GIS – Based Logistic Regression Models for Landslide Hazard Mapping: A Case Study From Ensaro District, North Ethiopia

Abdurohman Adem (✉ abdurohman2015@gmail.com)

Debre Berhan University <https://orcid.org/0000-0002-8495-2644>

Suryabhagavan Venkata Karuturi

Addis Ababa University

Tarun Kumar Raghuvanshi

Addis Ababa University

Research

Keywords: Logistic regression, Landslides hazard zone, NDVI, AUC, Receiver Operating Characteristic curve

Posted Date: October 5th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-80954/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Title: GIS – Based Logistic Regression Models for Landslide Hazard Mapping: A case study from Ensaro District, North Ethiopia

Abstract

The present study was undertaken to identify landslides hazard prone areas in North Ethiopia. The landslide hazard in the present study area was evaluated by using the logistic regression model. Seven landslide causative factors were used for the landslide hazard evaluation, these are; slope gradient, slope aspect, elevation, proximity to streams, land-use/ land-cover, lithology and Normalized Difference Vegetation Index. Besides, for the present study landslides inventory data for the period of 2000 to 2018 was collected from the field survey and the Google earth image interpretation. The coefficient for the considered causative factors and classes were used for the identification of landslides hazard index using raster tool in ARCGIS environment. The prediction of the logistic regression model reveals that one third of the study area (32%) is under high hazard zone and the steep slopes and the elevated areas are most susceptible areas. The predicted landslides hazard zonation map is highly correlated with the training data set where 74% of it lies in the very high and high landslide hazard zones. Results of the area under the Receiver Operating Characteristic curve for the training sample, was found to be 0.76 while the area under the ROC curve of the validation sample was 0.71. Thus, the validation results has confirmed the rationality of adopted methodology, considered causative factors and their evaluation in producing LHZ map for the area. Further, the study has forwarded recommendations that can be followed to prevent and mitigate the adverse impact of landslides in the study area.

Keywords: Logistic regression; Landslides hazard zone; NDVI; AUC; Receiver Operating Characteristic curve

Introduction

Landslides are one of the destructive geological processes which have caused major problems in mountainous terrains by killing hundreds of people every year besides damaging roads, bridges, and property (Sarkar and Kanungo, 2004). According to the United Nation International Strategy Disaster Reduction (UNISDR, 2006), Landslide is ranked third in terms of a number of deaths among the top ten natural disasters. In addition, Pradhan and Youssef (2010) indicated that globally landslides account for approximately 1,000 deaths and about \$4 billion property losses every year.

Countries like Ethiopia where most of their terrains in mountainous areas have been seriously affected by landslides because of the terrain related factors and activated by heavy rainfall or earthquake.

Most of the terrains in mountainous areas like Ethiopia have been subjected to landslides under the influence of a variety of terrain factors and triggered by events such as extreme rainfall or earthquake. The frequency and the magnitude of landslides can increase due to human activities, such as deforestation or urban expansion (Subramani and Krishnan, 2015). In particular, in developing countries like Ethiopia; the rapid rate of population growth has demanded new agricultural lands and settlements. As a result the expansion of settlements and life-lines over hazardous areas are increasing the impact of natural disasters like landslides. In many countries, the economic losses and casualties due to landslides are greater than commonly recognized and cause a yearly loss of property larger than any other natural disaster (Schuster and Fleming, 1986; Glade, 1998).

According to Kifle Woldearegay (2013), the hilly and mountainous terrains of the highlands of Ethiopia are affected by the common types of landslides which are triggered by heavy rainfall. According to Bekele Abebe et al. (2010), rugged morphology, high relief energy, and the nature of the outcropping rocks are among the main predisposing factors for the widespread distribution of landslides in Ethiopia. Further, particularly in the last five decades, Landslide in Ethiopia has resulted in the loss of human lives, properties, and infrastructures. From 1960 to 2010 alone, about 388 people were dead, 24 people were injured and a lot of agricultural lands, houses and infrastructures were affected (Lulseged Ayalew, 1999; Berhanu Temesgen et al., 2001; and Kifle Woldearegay, 2008). Similarly, the present study area, the Ensaro District, has been experiencing

landslides hazards for the last few decades. However, in the recent few years, the hazards related to landslides are causing great damage and catastrophe to the local people. Moreover, little attempt has been made to predict them, or to prevent damage arising from landslides.

Thus, the main objective of the present study is to prepare landslides hazard zone using logistic regression. Further, an attempt had been made to validate the predicted landslides hazard model using ROC curve.

2. Overview of the Study area

2.1 Location

Ensaro is one of the Districts in Northern Ethiopia. The District is found in the Amhara Regional state specifically in North Shewa Zone. Ensaro is bordered on the west and south by the Oromia Regional State, on the north by the Jamma River which separates it from Merhabiete District, on the northeast by Moretna Jiru District, and on the east by Siyadebir District. Lemi is the main town of the District. The Geographical location of the study area lies between 9°42'00"–9°58'00"N latitude and 38°47'00"–39°4'00"E longitude covering an area of 432.17 km² (Fig. 1).

[Insert Fig. 1 here]

2.2. Climate and Vegetation

The mean annual temperature and rain fall of the study area range from 20°C to 25°C and 500 to 1600 mm, respectively (ERAO, 2018). The climate of the area ranges from humid to arid climatic condition. Two third of the District area lies in the “kola” agro climatic zone and the rest of the area lies in the “Dega” and “Weina Dega” agro climatic zones. The dominant vegetation types of the area are represented by eucalyptus trees, junipers, scattered acacia girar, thorn bushes, small shrubs and many undifferentiated ever green plants. Green acacia and scrubs are dominantly found in the “kola” agro-ecological zone and along stream and ridges of the District (ERAO, 2018).

2.3. Physiography and geology

Ensaro District is characterized by flat lying topography up lifted up to 2675 m asl with undulating ridge chains and deeply dissected valleys which reaches an altitude of 1171 m asl. The flat lying topography is mostly represented by the tertiary basalts. The deeply dissected

valleys located in south east and the northwestern parts of the study area, which is represented by Jemma River and its tributaries, where most of the mesozoic sediments exposed.

In the present study area, the Mesozoic sediments are exposed in the highly dissected plateau area in the study area(Fig. 2). The typical succession in the study area is stratified sandstone and mudstone of Mesozoic age, situated within the gorge and canyon of Jemma drainage basin. Debrelibanos sandstone is the oldest lithologic unit outcropping in the study District. This unit is exposed in the deep gorges in the study area. It forms very steep cliffs along the stream cuts with a thickness of 50 m at Jemma River section (Tarekengn Taddesse, 2005). The Debrelibanos Sandstone conformably overlies the Muger mudstone and is in turn unconformably overlain by the Tertiary volcanic rocks. The rock is medium-grained, light red (fresh color) to reddish/ yellowish/ pinkish brown (weathering color), sub-rounded grains, poorly to well sorted, compacted and thickly bedded to laminated, including the dominant cross bedding structure. The Muger Mudstone is dominantly found in within the Jemma River in the study area. It forms the gentle slopes and conformably underlies the sandstone. The rock is fine to medium grained, light gray (fresh color) to light/dark brown (weathering color)(Getachew Lemesa, 2008).

The Tertiary volcanic rocks are the earliest volcanic rocks in the study area that occur along the plateau bordering the rift, which were erupted from fissures due to extensional tectonic activity. (Birhanu Ermias et al., 2017).The Lower and Middle Basalt rocks cover the majority of the plateau in the study area. The rocks include fine, medium to coarse grained, dark gray (fresh color) to light/ reddish/ dark/ yellowish brown (weathering color) and aphanitic to porphyritic basalts. Olivine-plagioclase phyric basalt which is also named as lower basalt is characterized by medium to coarse grained, dark gray (fresh color) to dull gray (weathering color), phyric with dominant plagioclase and few olivine phenocrysts. It is massive and blocky in appearance, exhibiting spheroidal weathering. The rock exhibits porphyritic to seriated textures. Whereas Olivine-phyric basalt (middle basalt) is coarse grained, greenish gray (fresh color) to gray (weathering color) and forms a ridge with blocky appearance (GSE, 2010). Further, in the study area, the Quaternary superficial deposits (soil units) predominantly cover the low-lying flat alluvial plains, intermountain depressions and river channels (bed, bank, flood plain). Yet, in hilly relief with medium and even high relief residual, colluvial soils can be also found. They are

derived from the weathering, transportation and reworking of different rocks from the plateau (Jiri Sima et al., 2009).

3. Methodology

Statistical landslide hazard assessments have become very popular, especially with the use of geographic information systems (GIS) and reduce the subjectivity and ensure better reproducibility of the hazard zonation processes. Statistical methods in landslides hazard assessments are used to estimate the relative contributions of the factors responsible for landslides and to generate some predictions based on these factors (Van Westen et al., 2008; Guzzetti et al., 1999).

In this research paper, logistic regression (LR) model which is one of the multivariate statistical methods had been applied to build the landslide hazard zonation map. The model is one of the best-fitting models to explore the relationship between a dependent variable and a set of independent explanatory variables, which could be categorical or continuous (Atkinson and Massari, 1998; Gayen and Saha, 2018).

Its simplest form, the logistic model can be written as:

$$P=1/(1+e^{-Z}) \quad (1.1)$$

Where, P is the estimated probability of landslide occurrence. The probability lies between 0 to 1 having S-shaped curve and Z is the linear combination and its equation is:

$$\text{Prediction (Z)} = (B_0 + B_1X_1 + B_2X_2 + B_3X_3 + \dots + B_n X_n) \quad (1.2)$$

Where, B_0 is the intercept of the logistic model, n is the number of independent variables, B_i ($i = 1, 2, 3, \dots, n$) is the slope coefficient of the model and X_i ($i = 1, 2, 3, \dots, n$) is the independent variable.

4. Landslide Inventory

Landslide inventory can be developed from field surveys, by interpretation of remotely sensed images and Google images interpretation (Xu et al., 2013). For the present study, landslides inventory data for the period of 2000 to 2018 was collected from the field survey and the Google earth image interpretation. Besides, local respondents were also interviewed as a source of

information for the identification of those areas which had experienced landslides in the past (Fig 2).

Hence, in the present study 36 landslide points were identified and divided into two subsets *viz.*, 75% of the landslide data were used for training and the remaining 25% for validating the developed model (Fig 3). In the same way, 36 non-landslide points were randomly selected within the research area and separated also into training and validating data. The landslide data from Google Earth was saved as compatible format (kml) and the data was subsequently changed into shape file in ArcGIS environment.

[Insert Fig. 2 here]

5. Thematic Data Layers

Instability in a terrain is primarily governed the following factors such as geology, slope morphology, drainage, land use, anthropogenic activity, seismicity, and climatic condition. These factors were grouped into intrinsic and external triggering factors. The intrinsic factors include geological factors, geometry of the slope (slope inclination, aspect, elevation and curvature) and land-use/ land-cover and the external trigger landslides factors are rainfall, seismicity (Wang and Niu, 2009; Raghuvanshi et al., 2014). Further, the factors considered in the present study are geology, slope, aspect, elevation, and NDVI, drainage, land-use/ land-cover and profile curvature (Fig. 3a–f). These data layers were derived from different sources. The topographic data used in the analysis include elevation, slope, and aspect (Fig. 3a–c). These data were derived from the ASTER DEM using raster surface in arc toolbox of ArcGIS 10.5 software.

[Insert Fig. 3 here]

The lithology of the study area was digitized from the geological map of Addis Ababa and Debre Birhan which were produced by the Geological Survey of Ethiopia (GSE, 2010). The digitized map was converted in to raster surface using ArcGIS software. Further, the land- use/ land-cover map (Fig.3E) was derived from the Landsat 8 image with 30 × 30 m pixel resolution. The land-use/ land-cover map had been done using the Google Earth images and training area in the field

as a reference for the supervised maximum likelihood classification through ERDAS Imagine software of version 2014. In addition, the Landsat 8 image was used to generate the Normalized Difference Vegetation Index (NDVI) which is one of the indications of the vegetation cover of the earth's surface. It is computed with the equation (eq.3.3) using the ERDAS Imagine software version 2014 (Lillesand et al., 2007).

$$NDVI = (NIR - R) / (NIR + R) \quad (1.3)$$

Where, NDVI is Normalized Difference Vegetation Index, NIR is near infra red band, R is the red band.

6. The computation of Landslides hazard Index

Since the logistic regression has binary dependant outcome (the presence and the absence of landslides), the landslides points and the non landslides points were coded as 1 and 0, respectively. Further, the sub class of categorical thematic data layers such as; land-use and land-cover, Aspect and geology were coded using dummy variables that are using 0 (zero) and 1. However, the value of the continuous thematic layers such as; the NDVI, Slope, Proximity to streams, Elevation was used as it was.

In order to generate the presence or the absence of landslides points from each thematic layer, all thematic layers were loaded in the ArcGIS 10.5 environment. After loading the layers, using the “multipoint extraction” tool from the arc tool box in ArcGIS, “landslide” and “no landslide” features were selected as a target source to extract the information from each thematic layer one by one. Then, the extracted data from each thematic layer were exported as dbf format for further analysis in statistical software of IBM SPSS. Using logistic regression analysis tool in SPSS, the co-efficient of each thematic layers and the intercept were determined. Therefore, the prediction map of the study area was produced using equation (1.1).

$$P(\text{Landslide probability}) = 1 / 1 + \exp - [B_0 + (B_1 * \text{Slope}) + (B_2 * \text{stream dis}) + (B_3 * \text{NDVI}) + (B_4 * \text{Elevation}) + (\text{Aspect}_c) + (\text{Geology}_c) + (\text{Landuse}_c)]$$

7. RESULTS AND DISCUSSIONS

7.1 Slope Gradient

Slope gradient is important with regard to landslide initiation (Raghuvanshi et al., 2015; Fikre Girma et al., 2015; Gemechis Chimidi et al., 2017). The digital elevation model (DEM) is used to extract the slope gradient of the study area. Thus, the slope gradient map of the study area was classified in to six classes; 0 - 7°, 7 - 15°, 15 - 25°, 25 - 35°, 35 - 45° and >45° as shown in Fig. 3b.

From the past landslides inventory data, most of the landslides (57%) has occurred in the slope classes 15 - 25° and 25 - 35° (Fig.4). About 16% of the landslides have occurred in the slope class 25 - 35° and 13% of it has occurred in the slope class above 45°. The slope class in which least landslides (3%) was observed is the slope class 0 - 7°. Hence, it is clear that the slopes which are gentle have experienced the highest proportion of the landslides. This might be because such slope gradients are favorable for agricultural practice. As Raghuvanshi et al. (2014) stated, the increase in the soil moisture because of irrigation practice in the gentle slope may initiation landslides by creating pore pressure within the slope material.

[Insert Fig. 4 here]

7.2 Elevation

The elevation map of the study area was derived from the DEM and the maximum and the minimum elevation of the study area is 2675 m asl and 1171 m asl, respectively (Fig.3a). The elevation map was classified on the basis of natural breaks in to four classes. The elevation class in the range of 1171–1587 m asl covers the largest area followed by the elevation class in the range of 1587–1936 m asl which accounts for 28.28% of the study area. About 43% of the area lies in the remaining elevation classes of the study area (Fig. 3a).

With regards to the past landslides distribution, the highest distribution of landslides (52.54%) was observed in the elevation class of 1935-2361 m asl. About 26% of the past landslides have occurred in the elevation class 2361–2675 m asl. The elevation class 1171–1587 m asl had experienced the least (2.1%) landslides in the study area (Fig.7).

Thus, the result of the present study reveals that landslides and elevation have a strong relationship. That is with the increment of elevation in the study area, the occurrence of past landslides was getting higher (Fig.5). Even in most slope instability studies, elevation is considered to be the principal factors for landslides initiation (Raghuvanshi et al., 2015; Gemechis Chimidi et al., 2017). That is the amount of precipitation, weathering processes and resulting weathering depths and land use are all influenced by elevation (Guzzetti, et al, 2001; Dai and Lee, 2002). Therefore, the more intense erosion and weathering, the more will be the influence of elevation on landslides (Raghuvanshi et al., 2015).

[Insert Fig. 5 here]

7.3 Aspect

Aspect is the geographical direction in which slope face is oriented. Aspect is considered to be another important factor that influences the stability of a slope (Raghuvanshi et al., 2015). Depending on the slope aspect, the amount of rainfall on a particular slope may vary. In general, by controlling the concentration of the soil moisture, slope aspect can have an influence on the density and distribution of landslides in an given area (Wieczorek et al., 1997). Further aspect of the slopes can also control the orientation of tectonic fractures. Ayalew and Yamagishi (2002) have reported variation in the degree of weathering and basal erosion as an effect of slope aspect. Thus, aspect has an significant effect on landslide distribution (Raghuvanshi et al., 2015).

The Aspect map of the study area was classified in to ten classes, as indicated in Fig.3c. Relatively, the most dominate slope aspect with respect to their area coverage are slopes that are oriented towards northeast (15%), west (14%) and the northwest (13%). In terms of the past landslides distribution, the slope that are oriented towards south (20%) experienced the highest number of landslides followed by the slopes that are oriented towards east (14%) and in the southwest (14%) directions. The aspect class flat did experienced little number of landslides.

Further, Dai and Lee (2002) stated that aspect influences local climatic factors such as exposure to sunlight, dry wind and amount of moisture. Hence, the slopes oriented towards south in the study area might experience more moisture which possibly may initiate the landslides. However, it can be said that slope aspect and landslides in the present study area do not demonstrate strong

correlation since most of the aspect class has experienced more or less similar percentage of past landslides.

7.4 Land-use/land-cover

Land-use/ land-cover is one of the potential factors for instability of slopes (Anbalagan, 1992; Raghuvanshi et al., 2014; 2015). The land-use/ land - cover classes that were identified in the present study area are vegetation, barren land, drainage land, settlement and agricultural Land (Fig. 4.6). Among the land-use/ land-cover classes, agricultural land (40%) and barren land (32%) constituted the largest area coverage in the study area (Fig.3e). With respect to the past landslides distribution, the highest number of landslides (43%) has occurred in the agricultural land followed by the barren land (24%).

High number of landslides in agricultural land in the study area may be related to the fact that improper cultivation and barren areas demonstrate more erosion, thus greater chance will be for the occurrence of landslides as compared to the protected forests which minimize the effect of climatic agents on the slope faces (Engdawork Mulatu et al., 2009).

[Insert Fig. 6 here]

Further, the high concentration of landslides within the agricultural land is possibly related to two aspects; firstly most of the cultivated lands have weak soils, which are generally highly susceptible to landslides and secondly activities related to cultivation for example; irrigation results into saturation of soil mass which resulting into reduction of shear strength of soil mass (Raghuvanshi et al., 2014; 2015).

7.5 Normalized Difference Vegetation Index (NDVI)

Vegetation has great role in bonding the slope material so as to reduce slope instability. Hence, a slope with dense vegetation is less likely to the incident of shallow landslides than barren slopes, while all other factors remain constant (Richard, 2005). According to Lillsand et al. (2007), the value of NDVI ranges from -1 to $+1$; where the values closer to $+1$ means higher vegetation cover whereas values closer to 0 means little or no vegetation cover.

For the present study area, the NDVI value falls in the range -0.128 and 0.530 (Fig.3f). The past landslides distribution showed that there was a decreasing trend of landslides with an increase in

the NDVI values. High percentage of landslides (31.17%) was observed in the NDVI range of -0.129 to 0.101 whereas, in the NDVI class of $0.225-0.531$, the lowest number of landslides (18.18%) was observed in the study area (Fig. 7).

The results of the present study reveals that the NDVI values and the distribution of the past landslides have relatively strong relationship; where the higher NDVI values were determined, the lower past landslides occurrence was observed and vice versa.

[Insert Fig. 7 here]

7.6 Proximity to streams

In general, proximity to the streams may trigger the landslides. Streams can cut deep in to the slopes on the sides of the stream banks and makes them susceptible for instability. Also, the toe of slopes along the stream course may be removed by stream water that may make slope to overhang. Thus, the above slope material may become susceptible for failure (Raghuvanshi et al., 2014; Gemechis Chimidi et al., 2017).

In order to incorporate effect of streams in landslide evaluation process, proximity to streams to slopes was worked out by using Euclidean distance function in ArcGIS 10.2. The proximity to stream in the study area was classified in to six categories; 0–300 m, 300–600 m, 600–900 m, 900–1200 m and streams at distance greater than 1200 m (Fig.8).

[Insert Fig. 8 here]

The proximity to streams in the class range between 0–300 m has the largest areal coverage (57.42%) whereas the proximity to streams which are greater than 1200 m covers very small area (12.25%). Further, the distribution of past landslides showed that most of the landslides (51%) have occurred in the class 0–300m and about 30.55% of the landslides have occurred in the class 300–600 m. The lowest landslides distribution (2.78%) occurred in the class 900–1200 m (Fig. 9).

The present study has clearly showed that the presence of streams greatly influences initiation of landslides in the present study area. This was also observed during the field survey. As discussed by Mathew, et al. (2007) and Bathrellos, et al. (2009), presence of streams greatly influences the

stability by toe erosion or by saturating the slope material. Thus, the closer the slope to streams, the more likely will it get water and develop pore pressure.

[Insert Fig. 9 here]

7.7 Lithology

Lithology greatly controls the stability condition of the slopes. Since rocks have varying weathering process, their degree of vulnerability to landslides is different. For instance, rocks like quartzite, limestone and igneous rocks are mostly hard, massive and resistant to erosion, hence form steep slopes. Whereas, sedimentary rocks are more vulnerable to weathering and may easily contribute to landslides (Anbalagan, 1992; Raghuvanshi et al., 2014; Henok Woldegeorgis et al., 2014; Engdawork Mulatu et al., 2010).

The lithology map of the study area was derived from the Geological Map of Addis Abeba and Debre Birhan, prepared by Geological Survey of Ethiopia (GSE, 2010). The lithology present in the study area includes; Lower Basalt, Middle Basalt, Debre Libanos Sandstone, Muger Sandstone, Quaternary Deposit and Antalo Limestone (Fig.5.11). The largest portion of the study area (36%) is covered by lower and middle basalt followed by Muger sandstone which constituted 22% of the study area. Further, the quaternary superficial deposit covers the smallest area (6%) (Fig.14). Further, past landslide distribution among various lithology showed that 41% of the landslides have occurred in Debre libanos sandstone, 30% within Middle basalts and 22% were recorded in the Lower basalts. The least number of landslides were observed in the Muger mudstone (6%) in the study area (Fig. 10).

In the present study, Debre libanos sandstone, which is the sedimentary rock, has experienced maximum numbers of landslides. On the other hand, as it was observed in the field, Lower basalt and the Middle basalt also experienced relatively higher number of landslides. These lithological units form the steeper slope sections in the study area.

[Insert Fig. 10 here]

8. Profile curvature

Profile curvature is the curvature in the downslope direction along a line formed by intersection of an imaginary vertical plane with the ground surface (Ohlamacher 2007). Slope shape has a strong influence on slope stability in steep terrain by converging or diverging surface and primarily subsurface water in the landscape. Concavity or convexity of the surface is determined by the sign of the curvature value is important for determining i.e. negative and positive values represent concave and convex surfaces, respectively. (Pradhan, 2010).

[Insert Fig. 11 here]

The map of profile curvature of the study area (Fig.11) shows that more than two third of the study area (65%) fall in the plain curvature and convex curvature covers the least in area coverage (11%). Regarding to the past landslides distribution, concave curvature experienced relatively the largest landslides (46%) followed by the plain curvature (33%).

Thus, the profile curvature of the study area shows similar result as others research finding revealed (Lulseged Ayalew and Yamagishi 2005;; Matebie Meten et al., 2015). Hence, concave surface tend to concentrate subsurface water into small areas of the slope, thereby generating rapid pore water pressure increase during storms or periods of rainfall. If pore pressures develop in the hollow, the soil shear strength reduces to a critical level and a landslide can occur. The divergent or convex landform is most stable in steep terrain where the surface water will diverge. (Ohlamacher 2007).

10. Landslide hazard zoning (LHZ) mapping

In the present study, the spatial relationship between landslides occurrence and factors influencing landslides was assessed by using the logistic regression model. For analyzing landslide occurrence, logistic regression model is an appropriate tool, where the dependent variable should be categorical (e.g., presence or absence) whereas the explanatory (independent) variables can be categorical, numerical, or both (Chang et al., 2007). The logistic regression model has the form:

$$\text{logit}(Z) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i \quad (\text{eq.4.1})$$

Where; Z is the dependent variable (landslide occurrence), X_i is the independent or explanatory variable (potential causal factors of landslide phenomena), β_0 is a constant, and β_i is the i-the regression coefficient.

$$p = (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i) / 1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i) \quad (\text{eq.4.2})$$

$$p = 1 / 1 + \exp^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i)} \quad (\text{eq.4.3})$$

Hence, the result of the LR produced the intercept of the model and the value of the coefficients for all independent variables from the final step of the logistic regression analysis model.

These coefficient values (Table 4.1) were assigned as weights for the individual independent variables. Later, by using all these coefficients and intercept value, the predicted probability (LHZ) for the present study area was calculated by using the following equation (eq.4.3):

$$P (\text{Landslide probability}) = 1 / 1 + \exp - [-65.552 + (.318 * \text{Slope}) + (.126 * \text{stream dis}) + (\text{Aspect}_c) + (62.108 * \text{NDVI}) + (.021 * \text{Elevation}) + (\text{Geology}_c) + (\text{Landuse}_c) \dots \dots \dots (\text{eq.4.3})$$

Landslide hazard zoning (LHZ) map was classified into five categories of hazard such as; very high hazard zone (VHH) (0.81–0.99), high hazard zone (HH) (0.58–0.81), moderate hazard zone (MH) (0.37–0.58), low hazard zone (LH) (0.18–0.37) and very low hazard zone (VLH) (0.00–0.18), respectively (Table 2). From Table 4.2, it can be observed that 30.12% of the study area falls within very low landslides hazard zones and 21.13% of it in low hazard zone. High hazard zone (14.32%) and very high hazard (18.15%) zone cover relatively small areas.

Regarding the validating the predicted landslides hazard zoning using the training dataset (Table. 2), 52.12% and 22.10% of the observed landslide dataset were found in very high hazard (VHH) and high hazard (HH) zones whereas 2.67% and 12.11 % of the observed landslide points falls in the areas with very low (VL) and low hazard (LH) classes. Therefore, about 74% of the past landslide points were found to be concentrated in the areas that are delineated as high hazard zones and very high hazard zones.

[Insert Table. 1 here]

[Insert Table. 2 here]

Further, the goodness of fit of the model is evaluated by using pseudo- R^2 . The R^2 values of the logistic regression model for the training dataset were 0.619 and 0.826 for Cox and Snell R^2 and Nagelkerke R^2 , respectively. Thus, the regression model indicated a good performance (Table 3).

According to Lulseged Ayalew and Yamagishi (2005), the higher R^2 indicates the extent to which the model fits the data that is, if the R^2 value approaches to the value 1, the model fits the data perfectly whereas, if the R^2 approaches to 0, there is no relationship with the data.

[Insert Fig. 12 here]

[Insert Table. 3 here]

In addition to the R^2 , the classification matrix of the model has indicated that the model has predicted correctly 79.2% of the training datasets (Table 4).

[Insert Table. 4 here]

8.1 Model validation

One of the most recognized methods for obtaining a good internal validation of a model is data splitting in which the training data is randomly split into two parts; one to develop the model, and another to measure its performance (Chung and Fabbri, 2003). In the present study, the validation dataset has shown that 69% of the past landslides were observed in the high and very high hazard zones and 17% of it overlay in the very low and low hazard zones. The Receiver Operating Characteristic (ROC) is another summary measure of the model's predictive power (John, 2009) (Fig. 4.15). In the present study, the area under the ROC curve for the training sample was 76% while the area under the ROC curve for the validation sample is 71%. According to Hosmer and Lemeshow (2000), the value of the area under the ROC curve between 0.7 and 0.8 of the model is considered to have acceptable discrimination. Hence, the logistic regression model used in the present study, for the prediction of the Landslides hazard zonation (LHZ) is considered to be acceptable.

A perusal of Table 4.2 Clearly shows that about 74.12% of the past landslides validates with the prepared LHZ map, as about 74.12% of past landslides fall within very high or high hazard zones. The remaining 25.88% landslides, which do not validate and fell within moderate hazard,

low hazard or no hazard zones, may possibly be on account of limitations of the methodology that was followed during the present study.

[Insert Fig. 13 here]

Further, the logistic regression statistical approach followed during the present study was applied at a medium scale (1:50,000) and certain factors, which may be responsible for instability, cannot be considered at this scale (Shiferaw Ayele et al. 2014; Fikre Girma et al., 2015; Gemechs Chimidi et al., 2017). The stability condition of a slope is affected by factors such as shear strength of slope material, water forces within soil and rock mass, and characteristics of discontinuities. However, these factors can only be considered when deterministic approach at detailed scales (> 5000) is followed on individual slope basis (Tilahun Hamza and Raghuvanshi 2017; Gemechs Chimidi et al., 2017). Thus, the validation results has reasonably confirmed the rationality of adopted methodology, considered causative and triggering factors and their evaluation in producing LHZ map for the present study area. Thus, the high hazard zone depicted in the LHZ map may be safely considered vulnerable for any future development in these zones and may require further detailed slope stability studies for the implementation of any developmental activities within very high or high hazard zones.

Conclusion

Landslide hazard zonation maps provide fundamental knowledge of the causes and effective factors on landslide occurrence and can be effective in hazard management and its mitigation measures. Therefore, in the present research an attempt was made to identify landslides hazard areas in Ensaro District of Northern Ethiopia. The present study area, Ensaro District, is specifically found in the North Shewa Zone of the Amhara Regional State in Ethiopia. The study area is situated at the plateau flat lying topography which is up lifted to 2675 m asl., with undulating ridge chains and deeply dissected valleys which reaches an altitude of 1171 m asl. The flat lying topography is mostly represented by the Tertiary basalts. The deeply dissected valleys which are located in south east and the northwestern parts of the study area are represented by Jema River and its tributaries, where most of the Mesozoic sedimentary rocks are exposed.

Landslide hazard map of the study area was made by establishing the relationship between landslides and casual factors. In the present study, eight landslide controlling factors were considered for the analysis which includes: slope gradient, slope aspect, elevation, proximity to streams, land use/ land cover, lithology , Normalized Difference Vegetation Index (NDVI) and profile curvature. For this, the database on landslide inventory and the causative factors were collected from various sources, both primary and secondary sources and were later preprocessed. The topographic data used in the analysis include elevation, slope, aspect and curvature. These data were derived from the ASTER DEM using raster surface in arc toolbox of ArcGIS 10.5 software. The lithology of the study area was digitized from the geological map of Addis Ababa and Debre Birhan which were produced by the Geological Survey of Ethiopia. Further, the land-use/ land-cover map was derived from the Landsat 8 image with 30×30 m pixel resolution. The land- use/ land-cover map had been done using the Google Earth images and training area in the field as a reference for the supervised maximum likelihood classification through ERDAS Imagine software of version 2014. In addition, the Landsat 8 image was used to generate the NDVI. In order to incorporate effect of streams in landslide evaluation process, proximity to streams to slopes was worked out by using Euclidean distance function in ArcGIS 10.5.

In the present research, logistic regression analysis was implemented in order to obtain the weights for every factor and class using direct pairwise comparison by using statistical package software IBM SPSS , later based on these weights, thematic maps of factors were combined by weighted overly techniques and the landslide hazard zonation map of the study area was created. The relationship between the past landslides distribution and the causative factors reveals that slope gradient, elevation and NDVI have a strong positive relationship than the rest of the environmental factors considered in the study. In addition, agricultural and barren land had strong statistical significant relation to past landslides distribution where 43% of landslides have occurred within the agricultural land and 24% of the landslides has occurred on barren land.

The predicted map of the landslides hazard zonation was classified into five hazard zones which specified that the high and very high hazard zones are distributed over about 32% of the total area, while about 41% of the area falls into low and very low susceptible zones and 27% of the study area falls under moderately hazard zone. Further, the landslide hazard zonation map was validated by

using the training and validation landslides data based on the area under the ROC curve method, by which the prediction precision of 76% and 71% was established, respectively.

The results showed that about 74.12% of the past landslides validates with the prepared LHZ map. The remaining 25.88% landslides, which do not validate and fell within moderate hazard, low hazard or no hazard zones, may possibly be on account of limitations of the methodology that was followed during the present study. Factors such as shear strength of slope material, water forces within soil and rock mass, and characteristics of discontinuities are responsible in defining stability condition of a slope. Thus, the validation results has reasonably confirmed the rationality of adopted methodology, considered causative and triggering factors and their evaluation in producing LHZ map for the present study area. The very high and high hazard zone depicted in the LHZ map may be considered vulnerable for any future development and may require more detailed studies before the implementation of developmental activities.

Declaration

Acknowledgements

The authors are thankful to the local respondents and the community residing in the area for providing important information and the support in the field data collection process.

Authors' contributions

The first author, Mr. Abdurohman Yimam, has mostly participated in the whole process including fieldwork, data collection, database preparation, compiling the results, and taking comments from Dr. **Suryabhagavan and**, address the comments and finalize the draft for journal submission after a consensus is reached with them. The Second Author, Dr. **Suryabhagavan**, has participated in the inception and design of this paper and helped greatly in data preparation and analysis. The third Author, Dr. **Tarun**, has also been involved in a detailed review of the manuscript before submission.

Funding

No funding was used for this manuscript preparation.

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Competing interests

The authors have no competing interest.

Authors' information

Not applicable.

List of Abbreviation

AUC	Area Under Curve
DEM	Digital Elevation Model
GIS	Geographic Information System
GSE	Geological Survey of Ethiopia
GSI	Geological Survey of Ireland
LHZ	Landslide Hazard Zonation
LR	Logistic Regression
LULC	Land Use and Land Cover
NDVI	Normalized Difference Vegetation Index
OLI	Operational Land Imagery
ROC	Receiver Operating Characteristic
USGS	United States Geological Survey

References

- Anbalagan, R. (1992). Landslide hazard evaluation and zonation mapping in mountainous terrain. *Eng. Geol.* **32**: 269-277.
- Atkinson, P.M.; Massari, R. (1998) Generalised linear modelling of susceptibility to landsliding in the central Apennines, Italy. *Comput. Geosci.*, 24, 373–385.
- Bathrellos, G.D., Kalivas, D.P. and Skilodimou, H.D. (2009).GIS-based landslide susceptibility mapping models applied to natural and urban planning in Trikala, Central Greece. *Estudios Geológicos.* **65** (1): 49–65.
- Bekele Abebe, Dramis, F. , Fubelli , G., Mohammed Umer and Asfawossen Asrat. (2010). Landslides in the Ethiopian highlands and the Rift margins. *Journal of African Earth Sciences* **56**:131–138.
- Berhanu Temesgen, Mohammed Umar and Tesfaye Korme .(2001). Natural hazard assessment using GIS and remote sensing methods with particular reference to the landslides in Wondogenet area, Ethiopia. *Physical chemistry Earth.* **26** (9): 665- 675.
- Birhanu Ermias, Raghuvanshi,T.K and Bekele Abebe. (2017). Landslide Hazard Zonation around the Route from Alemketema Town to Ambat Village, North Shewa, Ethiopia.GIS based expert evaluation approach,International. *Journal of Earth Sci. & Engineering.***10**(1): 33-44.
- Chung, C.J. F. and Fabbri, A. G. (2003). Validation of spatial prediction models for landslide hazard mapping, Natural Hazards. **30**(3):451–472.
- CSA (Central Statistical Authority) .(2008). Summery and Statistical Report of the 2007 Population and Housing Census. Addis Ababa , Ethiopia
- Dai, F.C. and Lee, C.F.(2002). Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology.* **42**:213–228.
- Engdawork Mulatu, Raghuvanshi,T.K., and Bekele Abebe .(2009).Landslide Hazard Zonation Around Gilgel Gibe-II Hydroelectric Project, Southwestern Ethiopia Sinet: *Ethiop. J. Sci.* **32**(1):9–20.
- Fikre Girma, Raghuvanshi, T.K., Tenalem Ayenew and Trufat Hailemariam. (2015). Landslide hazard zonation in Ada Berga District, Central Ethiopia – a GIS based statistical approach. *J. Geomatics.* **90**:25–38.
- Henok Woldegeorgis , Raghuvanshi T.K., and Balemwal Atnafu. (2014). Landslide hazard zonation using expert evaluation technique: a case study of the area between Gohatsion

- town and the Abay (Blue Nile) River, Central Ethiopia. *SINET: Ethiopian Journal of Science*. **37**(2):75-94
- Kifle Woldearegay (2008) Characteristics of large-scale landslide triggered by heavy rainfall in Tarmaber area, Central Highlands of Ethiopia. *Geophysical Research* **10**:
- Kifle Woldearegay (2013). Review of the occurrences and influencing factors of landslides in the highlands of Ethiopia: With implications for infrastructural development. *Momona Ethiopian Journal of Science (MEJS)*, **1**:3–31.
- Gayen, A.; Saha, S.(**2018**) Deforestation probable area predicted by logistic regression in Pathro river basin: A tributary of Ajay River. *Spat. Inf. Res.*, 26, 1–9.
- Gemechis Chimidi, Raghuvanshi, T.K , Suryabhagavan, K.V.(2017). Landslide hazard evaluation and zonation in and around Gimbi town, western Ethiopia – a GIS-based statistical approach. *Applied Geomat.* 9(4): 219-236
- Glade,T.(1998).Establishing the frequency and magnitude of landslide-triggering rainstorm events in New Zealand. *Environmental Geology*. **35**:2–3.
- GSE (Geological Survey of Ethiopia) (2010). Geological Mapping of Debre Birhan Map Sheet (NC 37-11). Compilers, Daniel Meshesha, Dejene Hailemariam and Abrham Mamo Unpublished report, Addis Ababa, Ethiopia.
- Getachew Lemessa (2008). Summary review and new perspectives in Geotechnical Investigation & Mineral exploration at the Geological Survey of Ethiopia (GSE). Unpublished pdf slides presented at Bangkok, Thailand, 31 pp.
- Guzzetti, F., Carrara, A., Cardinali, M., Reichenbach, P. (1999). Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, central Italy. *Geomorphology*. **31** (1–4): 181–216.
- Jiri Sima et al. (2009). *Water Resources Management and Environmental Protection Studies of the Jemma River Basin for Improved Food Security*. AQUATESST a.s., Geologicka 4, 152 00 Prague 5, Czech Republic, (1st edition), ISBN 978-80-254-5021-5, 225 pp.
- Lemeshow,S. and Hosmer, D. W. (2000). The use of goodness-of-fit statistics in the development of logistic regression models, *American Journal of Epidemiology*. **115**:92–106.
- Lillesand, T.M., Kiefer, R.W. and Chipman, J. (2007). *Remote Sensing and Image Interpretation* .6th ed. John Wiley and Sons, New York.

- Lulseged Ayalew (1999) The effect of seasonal rainfall on landslides in the highlands of Ethiopia. *Bull Eng Geol Env.* **58**:9–19, 7
- Lulseged Ayalew and 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.
- Matebie Meten, Bhandary, NP. and Yatabe, R. (2015). GIS Based Frequency ratio and logistic regression modeling for landslides susceptibility mapping of Debre Sina Area, North Ethiopia. *Journal of mountain science.* **12**(6):1355-1372.
- Mathew, J., Jha, V. K. and Rawat, G. S. (2007). Weights of evidence modelling for landslide hazard zonation mapping in part of Bhagirathi valley, Uttarakhand. *Current Science.*, **92**(5): 628–638.
- Ohlamacher, G.C. (2007). Plan curvature and landslide probability in regions dominated by earth flows and earth slides. *Eng Geol* 91:117–134
- Pradhan, B. (2011). Manifestations of an advanced fuzzy logic model coupled with Geo-information techniques to landslide susceptibility mapping and their comparison. *Environ Ecol Stat.* 18:471-493.
- Pradhan, B. and Youssef, A. M. (2010). Manifestation of remote sensing data and GIS on landslide hazard analysis using 430 spatial-based statistical models. *Arab J Geosci.* 3:319-326
- Raghuvanshi, T.K., Lensa Negassa, Kala P.M. (2015). GIS based Grid overlay method versus modeling approach: A comparative study for landslide hazard zonation (LHZ) in Meta Robi District of West Showa Zone in Ethiopia. *The Egyptian Journal of Remote Sensing and Space Sciences.* **18**:235–250.
- Raghuvanshi, T.K., Jemal Ibrahim, and Dereje Ayale. (2014). Slope stability susceptibility evaluation parameter (SSEP) rating scheme – an approach for landslide hazard zonation. *Journal. Afr. Earth Sci.*, **99**:595–612.
- Raghuvanshi, T.K. (2019). Governing factors influence on rock slope stability – Statistical analysis for plane mode of failure. *Journal of King Saud University*, Published online January 7, 2019. <https://www.sciencedirect.com/science/article/pii/S1018364718314320>
- Sarkar, S. and Kanungo, D.P. (2004). An Integrated Approach for Landslide Susceptibility Mapping Using Remote Sensing and GIS. *Photogrammetric Engineering & Remote Sensing.* **70**(5):617–625.

- Schuster, R.L. and Fleming, R.W. (1986). Economic Losses and Fatalities due to Landslides, *Bulletin of the Association of Engineering Geologists*. **23**(1): 11-28.
- Subramani, T. and Krishnan, N. (2015). Landslides Hazardous Zones by Using Remote Sensing And GIS. *International Journal of Application or Innovation in Engineering & Management (IJAIEEM)*. **4**(5): 211-222.
- Tarekegn Taddesse (2005). Geology. 2005 Geology and Geomorphology EAE Encyclopedia, pp. 1-3.
- United Nations International Strategy for Disaster Reduction (UNISDR) (2006). Making the case for disaster risk reduction in Africa , accessed in November 21, 2018 from <https://www.unisdr.org>
- Van Westen, C. J., Castellanos, E., & Kuriakose, S. L. (2008). Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. *Engineering Geology*. **102**(3-4):112-131.
- Wieczorek, G.F., (1984). Preparing a detailed landslide-inventory map for hazard evaluation and reduction. *Bulletin Association of Engineering Geologists*. **21**(3):337-342.
- Wang, X. and Niu, R. (2009). Spatial forecast of landslides in three gorges based on spatial data mining. *Sensors*. **9**:2035-2061.
- Xu, C., Xu, X.W., Dai, F.C., Wu, Z.D., He, H.L., Wu, X.Y., Xu, S.N. and Shi, F. (2013). Application of an incomplete landslide inventory, logistic regression model and its validation for landslide susceptibility mapping , Wenchuan. *China Nat. Hazard*. **68** (2):883-900.

Figures

Figure 1

Location map of the study area

Figure 2

Landslides location map

Figure 3

Causative factor maps; (a) Elevation, (b) slope gradient , (c) aspect, (d) NDVI, (e) Lithology and (f) land use/land cover

Figure 4

Distribution of the past landslides with respect to slope gradient

Figure 5

Past landslides in relation to Elevation

Figure 6

Past landslides and areal coverage in relation to land use/cover

Figure 7

Past landslides distribution in relation to NDVI

Figure 8

Proximity to streams map of the study area

Figure 9

Past landslides distribution in relation to proximity to stream

Figure 10

Past landslides distribution in relation to the Lithology

Figure 11

The profile curvature map of the study area

Figure 12

Landslides Hazard Zonation Map of the study area

Figure 13

Receiver operating characteristic (ROC) curve of the training sample of the model