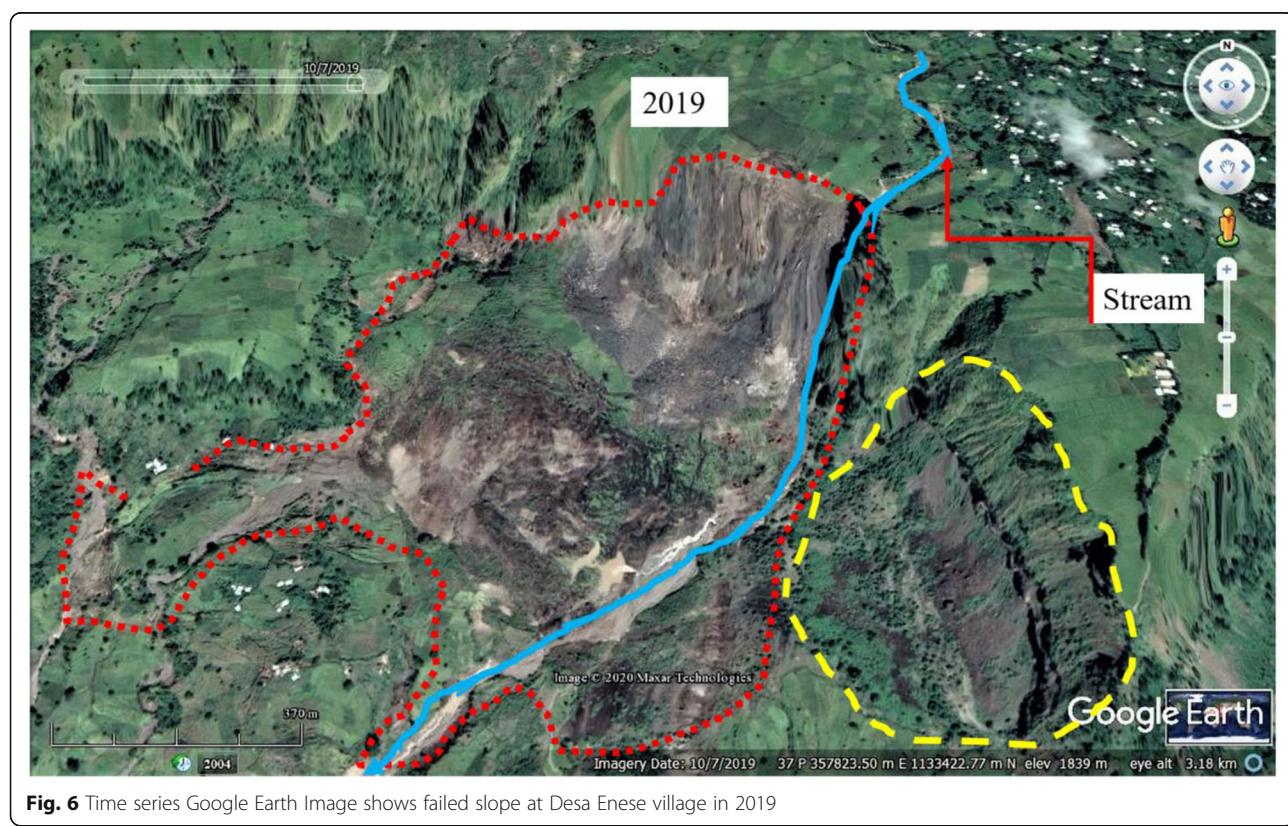
Where Conditional probability is the ratio of the pixel of a landslide in class to the pixel of a class

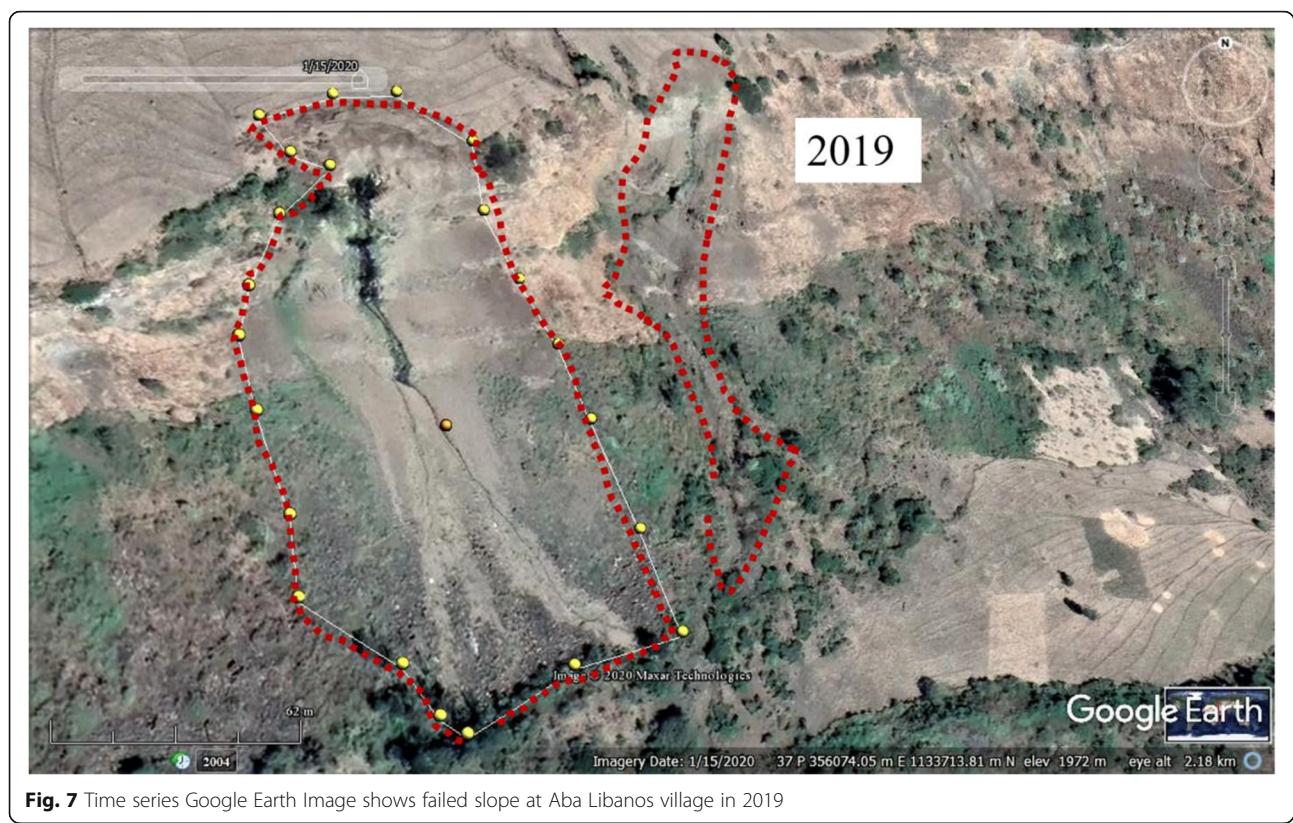
and prior probability is the ratio of the total number of pixels of landslide to the total number of pixels of the study area. $Nsdpix$ is landslide pixel/area in a landslide factor class. $Ntsdpix$ is the total area of a landslide in the entire study area. $Ncpix$ is the area of the class in the study area and $Ntcpix$ is the total pixel area in the entire study area. When the $IV > 0.1$, the landslide occurrence with the factor classes have a high correlation, means it will have a high probability of landslide occurrence however when the $IV < 0.1$ or $IV < 0$, it is low coloration between landslide factors and landslide occurrence which indicated a low probability of landslide occurrence. After calculation of the information value for each

Table 3 Statistical results of landslide susceptibility map produced by frequency ratio, information value, and certainty factor methods

	LSI Value	LSI	Factor class area (%)	Validation data set (%)	Training data set (%)	AUC for validation landslide	AUC for training landslide
Information value method	-0.5-0.9	VLS	15.5	3.9	6.3	0.848323	0.808265
	0.9-1.5	LS	24.3	7.9	11.8		
	1.5-2	MS	31.5	20.1	23.8		
	2.0-2.6	HS	21.1	40.8	31.8		
Certainty Factor (CF)	2.6-4.1	VHS	7.6	27.3	26.3		
	-2.2- -0.97	VLS	17.8	4.7	6.0	0.870348	0.871933
	-0.97- -0.47	LS	31.0	12.3	16.4		
	-0.47-0.04	MS	28.8	17.5	24.8		
	0.04-0.74	HS	19.0	34.8	33.0		
Frequency Ratio (FR)	0.74-2.61	VHS	3.4	30.7	19.7		
	3.1-4.3	VLS	22.7	5.4	9.3	0.888337	0.832718
	4.3-4.8	LS	30.8	14.7	17.8		
	4.8-5.3	MS	22.4	19.5	20.0		
	5.3-6	HS	19.3	43.7	35.0		
	6-7.7	VHS	4.8	16.7	17.8		

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





landslide factor class using Microsoft excel and GIS, the information value for each factor class assigned through the join in the ArcGIS tool. Then, the weighted landslide factors are rasterized using the lookup tool in spatial analysis and the landslide susceptibility index (LSI) of the study area calculated as in Eq. 4.

$$LSI = \sum_{i=1}^n IV_i X_i \quad (4)$$

$LSI = IV * Slope\ raster + IV * Slope\ aspect\ raster + IV * Slope\ curvature\ raster + IV * Lithology\ raster + IV * Land\ use\ raster + IV * Distance\ to\ stream\ raster$. Where LSI is 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.

Certainty factor model

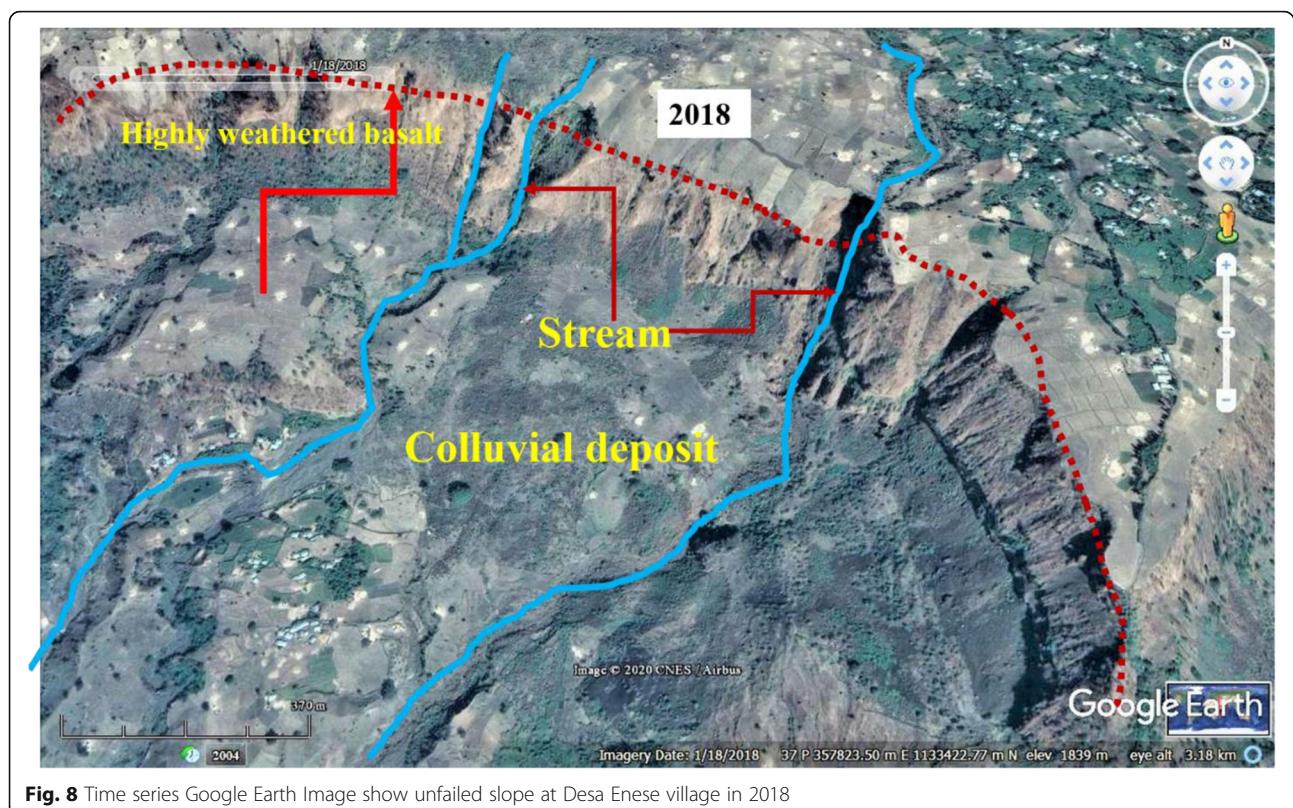
The certainty factor is one of the probabilistic methods that widely used for landslide susceptibility mapping for different data (Kanungo et al. 2011; Sujatha et al. 2012; Pourghasemi et al., 2013; Liu et al. 2014). Shortliffe and Buchanan (1975) proposed the certainty factor (the probability function) for landslide susceptibility mapping

later Heckeman (1986) improved it and it expresses mathematically as:

$$CF = \begin{cases} \frac{PP_a - PP_b}{PP_a(1 - PP_b)} & \text{if } PP_a \geq PP_b \\ \frac{PP_b - PP_a}{PP_a(1 - PP_b)} & \text{if } PP_a \leq PP_b \end{cases} \quad (5)$$

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 of 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 (Sujantha et al. 2012; Dou et al. 2014).

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



$$Z = \begin{cases} X + Y - XYX, & Y \geq 0 \\ \frac{X + y}{1 - \min(|X|, |Y|)} & X * Y < 0 \\ X + Y + XYX, & Y < 0 \end{cases} \quad (6)$$

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

$$LSI = \sum_{i=1}^n CF_i X_i \quad (7)$$

$$\begin{aligned} LSI = & CF * \text{Slope raster} + CF * \text{Slope aspect raster} \\ & + CF * \text{Slope curvature raster} \\ & + CF * \text{Lithology raster} + CF * \text{Land use raster} \\ & + CF * \text{Distance to stream raster}. \end{aligned}$$

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

Model validation

Landslide susceptibility map without validation has no sense in the scientific world (Wubalem and Meten 2020). Therefore, validation of the landslide susceptibility model is very important to evaluate the degree of accuracy of modeling using different validation techniques (Gorsevski et al. 2000; Chung and

Fabbri 2003). For this purpose, the landslide area classified based on time, space, and random partition (Chung and Fabri, 2003, Lee and Pradhan, 2007, and Meten et al., 2015). In this case, the landslide in the study area classified into 70% (359) training landslide data sets and 30% (155) validation landslide data sets randomly keeping their spatial distribution. As stated by Yesilnacar and Topal (2005), the area under the curve (AUC) value used to evaluate the performance of the model, and its value range from 0.5–1. When the AUC value in between the range of 0.9–1, the model has excellent performance; if an AUC value in between the range of 0.8–0.9, the model has very good performance. If the AUC value between the range of 0.7–0.8, the model has good performance. If the AUC value between the range of 0.6–0.7, the model has an average performance. However, if AUC values between the range of 0.5–0.6 and equal to 0.5 or less than 0.5, the model has poor performance (Yesilnacar and Topal 2005).

In the present work, the landslide area randomly classified as 70% landslide for training and 30% landslide for model validation by keeping their spatial distribution into the account using the random partition technique (Chung and Fabri, 2003, Meten et al., 2015). After model development, the models validated by Receiver Operating Characteristics (ROC) curves.

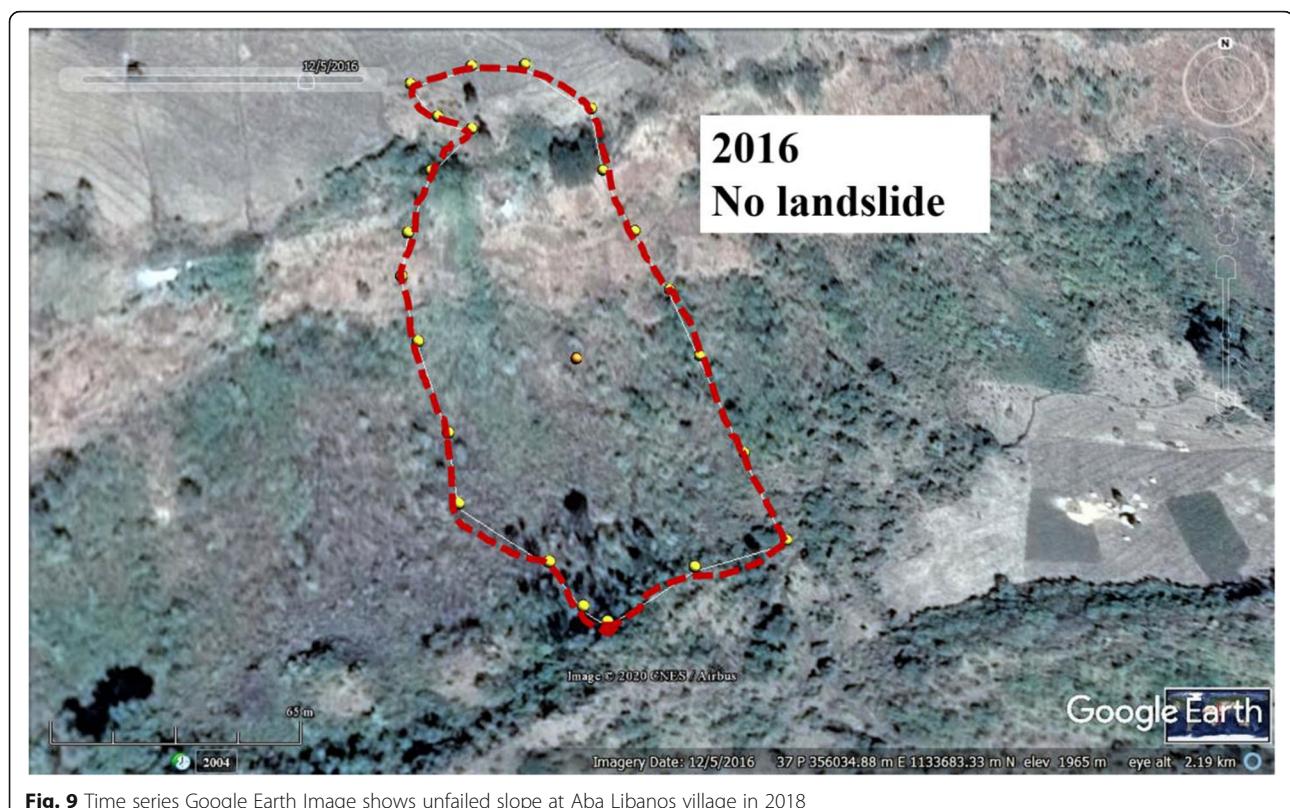


Fig. 9 Time series Google Earth Image shows unfailed slope at Aba Libanos village in 2018

Results and discussions

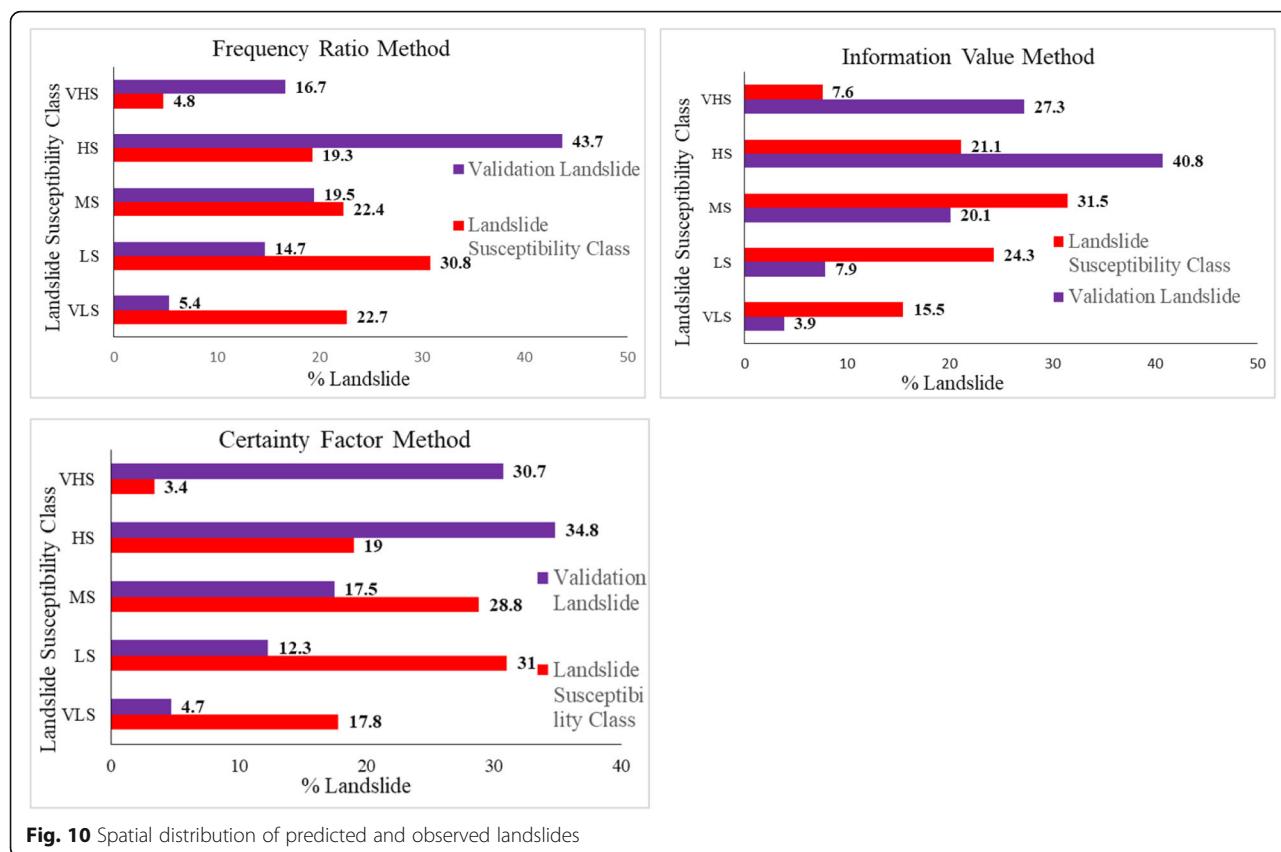
Results

As shown in Table 2, the correlation between landslide locations and landslide driving factor classes were determined using FR (Eq. 1), IV (Eq. 3), and CF (Eq. 5). The higher value of the FR, IV, and CF, indicated the strong correlation between the landslide and landslide factor classes. The detail for FR, IV, and CF has described in the following sections.

Frequency ratio (FR)

To understand the significance of landslide factor classes for landslide occurrence, weight value computed using frequency ratio methods as shown in Eq. 1 (Table 2). The frequency ratio for all landslide factor classes was rating and show important effects of each factor class on slope instability (Table 2). As it can be observed from Table 2, the lithology class colluvial deposit and weathered basalt have a high value of the FR (1.3 and 1.1 respectively) which is > 1 , indicated high landslide probability, but sandstone class has low FR value (0.6) which is < 1 , indicated a low probability of landslide occurrence. Because from field observation, it has been seen that the colluvial deposit is a recent deposit in the study area, which characterized by loose/unconsolidated, low shear strength, and a series of spring water. The presence of spring water has been reducing the normal force in the

slope material when the pore space in the soil grain filled with water, it will be generated pore water pressure. Besides this, a series of the stream has passed through the slope toe of this loose soil deposit, which caused the removal of the slope toe, by the stream bank erosion. This resulted in the reduction of resisting force in the slope material when the slope toe eroded. As we know that landslide may have occurred when the driving force exceeds the resisting force in the slope material. This is happening due to various constraints. In this research case, slope toe erosion by a stream is the key element to driving landslide incidence in the study area. Basalt rock has a high positive relation to landslide occurrence than sandstone due to the effects of weathering because the basalt rock in the study area is highly affected by weathering, but sandstone has a low degree of weathering because of the presence of quartz cement. As designed in Table 2, the slope class 0°–7°, and 7°–14° have low FR value (0.89 & 0.76, respectively) and high value of the FR (1.04, 1.3, & 2.09) for slope classes 14°–21°, 21°–28°, 28°–68°, respectively. This correlation indicated that landslide probability increase as the slope gradient increases (Sun 2009), however, it may not be always true when the steep slope comprised of massive and strong slope material. Landslide may have occurred in a gentle slope when the slope material is loose and the slope subjected to modification due to

**Fig. 10** Spatial distribution of predicted and observed landslides

anthropogenic and natural activity (Wubalem & Meten 2020). However, in the present study area, the result of the FR value indicated in Table 2, as the slope angle increased, the landslide probability is increased. This is because of the presence of shallow loose soil deposit, highly weathered rock, active soil erosion, and improper land use practice. In the case of the slope aspect factor class, the FR value is > 1 for south-facing (1.33), southwest facing (1.68), and west-facing (1.41), indicated 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 class of -26- -2 (1.42) & 2-23 (1.32) is > 1 , indicated high landslide probability. This is because of the effects of slope shape for rainwater impounding and gravity effect. However, the slope curvature class -2-2 has the FR value (0.85) is < 1 , which indicated a low probability of landslide occurrence. In the case of distance to stream, as designated in Table 2, as a distance to stream increase, the probability of landslide occurrence decrease. At a distance of 0–50 m, 50–100 m, and 100–150 m, the value of the FR (1.2) is > 1 , indicated high landslide probability, however, at a distance > 150 m, the value of the FR is < 1 , indicated the low landslide probability. This is because of the effects of slope modification, gully erosion,

riverbank erosion, and river undercutting. As noticed in Table 2, the value of the FR for land use/cover class of agriculture land (1.1) and bar land (10.7) is > 1 , indicated high landslide probability. This is because the cultivated land has increased soil moisture. Whenever the soil moisture increased in the slope, the weight of slope material and the pore water pressure in the slope material increased in parallel. This could have resulted 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 has shown a higher correlation to the probability of landslide occurrence. Hence, bare land in the study area is highly affected by a gully soil erosion, which caused a reduction of shear strength of soil material. The remaining classes, including settlement and grazing land, have FR value < 1 , indicating a low probability of landslide occurrence. Because of settlement and grazing land have been practicing in gentle slope gradient parts of the study area.

Information value (IV)

The information value rating for different landslide factor classes calculated by overlay landslide raster with landslide factor raster layer and it shows the important effects of each factor class on slope instability (Table 2).

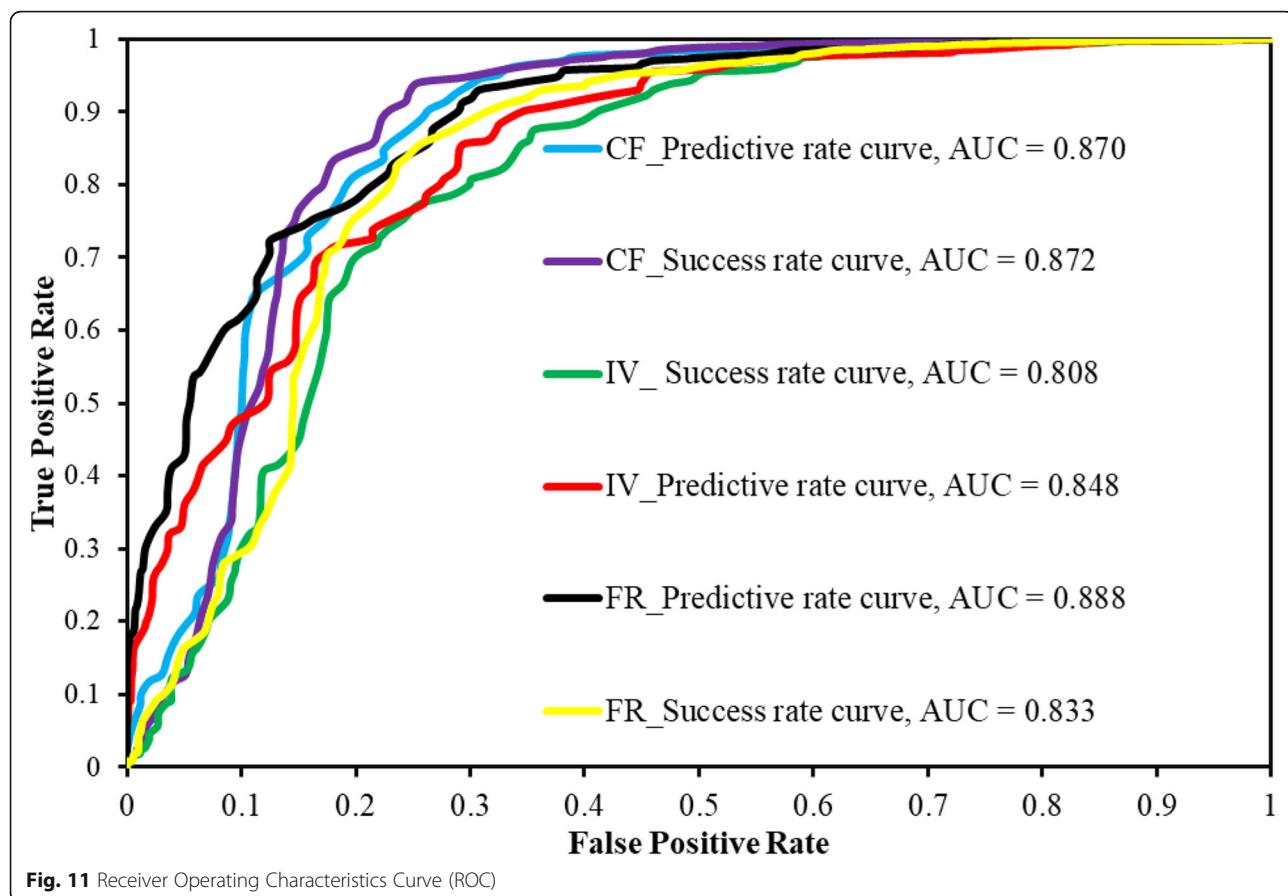


Fig. 11 Receiver Operating Characteristics Curve (ROC)

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 designated in Table 2, the $IV > 0.1$ for lithology class such as colluvial deposit and weathered basalt (0.27 and 0.12 respectively), indicated high landslide probability, but the $IV < 0.1$ for sandstone class (-0.5) which indicated a low probability of landslide occurrence. As observed in Table 2, the $IV < 0.1$, for slope class 0° - 7° , 7° - 14° and 14° - 21° ($IV = -0.12$, -0.28 & 0.04 , respectively), indicated low landslide probability and $IV > 0.1$ for slope classes, 21° - 28° and 28° - 68° , respectively ($IV = 0.27$ & 0.74), indicated high landslide probability. In the case of slope aspect factor class, the $IV > 0.1$ for south-facing ($IV = 0.28$), southwest facing ($IV = 0.52$) and west-facing ($IV = 0.34$), 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 -26°-2° ($IV = 0.35$) & 2°-23° ($IV = 0.28$), indicated high landslide probability. However, the $IV < 0.1$ for the slope curvature class -2°-2° ($IV = -0.17$), indicated low probability of landslide occurrence. At a distance of 0–50 m and 100–150 m, the value of the $IV > 0.1$, which is 0.2 and 0.16, indicated high landslide probability, however,

at a distance 50–100 m and > 150 m, the $IV < 0.1$, indicated the low landslide probability. As noticed in Table 2, the value of IV for land use/cover class of agriculture land (0.07) and bar land (0.91) is > 0.1 , indicated high landslide probability. The IV for the remaining factor classes like settlement, scatter bush and grazing land is < 0.1 , indicated a low probability of landslide occurrence.

Certainty factor (CF)

The certainty factor rating for different landslide factor classes calculated by overlay landslide raster with landslide factor raster layer using Eq. 5 & 6 and it shows the important effects of each factor class on slope instability. As designated in Table 2, the lithology class such as colluvial deposit and weathered basalt have a positive and high value of CF (0.24 and 0.11, respectively), indicated a high landslide probability, but sandstone class has negative CF value (-0.4) which indicated a low probability of landslide occurrence. As observed in Table 2, the slope class 0 - 7° , and 7° - 14° have a negative CF value (-0.11 & -0.25, respectively), indicated the low landslide probability and positive value of CF (0.04, 0.24 & 0.54) for slope classes, 14° - 21° , 21° - 28° , and 28° - 68° , respectively, indicated high landslide probability. In the case of

slope aspect factor class, the CF value is positive for south-facing (0.25), southwest facing (0.42), and west-facing (0.3), indicated high landslide 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 -26°-2 (0.31) & 2-23 (0.25) is positive, indicated high landslide probability. However, the slope curvature class -2-2 has a negative CF value (-0.16), indicated a low probability of landslide occurrence. At a distance of 0–50 m and 100–150 m, the value of CF (0.19 and 0.25) is positive, indicated high landslide probability, however, at a distance 50–100 m and >150 m, have negative value, indicated the low landslide probability. As noticed in Table 2, the value of CF for land use/cover class of agriculture land (0.07) and bar land (0.91) is positive, indicated high landslide probability. The remaining factor classes as settlement, scatter bush and grazing land have negative CF value indicated a low probability of landslide occurrence.

Landslide susceptibility mapping

After the calculation of the landslide susceptibility index, it is important to classify the LSI into different susceptibility classes based on the LSI value. The landslide susceptibility index map of the study area of the information value method, certainty factor method, and frequency ratio method was classified into five levels of susceptibility classes using the natural break method in ArcGIS 10.1. Using the natural breaks method in ArcGIS 10.1, the landslide susceptibility map generated with the information value model reclassified into five classes like very low, low, moderate, high, and very high landslide susceptibility classes (Fig. 4c). From the results of the analysis (Table 3), 15.5% and 24.3% of the study area fall in very low and low susceptibility classes. Moderate, high and very high landslide susceptibility classes have comprised 31.5%, 21.1%, and 7.6% of the study area, respectively. As designated in Table 3, 6.3% and 11.1% of the landslides fall in very low and low susceptibility classes of the study area, respectively. The remaining 23.8%, 31.8%, and 26.3% of landslides fall into moderate, high, and very high landslide susceptibility classes. A landslide susceptibility map produced using certainty factor model (Table 3), very low and low susceptibility classes cover 17.8% and 31.0% of the total study area, however, 28.8%, 19.0% and 3.4% of the total area fall into moderate, high and very high landslide susceptibility classes, respectively. As indicated in Table 3, 4.7% and 12.3% of the landslides fall in very low and low susceptibility classes of the study area, respectively. The remaining 17.5%, 34.8%, and 30.7% of landslides fall into moderate, high, and very high landslide susceptibility classes, respectively. As it

observed from Table 3, the landslide susceptibility map produced using the frequency ratio model, very low and low landslide susceptibility classes cover 22.7% and 30.8% of the total area, however, 22.4%, 19.3% and 4.8% of the total area fall into moderate, high and very high landslide susceptibility classes, respectively. As designated in Table 3, 5.4% and 14.7% of the landslides fall in very low and low susceptibility classes of the study area, respectively. The remaining 19.5%, 43.7%, and 16.7% of landslides fall into moderate, high, and very high landslide susceptibility classes, respectively. For the three models, greater than 60% of validation landslides fall in high and very high susceptibility classes, which is again, confirms that models have very good accuracy (Fig. 10).

Model validation

Model validation is the last step in landslide susceptibility mapping, which helps to evaluate the accuracy of the model, generated using different statistical methods. Various model validation techniques are available like success and predictive rate curve, landslide relative density index (R – index), receiver operating characteristic curve (ROC), and area under the curve (AUC). However, in the present research work, the receiver operating characteristics curve and the area under the curve used to evaluate the accuracy of the landslide susceptibility model generated by frequency ratio, information value, and certainty factor methods. The three models 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 generated 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 (Meten et al. 2015; Silalahi et al. 2019). The prediction rate curve shows how well the models can predict the unknown forthcoming landslide events (Mezughi et al. 2011; Silalahi et al. 2019). In this study, the success rate and prediction rate curve calculated by reclassifying the landslide susceptibility index values into 100 for all cells and sorting in descending order and compare with both training and validation landslide data sets. Finally, the AUC and ROC curve for the three models were calculated using Real Statistics software in add to excel. As the results of the analysis shown in Fig. 11 and Table 3, the closer the ROC curve to the left of the top of the curve, indicating the higher the accuracy of the model. As indicated in Table 3, the AUC value is closer to one, indicating the higher accuracy of the model. The AUC value for CF is 0.870 and 0.872 of the predicted rates and Success rate curve, respectively. This means more or less the AUC value for two data set indicated closer to each other.

In the case of FR, the AUC value is 0.888 and 0.833 for the predictive rate curve and the success rate curve, respectively. The AUC value for IV is 0.848 and 0.808 for the predictive rate curve and the success rate curve. These results indicated that the FR, CF, and IV models have successfully estimated the landslide susceptibility classes of the region, and these models, which have employed in this study, have reasonable accuracy in predicting the landslide susceptibility classes of the study area. However, based on AUC values CF and FR models revealed that a little better result than the IV model for landslide susceptibility mapping in the study area (Fig. 11).

Discussion

Landslide susceptibility maps can forecasting/providing important information where the landslides occur 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 have occurred. However, the maps could not be forecasting the volume of material to displace, the time, and how often the landslide will occur. Nevertheless, the predictive models can be important for the regional land use planning of landslide hazard mitigation and prevention relief (Fell et al. 2008; Oh et al. 2009; Yilmaz and Kkskin 2009; Mezughi et al. 2011; Das and Lepcha 2019; Mandal and Mondal 2019; Silalahi et al. 2019). The landslide susceptibility maps of the study area classified into fivefold classification schemes of very low, low, moderate, high and very high susceptibility classes using natural break method, which is applicable to classify unevenly distributed data, and it is capable of classifying landslide susceptibility index map into different categories considering the inherent data value similarity. The resulted maps were validated using training and testing/validation landslide data sets through the success rate curve and predictive rate curve. The success rate curves for the three models generated from the training landslide data sets through combining tools with Landslide susceptibility classes, which used to evaluate how well the models classified the region based on the existing landslide events (Meten et al. 2015; Silalahi et al. 2019). While the prediction rate curve for the three models was generated from the validation landslide data sets through combining tools with landslide susceptibility classes which are used to evaluate how well the models can predict the unknown forthcoming landslide events (Mezughi et al. 2011; Silalahi et al. 2019). High and very high susceptibility classes in the region are falling in a steep slope, which covered with very loose 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 is fall in the area of highland landscapes. Low

and very low susceptibility of a region falls in the area of low plain landscapes and areas, which have covered by massive weathering resistant rock masses.

Although the three models commonly applied in landslide susceptibility mapping, comparison among them did not work yet. There is some literature regarding the comparison of the frequency ratio method with the information value method, the certainty factor method with the information value method, and the certainty factor method with the frequency ratio method. Zine et al. (2019) state that the information value and frequency ratio methods shown a closer high prediction accuracy ($AUC = 89.05\%$) and $AUC = 85.57\%$, respectively). Similarly, in this study, the frequency ratio method has shown better performance for both success rates ($AUC = 83.27\%$) and predictive rate curve ($AUC = 88.8\%$) more or less similar to the information value methods with success rate curve ($AUC = 80.8\%$) and predictive rate curve ($AUC = 84.8\%$). Even though the frequency ratio model showed a little bit different in AUC value in general, the accuracy of the two models falls in the same ranges, which is a very good performance. As shown from the work of Wang et al. (2019), the certainty factor model showed a high predictive accuracy of AUC value of 75% compared to the information value model with prediction rate curve value ($AUC = 64.08\%$), but their accuracy value is fall in the same ranges which is a good performance. Similarly, in the present model, the certainty factor model also showed a relatively higher prediction rate value ($AUC = 87.03\%$) than the information value model with relatively low prediction rate value ($AUC = 84.8\%$), but they have same accuracy range which is a very good performance. From the work of Haoyuan et al. (2016), based on the predictive rate value of the area under the receiver operating characteristic curve (AUC), the frequency ratio and certainty factor models have shown more or less similar predictive capacity, which is 81.18% for the certainty factor model and 80.14% for the frequency ratio model. However, CF has shown a bit of little performance than the Frequency ratio model. In the present work, the two models showed almost similar AUC value of the prediction rate curve (87.03% for the certainty factor model and 88.8% for the frequency ratio model). Generally, the three bivariate statistical methods in literature and this study showed, the closer prediction capacity with $AUC > 64\%$ and $AUC > 80\%$, respectively falls in the range of good and very good performance (Yesilnacar and Topal 2005). In this study, high and very high landslide susceptibility class covered more than 20% of the study area (Fig. 4) and the percentages of high and very high susceptibility class of a region are more or less similar which are 4.8%, 3.4%, 7.6%, 19.3%, 19% and 21.1% for FR, CF and IV methods, respectively. The landslide validation results for three

models are closer to each other and it falls in the same range of very good performance. Besides this, the percent of landslides that fall in the high and very susceptibility classes are also more or less the same (60.4%, 65.5% & 68.1% for FR, CF, and IV, respectively). Therefore, from these results, the research work finds out that in landslide susceptibility mapping, the three 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 moderate, high, and very high susceptibility area coverage of the IV models showed few differences compared to the FR and CF 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. Based on the prediction accuracy of AUC value, FR and CF models are relatively better for regional land use planning, landslide hazard mitigation, and prevention purposes.

Conclusion

The study area (Uatzau) is characterized by recent unconsolidated soil deposits, rugged topography, active gully, and riverbank erosion, and improper land use practice which makes it very prone to different landslides, including soil slide, weathered rockslide, debris flow, earth flow, earth fall, and soil creep. A landslide can well be thought out the most serious natural hazards in the Uatzau basin. To determine the landslide susceptibility prone areas, Frequency ratio (FR), Certainty factor (CF), and information value (IV) models were applied. The landslide susceptibility maps of the Uatzau basin were categorized into very low, low, moderate, high and very high susceptibility classes. The high and very high susceptibility classes are high in the seven villages including, Desa Enese, Moching, Yewebi Enefoch, Aratu Amba, Aba Libanos, Denba, and Kebi in order of decreasing the risk of landslide incidence due to the presence of active riverbank erosion; loose 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 evaluated using the receiver operating characteristics (ROC) curve through comparison of training and validation landslide raster with the models. The prediction rate curve value of AUC for three models is closing in 1, indicating very good accuracy of the models. Based on the AUC value of the results and > 60% of observed validation landslides which fall in high and very high susceptibility classes, the statistical methods can be proved the most economical and effective methods in landslide susceptibility mapping in the similar regions as the Uatzau area. The

models, which generated using the three 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 Uatzau basin. 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, I would like to thank the almighty God who allowed me to accomplish this research work. Next, I would like to give my gratitude to my lovely wife Ms. Seble Asmare, and my brother Mr. Zemene Wubalem for their valuable support and advice during this research work. I would like to give my special thanks to Dr. Veera and my friends for their continuous support during the research work. Finally, I would like to thank the University of Gondar for geological equipment; National Meteorological Agency staffs, Geological Survey of Ethiopia, disaster head office, natural resource management head office, and the rural community, for their valuable data.

Author's contributions

All activities starting from the conception and design of the work, the developments of the models as well as the statistical analysis and interpretations of the results has done by me.

Funding

It is not applicable in this case.

Availability of data and materials

All the datasets that have been used and analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

I have declared that there are no competing interests.

Received: 14 February 2020 Accepted: 8 August 2020

Published online: 05 January 2021

References

- Akgun A, Dag S, Bulut F (2008) Landslide susceptibility mapping for a landslide-prone area (Findikli, NE of Turkey) by likelihood-frequency ratio and weighted linear combination models. *Environ Geol* 54:1127–1143
- Aleotti P, Chowdhury R (1999) Landslide hazard assessment: summary review and new perspectives. *Bull Eng Geol Environ* 58:21–44
- Anis Z, Wissem G, Vali V, Essghaier HSGM (2019) GIS-based landslide susceptibility mapping using bivariate statistical methods in North-western Tunisia. *Open Geosciences* 11:708–726. <https://doi.org/10.1515/geo-2019-0056>
- Ayalew L (1999) The effects of the seasonal rainfall on a landslide in the highland of Ethiopia. *Bull Eng Geol Environ* 58:9–19
- Ayalew L, Yamagishi H (2005) The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology* 65:15–31
- Bednarik M, Yilmaz I, Marschalko M (2012) Landslide hazard and risk assessment: a case study from the Hlohovec-Sered' landslide area in south-West Slovakia. *Nat Hazards*, Switzerland. <https://doi.org/10.1007/s11069-012-0257-7>
- Bonham-Carter GF (1994) Geographic information systems for geoscientists. In: Modeling with GIS. Pergamon, Oxford, p 398
- Brabb EE (1984) Innovative approaches to landslide hazard mapping. In: Proceedings 4th international symposium on landslides, vol 1, Toronto, pp 307–324

- Brunsdon D (1979) Mass movement. In: Embleton C, Thornes J (eds) Processes in Geomorphology. Edward Arnold Ltd, London, p 436
- Chandak PG, Sayyed SS, Kulkarni YU, Devtale MK (2016) Landslide hazard zonation mapping using information value method near Parphi village in Garhwal Himalaya. *Ljemas* 4:228–236
- Chen Z, Wang J (2007) Landslide hazard mapping using a logistic regression model in Mackenzie Valley, Canada. *Nat Hazard* 42(1):75–89
- Chung CJ, Fabbri AG (2005) Systematic procedures of landslide hazard mapping for risk assessment using spatial prediction models. In: Glade T, Anderson MG, Crozier MJ (eds) Landslide hazard and risk. Wiley, New York, pp 139–177
- Chung CJF, Fabbri AG (2003) Validation of spatial prediction models for landslide Hazard mapping. *Nat Hazards* 30(3):451–472
- Corominas J, Van Westen C, Frattini P et al (2014) Recommendations or the quantitative analysis of landslide risk. *Bull Eng Geol Environ* 73:209–263
- Cruden DM (1991) A simple definition of a landslide. *Bull Eng Geol Environ* 43(1): 27–29
- Dai FC, Lee CF (2002) Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology* 42:213–228
- Das G, Lepcha K (2019) 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* 1:1453. 10.1007/s4245 2-019-1499-8
- Donati L, Turrini MC (2002) 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). *Eng Geol* 63:277–289
- Dou J, Oguchi T, Hayakawa YS, Uchiyama S, Saito H, Paudel U (2014) 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, Amsterdam, pp 419–424
- Duman TY, Can T, Gokceoglu C, Nefesioglu HA, Sonmez H (2006) Application of logistic regression for landslide susceptibility zoning of Cekmee area, Istanbul, Turkey. Verlag, Switzerland, pp 242–256
- Fell R, Corominas J, Bonnard C, Cascini L, Leroi E, Savage WZ (2008) Guidelines for landslide susceptibility, hazard, and risk zoning for land-use planning, joint technical committee (JTC-1) on landslides and engineered slopes. *Eng Geol* 102:85–98
- Gorsevski PV, Gessler P, Foltz RB (2000) Spatial prediction of landslide hazard using discriminant analysis and GIS. In: *GIS in the Rockies 2000 Conference and Workshop*
- Gutiérrez F, Linares R, Roqué C, Zarroca M, Carbonell D, Rosell J, Gutiérrez M (2015) Large landslides associated with a diapiric fold in Canelles reservoir (Spanish Pyrenees): detailed geological-geomorphological mapping, trenching and electrical resistivity imaging. *Geomorphology* 241:224–242
- Haoyuan H, Knarr BA, Pohlig RT, Higginson JS, Binder-Macleod SA (2016) Mechanisms used to increase peak propulsive force following 12-weeks of gait training in individuals poststroke. *J Biomech* 49(3):388–395. <https://doi.org/10.1016/j.jbiomech.2015.12.040>
- Heckerman D (1986) Probabilistic interpretation of MYCIN's certainty factors. In: Kanal LN, Lemmer JF (eds) *Uncertainty in artificial intelligence*. Elsevier, New York, pp 298–311
- Hong H, Chen W, Xu C, Youssef AM, Pradhan B, Bui DT (2016) Rainfall-induced landslide susceptibility assessment at the Chongren area (China) using frequency ratio, certainty factor, and index of entropy. *Geocarto Int.* <https://doi.org/10.1080/10106049.2015.1130086>
- Ibrahim J (2011) Landslide assessment and hazard zonation in Mersa and Wurgesha, north Wollo, Ethiopia. Unpublished Master Thesis, School of Graduate Studies, Addis Ababa University, Addis Ababa, pp 1–10
- Jazouli E, Barakat A, Khelilou R (2019) GIS-multipriteria evaluation using AHP for landslide susceptibility mapping in Oum Er Rbia high basin (Morocco)
- Jia N, Xie M, Mitani Y, Ikemi H, Djamatuluddin I (2010) A GIS-based spatial data processing system for slope monitoring. *Int Geo Res Dev* J 1:4
- Kanungo DP, Arora MK, Sarkar S, Gupta RP (2009) Landslide susceptibility zonation mapping a review. *J South Asia Disaster stud* 2:81–105
- Kanungo DP, Sarkar S, Sharma S (2011) Combining neural network with fuzzy, certainty factor south-facing, and likelihood ratio concepts for spatial prediction of landslides. *Nat Hazards* 59(3):1491–1512
- Karimi Nasab S, Ranjbar H, Akbar S (2010) 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 1:3
- Kouhpeima S, Feizniab H, Moghadamnia AR (2017) Landslide susceptibility mapping using logistic regression analysis in Latyan catchment. *Desert* 22:85–95
- Lee S, Pradhan B (2006) Probabilistic landslide hazards and risk mapping on Penang Island, Malaysia. *J Earth Sys Sci* 115(6):661–667
- Lee S, Pradhan B (2007) Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides* 4:33–41
- Lee S, Talib JA (2005) Probabilistic landslide susceptibility and factor effect analysis. *J Environ Geol* 47:982–990
- Liu M, Chen X, Yang S (2014) Collapse landslide and mudslide hazard zonation; In *Landslide science for a safer geoenvironmental*. Springer International Publishing, Switzerland, pp 457–462
- Mandal S, Mondal S (2019) Probabilistic approaches and landslide susceptibility. Geoinformatics and modeling of landslide susceptibility and risk. In: *Environmental science and engineering*. Springer book series (ESE), New York pp 145–163
- Meten M, Bhandary NP, Yatabe R (2015) GIS-based frequency ratio and logistic regression modeling for landslide susceptibility mapping of Debre Sina area in central Ethiopia. *J Mountain Sci* 12(6):1355–1372
- Mezughi TH, Akhir JM, Rafek AG, Abdullah I (2011) Landslide susceptibility assessment using frequency ratio model applied to an area along the E-W highway (Gerik-Jeli). *Am J Environ Sci* 7:43–50
- Mohammad M, Pourghasemi HR, Pradhan B (2012) Landslide susceptibility mapping at Golestan Province, Iran: a comparison between frequency ratio, Dempster-Shafer, and weights-of-evidence models. *J Asian Earth Sci* 61: 22136
- Oh HJ, Lee S, Wisut C, Kim CH, Kwon JH (2009) Predictive landslide susceptibility mapping using spatial information in the Pechabun area of Thailand. *Environ Geol* 57:641–651
- Pourghasemi HR, Pradhan B, Gokceoglu C, Mohammadi M, Moradi HR (2012) Application of weights-of-evidence and certainty factor models and their comparison in landslide susceptibility mapping at Haraz watershed, Iran. *Arab J Geosci*. <https://doi.org/10.1007/s12517-012-0532-7>
- Pourghasemi HR, Pradhan B, Gokceoglu C, Mohammadi M, Moradi HR (2013) Application of weights-of-evidence and certainty factor models and their comparison in landslide susceptibility mapping at Haraz watershed, Iran. *Arab J Geosci* 6(7):2351–2365
- Pradhan B, Chaudhari A, Adinarayana J, Buchroithner MF (2012) Soil erosion assessment and its correlation with landslide events using remote sensing data and GIS: a case study at Penang Island, Malaysia. *Environ Monit Assess* 184(2):715–727
- Pradhan B, Lee S, Buchroithner MF (2010) Remote sensing and GIS-based landslide susceptibility analysis and its cross-validation in three test areas using a frequency ratio model. *Photogramm Fernerkun* 1:17–32. <https://doi.org/10.1127/14328364/2010/0037>
- Pradhan B, Mansor S, Pirasteh S, Buchroithner M (2011) Landslide hazard and risk analyses at a landslide-prone catchment area using the statistical-based geospatial model. *Int J Remote Sens* 32(14):4075–4087. <https://doi.org/10.1080/01431161.2010.484433>
- Rai PK, Mohan K, Kumra VK (2014) Landslide hazard and its mapping using remote sensing and GIS. *J Sci Res* 58:1–13
- Regmi AD, Yoshida K, Pourghasemi HR, Dhital MR, Pradhan B (2014) Landslide susceptibility mapping along Bhalubang-Shiwapur area of mid-western Nepal using frequency ratio and conditional probability models. *J Mountain Sci* 11(5):1266–1285
- Roy J, Saha S (2019) Landslide susceptibility mapping using knowledge-driven statistical models in Darjeeling District, West Bengal, India. *Geoenviro* Disaster 6:11
- Saha AK, Gupta RP, Sarkar I, Arora KM, Csaplovics E (2005) An approach for GIS-based statistical landslide susceptibility zonation with a case study in the Himalayas. *Landslides* 2(1):61–69
- Sarkar S, Kanungo D, Ptra A, Kumar P (2006) Disaster mitigation of debris flow, slope failure, and landslides. In: *GIS-based landslide susceptibility case study in Indian Himalaya*. Universal Acadamy Press, Tokyo, pp 617–624
- Sarkar S, Rjan MT, Roy A (2013) Landslide susceptibility assessment using information value method in parts of the Darjeeling Himalayas. *Geo Soc India* 82:351–362
- Shortliffe EH, Buchanan BG (1975) A model of inexact reasoning in medicine; math. *Biosci* 23(3):351–379
- Silalahi FE, Arifianti Y, Hidayat F (2019) Landslide susceptibility assessment using frequency ratio model in Bogor, West Java, Indonesia. *Geosci Lett* 6:10
- Sujatha ER, Rajamanickam GV, Kumaravel P (2012) Landslide susceptibility analysis using probabilistic certainty factor approach: a case study on Tevankarai stream watershed, India. *J Earth Syst Sci* 121(5):1337–1350

- Sun WF (2009) Study of landslide hazard assessment on typical loess area in Qianhe valley, Qianyang County (Ph.D. dissertation: Chinese Academy of Geological Science)
- Temesgen B, Mohammed U, Korme T (2001) Natural hazard assessment using GIS and remote sensing methods, with particular reference to the landslides in the Wondogenet area, Ethiopia. *Physics Chem Earth* 26:665–615
- Van Westen CJ, Castellanos E, Kuriakose SL (2008) Spatial data for landslide susceptibility, hazard, and vulnerability assessment: an overview. *Eng Geol* 102:112–131
- Varnes DJ (1984) Landslide hazard zonation, a review of principles and practice, International Association of Engineering Geology Commission on landslides and other mass movements on slopes. UNESCO, Paris, p 63
- Wang HB, Wu SR, Shi JS, Li B (2011) Qualitative hazard and risk assessment of landslides: a practical framework for a case study in China. *Nat Hazards*. <https://doi.org/10.1007/s11069-011-0008-1>
- Wang Q, Guo Y, Li W, He J, Wu Z (2019) Predictive modeling of landslide hazards in Wen County, northwestern China based on information value, weights-of-evidence, and certainty factor. *Geo Nat Hazards Risk* 10(1):820–835. <https://doi.org/10.1080/19475705.2018.1549111>
- Woldearegay K (2008) Characteristics of a large-scale landslide triggered by heavy rainfall in Tarmaber area, central highlands of Ethiopia *Geophysical Research Abstracts* 10
- Woldearegay K (2013) Review of the occurrences and influencing factors of landslides in the highlands of Ethiopia with implications for infrastructural development
- Wubalem A, Meten M (2020) Landslide susceptibility mapping using information value and logistic regression models in Goncha Siso Eneses area, northwestern Ethiopia. *SN Applied Sci*:807. <https://doi.org/10.1007/s42452-020-2563-0>
- Yalcin A, Reis S, Aydinoglu A, Yomralioglu T (2011) 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* 85:274–287
- Yesilnacar E, Topal T (2005) Landslide susceptibility mapping: a comparison of logistic regression and neural networks method in a medium scale study, Hendek region (Turkey). *Eng Geol* 79:251–266
- Yilmaz I, Keskin I (2009) GIS-based statistical and physical approaches to landslide susceptibility mapping (Sebinkarahisar, Turkey). *Bull Eng Geol Environ* 68:459–471
- Zhang K, Wu X, Niu R, Yang K, Zhao L (2017) The assessment of landslide susceptibility mapping using random forest and decision tree methods in the three Gorges reservoir area. *China. Environ Earth Sci* 76:405
- Zine EL, Abidine R, Abdel Mansour N (2019) Landslide susceptibility mapping using information value and frequency ratio for the Arzew sector (Northwestern of Algeria). *Bull Min Res Exp* 160:197–211. <https://doi.org/10.19111/bulletinofmre.502343>
- Zorgati A, Wissem G, Vali V, Smida H, Essghaier GM (2019) GIS-based landslide susceptibility mapping using bivariate statistical methods. *Open Geosci* 11: 708–726

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com