ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES DEPARTMENT OF EARTH SCIENCES

Landslide Susceptibility Modeling Using Logistic Regression and Artificial Neural networks in GIS: a case study in Northern Showa area, Ethiopia**.**

A thesis submitted to the school of graduate studies,
Addis Ababa University
In partial fulfillments for the degree of Master of Science in Remote Sensing and GIS.

Advisor: Seyd Ali (PhD.)

By Leta Alemayehu.

February 28, 2007.

Table of Contents

List of tables

List of figures

List of Annexes

Acknowledgment

Achievements and accomplishments whatever the results could be are the futile effects of cooperation and sharing of burden. If it had not been for the various institutes and individuals keen support and good will, I would not have come to the end of my thesis work. The cooperation of Dr. Seid Ali (Assoc. Prof, AAU) to become my advisor on such a short notice from the GIS unit of the Earth's Science Department has put in great momentum on my work. Also the initiator and my previous course coordinator Dr. Lulseged Ayalew (Assoc. Prof, AAU) deserves my keenest appreciation and acknowledgement for taking all the time to listen to my amateur's inception of the scientific work in earth sciences and his advice on the recent developments in the methodologies and provision of most valuable data. I fully acknowledge the effort of Dr. Dagnachew Legesse (Asst. Prof, AAU) to get me an adviser at critical times so that I can enjoy the opportunity to undertake a scientific investigation by myself for the first time. Thanks also goes to the institutes of Ethiopian Map Mapping Agency and the Geological Survey of Ethiopia for provision of the necessary manuscripts and the Addis Ababa University Graduate studies program and specially the Earth's Science Department for the partial funding to complete my thesis.

The technical support I got from Tilaye Yismaw, a graduate student at the University of Edinburgh, departments of Informatics, has been invaluable and useful to the extent I could not have imagined. The troubles taken by Endalkachew Getaneh, graduate student at ITC (Netherlands) in the GIS department, to send me satellite images necessary for this thesis work are unforgettable. Finally I wish all the best for those who supported me with numberless list of names: my parents, Colleagues, the cooperative secretaries of the Earth sciences department; and fellow students at remote sensing and GIS stream in AAU are heart fully thanked.

Abstract

Landslide susceptibility mapping has been undertaken using topographic parameters and lithologic information in north central Ethhiopia, Debresina area. The area has been one well known for landslide occurrences throughout its history.

Before conducting the GIS analysis, extraction of some of the remote sensing data has been undertaken involving rectification and enhancement of Landsat ETM+ image of May 2000 for lithologic units interpretations. Image classification of 5/7, 5/4, ¾ band combination resulting in 70% accuracy and another classification using the first principal componets which accounted for 98 % of the whole bands data was done resulting in 69% accuracy. Further visual interpretation of the lithologic units has been undertaken to produce the final lithologic map interpretation. In addition the digital elevation model (DEM) of the area was obtained at 20m resolution by vectorizing contours from the topographic map of the area to extract the topographic parameters.

In this study two susceptibility mapping methods have been employed: Logistic regression and Artificial Neural Network (ANN). Both have been used to generate the weights to represent the degree of contribution of selected seven parameters: Lithology, Slope, Aspect, Plan curvature, Profile curvature and flow accumulation in different platforms than geographic information system (GIS). Preceding to these the class weights of the various parameters were obtained by BSA method. Finally raster calculation of the seven layers of the parameters was conducted and two susceptibility maps were produced. The weights generated by ANN signified the contribution of Planar curvature, Aspect and Slope type towards landslide occurrence while that of the Logistic regression method signified Lithology and flow accumulation scoring higher in contribution towards landslide occurrence than the rest of the parameters. Finally, the outputs have been evaluated using the inventory of slope failure from the same period as those used for training the models and one from the recent massive landslide that occurred in the area.

1. Introduction

1.1. Rationale

Upon completion of the postgraduate study at Addis Ababa University a thesis work on Landslide susceptibility assessment has been undertaken. The thesis work focuses on Landslide susceptibility assessment of a vulnerable area: Debresina and the surrounding. The subject chosen is in such a way that it goes well with the authors' background, Geology. Hence knowledge gained from various term paper deliveries, classroom discussions, as well as some practical sessions and projects under the supervision of the instructors has been applied to the subject gaining momentum as more and more devastating Landslide hazards occur posing a serious problem to the society.

1.2. Introduction

1.2.1 Background

Slope instability (mass wasting) is a continuous and natural geologic process in countries like Ethiopia. It is the process of forming stable landform, which due to the internal processes, like volcanic activity, might form a rather dynamic (raised) landmass which is not stable. The brief view of the country's surface portrays many paleo-landforms indicative of the mass wasting process (Ayalew et.al, 2004). Many instances of such mass wasting processes (referred here after as landslide) occurrence have been reported (Tarekegn, 1993; EIGS*, 1979; Temesgen, 2001). They seldom result in infrastructure and property damage, resettlement and often casualty. The landslide processes are usually associated with topographic features nevertheless; other factors are also significantly associated with the process such as vegetation, drainage, geologic discontinuities (Temesgen, 2001; EIGS, 1979).

Landslide hazard assessment is the evaluation of occurrence of a certain probable landslide and its consequence on elements at risk (environment, people, property that can be endangered by a landslide) (M. Wise et al., manuscript). According to Turner (1996), [1]often assessment based on the probability of occurrence of a certain hazard distribution is a matter for physical models (Deterministic models) that needs enormous number of test data and determination of parameters for empirical formulas. In addition, most hazard models do not

* EIGS: Ethiopian Geological Survey

incorporate the role of temporal distribution because of the difficulty which according to Dikau et al. (1996), Ayalew (2005) they should have. Hence most hazard analysis hardly can produce hazard maps with this respect at the wanted fidelity. This in turn led to their limited application in GIS and Remote Sensing.

On the other hand, the use of statistical models in producing susceptibility maps has seen many improvements that it can serve as a basis for hazard-risk management (Ayalew et al., 2004). Susceptibility is the likelihood of occurrence of a phenomenon under given terrain condition regardless of the time scale within which the particular phenomena is likely to occur (Soeters et.al, 1996). Susceptibility mapping does not consider the role of triggering factors like Rainfall and Seismic records for two reasons. One is because of inflexibility of the method to properly incorporate them in the method. And the second one is the usually uniform nature of these parameters through out area of study. Malczewski (1999) recognizes four criteria to consider in selecting factors contributing to any physical phenomena in a GIS analysis: operational, uniformity, measurability, and non-redundancy. Hence by using one of the many possible methods and incorporating the appropriate contributory factors, it is possible to come up with reliable tool for producing susceptibility maps.

1.2.2 Problem Statement

Qualitative or subjective methods that involve simple overlaying operations either in GIS platform or classical manual map overlaying are the prevailing techniques used in most institutes of the country. But although regionally their guidance task is important, these subjective methods are no more sufficient to solve specific and Multicriterial mechanism of occurrence of the successively prevailing conditions of landslide hazards. This is mainly due to the increasing population density that has led to inhabiting more and more of new areas that are vulnerable to landslide and other environmental hazards. And this needs a reliable and well-established land use planning and management task. This calls first and for most for, the generation of reliable, good quality, and multi-criteria based assessment of environmental risk of the target areas. Of these, landslide susceptibility assessment is an important part, especially in the prone areas. The use of GIS comes in when we want to integrate the otherwise highly labor, finance, and time intensive task of data acquisition, processing and analysis.

1.2.3 Objective

1.2.3.1 General Objective

In this thesis work the slope instability problem of the selected area (study area) has been assessed using Remote sensing technique and GIS, and a methodology depicting the whole scenario would be applied. This would add to the few studies that have been done in the field of slope instability problem in a non-site specific situation which usually employ deterministic models in the frequently landslide affected area of Debresina and the surrounding as well as else where in the world.

1.2.3.2 Specific objectives

- Filter-out the most relevant contributing factors towards the slope instability process in the area.
- Evaluate the relative importance of the various factors
- Produce a ranked spatial distribution of predicted hazardous landslide occurrences using the neural nets and Logistic regression models.
- Compare the performance of the Logistic regression and Artificial Neural Networks analysis outputs.

1.3. Scope of the study

Assessment of the geo-environmental factors for predicting the occurrence of an event involves obtaining first and foremost a detailed inventory, accurate data/information on the geographic distribution and variation of the factors. In addition the prediction does not bear any direct relation to changes in time series of the factors generally recognized to initiate landslide, a most common draw back of the predictive model.

The above mentioned constraints have a great influence on the present study. As the area investigation has not been investigated before from geological point of view, data pertaining to the study are entirely acquired based on remote sensing data interpretation which are prone to errors, requiring either a very long and tedious interpretation coupled with repeated ground truthing. To overcome this difficulty the area of investigation has been kept to the minimum while still keeping the area to cover as most landslide inventories as possible. But the problem of time series analysis remains unaddressed for

1.4. Availability of data

Secondary data for the thesis-work is obtained from various government agencies like Ministry of Water Resources (MoW), Ethiopian Geological Survey (EGS), Addis Ababa University (AAU), Ethiopian Road Authority (ERA), Ethiopian Meteorology Agency (EMA), Ethiopian Map Making Agency (EMMA), Regional offices as well as Published and Unpublished papers form various professionals. Nevertheless, primary data pertaining to the theme of the study in the area come from the field activities that were undertaken for six days.

1.5. Literature Review

Slope failure in the Ethiopian plateau evolution is an important process. Apart from hazardous consideration of the slope failure or mainly landslide, Ayalew et. al. (2004), Woldearegay et.al. (2004) discuss the importance of landslides and rock falls as one of the agents that shape and are still shaping the northwestern Ethiopian plateau. The steep scarps and colluvial deposits at the foot of the escarpments as part of a long-lived landslide characterized by massive landmass movement, tension cracks, scree falls and mud flows have been considered contributory factors for the development of the blue Nile gorge as a whole (Ayalew, 2004; EIGS, 1998). The Blue-Nile basin is characterized by deep, narrow, sharp and relatively younger, thick sedimentary and volcanic terrain.

The study area lacks reliable documented geologic data being at very large scale when available. The stratigraphy of the area on a regional scale is comprised of: **the lower Mesozoic Sandstone** at the bottom, through the **Mesozoic Limestone and Gypsum** (which are exposed only in the western and mid western part of the western plateau where the river Abay (Blue Nile) has deeply dissected the terrain) to the **upper most (youngest) columnar jointed tertiary basalts.** Mass wasting process is not restricted to the Blue Nile basin in Ethiopia. Instead it is widely distributed along the north western shoulder of the main Ethiopian rift. To mention some Ankober (completely destroyed in the late 19th century, EIGS, 1979), Dessie, the highlands surrounding Ambassel and Woldia. Else where in the rest of the highland mass like the Lalibela-Sekota road, the Semien highlands, the surroundings of Gondar and Angereb, western Tigray, Adigrat-Adwa road, Gelemso and Harar regions, Sawla locality, Bonga area, Jima and Gilgel Gibe area, surroundings of Finchewa, and the Debre Libanos and Mugher localities (Ayalew, Manuscript) similar

problems are faced. The tectonic activity in the region has been considered as one of the preparatory factors of most landslides while anomalous intensity of rainfall and some earthquake tremors (EIGS, 1998; Ayalew, 1999) has been cited as triggering mechanism of most landslides in the Ethiopian highlands. Similar study of the occurrence of landslide activities has been conducted in different parts of the country by EGS (Bonga, Dessie, Sawla). Most of the landslides show reactivation of pre-existing mass movements under various triggering conditions (e.g. EIGS, 1996, 1998).

Engineering geophysical investigation in the Blue Nile area has been undertaken (EIGS, 1974) for the purpose of rerouting the unstable alignment of the main road to Bahirdar. It indicated the presence of basaltic bedrock with varying degree of weathering underlying the clayey soil material intermixed with rock fragments. Faults and weak zones were also identified in the bedrock endangering the proposed route. An investigation of integrated master plan in the blue nile basin has extensively gathered refraction seismic, VES, and Resistivity data in the Gilgel Abay, Ribb, North Chagni, Fetam and Guder area (EIGS 1994).

A team of geologists from Addis Ababa University has investigated a landslide event in the south central part of the country, Wondogenet area, where from their investigations the factors: shallow rooted plants, structural discontinuities, drainage distributions and geologic setting have been identified as the key elements governing the occurrence of the landslides (Temesgen, B., et al, 2002). Landslide process is not unique to the Ethiopian plateau. Instead, it is a commonly encountered phenomena elsewhere in countries of similar topographic disposition as that of the Blue Nile basin. For example studies concerning landslide hazard in Japan where it is a common problem (G. Zhou 2003 M. Xie 2003 K.T. Chang L. Ayalew 2005) have been conducted.

Coming to the method of analysis a number of methodologies and analytical procedures have been well developed which can be adopted for assessment in similar terrain conditions. An extensive review of landslide process and the so far developed analytical approaches of their studies is found in "Landslides: Investigations and Mitigations" (Turner et al, 1996), a special report based on the contribution from some of the selected authors of the subject matter from all over the world. The use of GIS and Remote Sensing in the assessment of risk evaluation (susceptibility mapping) as well as understanding of the slope process is explicitly discussed. Instead of classical simple classification of the mass wasting activities based on only the type

of material involved and type of motion exhibited, there is now a generally accepted classification scheme of mass movement. The scheme incorporates state of activity (Active, Suspended, Dormant, Reactivated, Abandoned) and style of activity (Complex, composite, Multiple, Successive). A unified Landslide Classification System (ULCS) has been suggested by Weiczorek (1997) where the age, type of material, and type of movement are used to identify a particular slope failure type. Landslide occurrences have been reported to be dominant in the lower slopes of the valley in most humid areas (Toyohico, Manuscript) following the slope profile, where a similar climatic condition prevails in Ethiopian highlands. For the study of landslides various models and analytical methods have been used and appraised. A brief review of the methodologies is given in L. Ayalew (Manuscript). According to the review, three basic types of approaches are recognized with each representing a group of individually independent methodologies. Of these, the widely used and predating of all is the semi-qualitative approach followed by the more popular quantitative approach. The later one includes deterministic, stochastic, and nowadays neural network and fuzzy classification systems.

The most utilized stochastic analytical methods include Discriminant analysis and Logistic regression. Both methods distinguish two sets of parameters: "Independent" variables or the so called priori considered causative factors, and the "dependent" variables which are out puts of the functions of the independent variables where they can assume two values either an occurrence (1) or absence (0) of an event. Nevertheless, the former methodology differs from the later as it requires a normal distribution among the input (independent) variables (Nuria. S., et al., 2003).

Variations of each type of methodologies exist. A more recent approach has been adopted in producing susceptibility mapping comprising the third approach. This has been termed as Hybrid index based and training based by combining appropriate methodologies categorized in the previous two types of methodologies. One of this is the Training based hybrid modeling method, which combines Bivariate Statistical Analysis (BSA) and Neural Networks (Lee et al., 2004).

The use of neural networks in physical sciences is new and becoming widely accepted which best depicts and/or models our world. It models the complex interaction of the environmental

systems interactions better than ever before since it allows modeling of complex systems without requiring the explicit formulation of the relationships that may exist between variables (Xiaogang, 2005).

Neural nets are composed of three layers: one layer of inputs that are in the raster format and presented as a single column which are connected to the output layer, which is the landslide inventory map, the hidden layers that are the connectors attaching weightings to the relationship layers and the out put layers (Jeff T, 2002).

There are variety of approaches of using neural networks: Competitive Learning, the Hopfield Network (Hopfield, 1982), the Kohonen network (Kohonen, 1988), and back propagation neural networks (Rumelhart, 1986) of which the back propagation neural networks method are the most popularly in used currently. According to Rumlehart (1986) the major components in this method are:

- a set of processing units,
- the state of activation of a processing unit,
- the function used to compute output of a processing unit,
- the pattern of connectivity among processing units,
- the rule of activation propagation,
- the activation function, and
- the rule of learning employed

In assessing the landslide hazard of south western Japan, Moxen Xie (2003) indicated the use of slope units for representing the various thematic maps in the analysis. He argues that slope units would have a more robust bearing on the natural conditions representation than a simple grid based representation which is on the other hand simpler in handling the spatial data distribution at the cost of illogical representation of the natural boundary of the physical parameters. He used the Archydro's "watershed" tool to generate basin boundary twice before and after inverting the DEM and unifying them later to obtain the full slope unit. Concerning the classification of feature classes fuzzy methods have been advocated in various research (Ercanoglu et al., Manuscript) which can be used to account for uncertainty and replace crisp classification method by providing class overlap that is more realistic for modeling continuous landscape patterns (Pece V. Gorsevski, et al., 2003). That means continuous factors parameters are classified into any of the classes based on a certain percentage of belongingness of the elements to any of the parameter classes instead of the usual either full or non membership to one of the parameter classes.

2. Study area overview

2.1. Study Area

The study area is bounded by the coordinates **9° 45' N to 10° 00' N and 39° 40' E to 39° 50' E** and is characterized by elevation ranges between **2800m** in the northern part to **1800m** in the southern section. It is located northeast of Addis Ababa city on the main road to Dessie, approximately 200 km from Addis Ababa. The accessibility of the area is limited being part of the highly dissected terrain of Ethiopian highlands (Showan Plateau). Access roads to the various parts of the study area are warn out gravel roads, when available and are possible to work on foot only (not workable for motor).

Generally, the study area is underlain by the **Tertiary Volcanic Rocks**, which have a thickness of **hundreds of meters** and are founded on top of the **Abbay basin sedimentary formations.** The prevailing climatic condition of the area is "**Dega" National meteorology Agency** (NMA) with monthly average **rainfall >1000mm/yr** and **temperature 10°c-15°c.** Scarce vegetation coverage of Eucalyptus in the high lands and Scrub in the low lying areas are dominant other than the extensively cultivated landmass in the plain areas. Two major drainage basin are located in the area: the Jemma river basin that is confluent with Abay river along with a number of its tributaries: Gado, Ayseram, Mofer wuha, Gedemsa, Lege yida, Dalecha and Beresa and the awash river basin in the south western part along with Zhema, Muter, Chewbele and Ayrara tributaries. The area has a bimodal rainfall distribution one starting from mid February to mid March, which is called Belg and the other from mid June to mid September which is called the Kiremt season. The population distribution of the area follows that of the topography: concentrating along the water divides mostly leading to location of almost all settlements and main roads along the divide of the highly dissected terrain.

2.1.2. Physiography, Climate and Vegetation

Physiographically the study area is located in the Showan highlands, the smallest highlands of the Ethiopian northwestern highlands, which also includes the Tigrean north central massif, South Western highlands of Gojam and Gondar. The Showan plateau is bounded by the Ethiopian rift on the eastern and south eastern sides while Abay gorge and Omo gorge border it on the north western and south western sides.

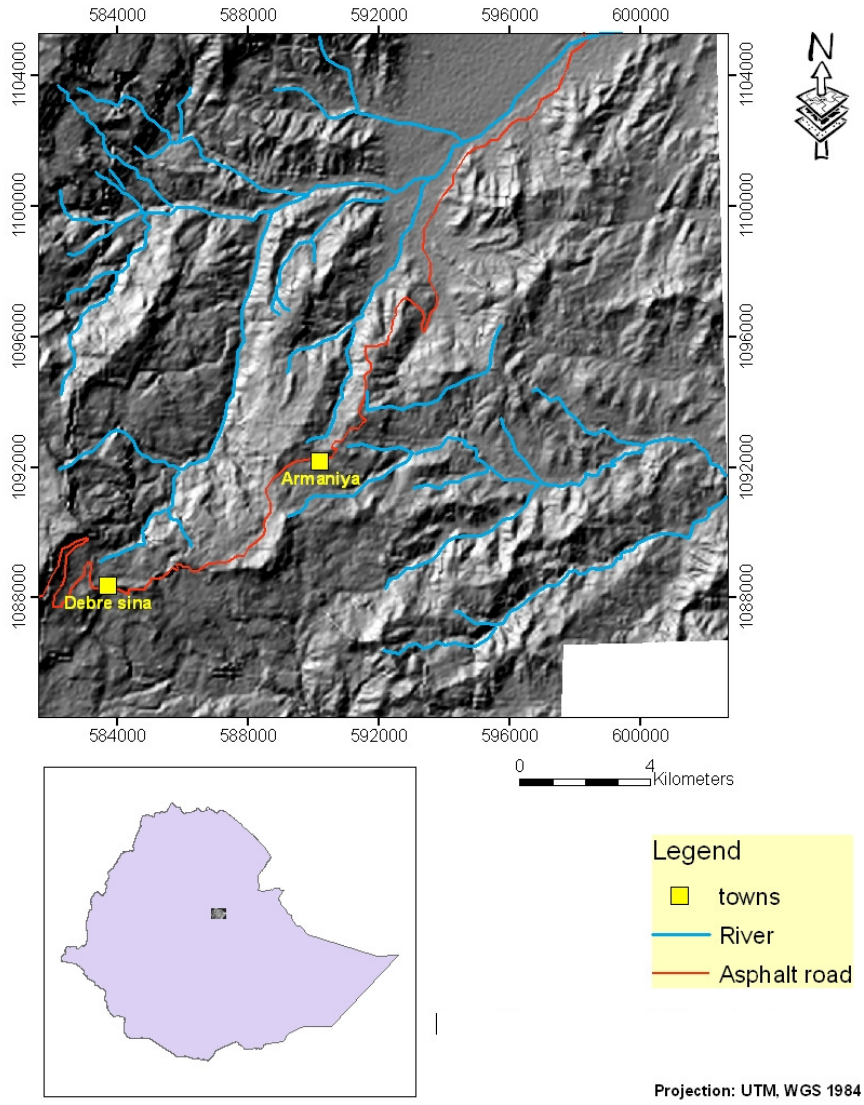


Figure 2.1Location map of the study area (Debresina).

Specifically the study area is located on the breeches dividing the Abay and Awash river basins. It is also a physiographically elevated landmass making the western steep boundaries of the Ethiopian rift. The extension of the North –south running graben known as Borkena graben in the high land characterizes it. In fact, it has been indicated that there is a triple junction of the Borkena graben, and the South west bearing faults of the rift valley around the vicinity of the study area.

The prevalent climatic condition of the area has been classified into warm, temperate by BCEOM (1999). The mean temperature of the coldest month is < 18 degree Celsius. The mean annual rainfall when expressed in relation to the mean annual temperature is:

RF (mm) > 20 * (T+14), where T stands for temperature

The region is characterized by Semi-bimodal rainfall pattern, i.e a dry season from October to January, separating the long and short rainy seasons of June to September and February to May respectively.

With the exception of a protected forest in the south western part of the study area, called "Wof Washa". The area is devoid of most of the natural vegetation and forest cover. This has been what used to be one of most vegetated places of the area from satellite images of the 1973 MSS. Currently shrub lands of less than 2m high, grass land and cultivated land dominate the land use/cover type.

Figure 2.2. The study area is located in a temperate like climate, with typical plant species like Lobelia

2.3. Socio-Economics condition of the study area

The study area is located in the part of the country characterized by high increase in population growth rate. This has led to a very tense economic activity through farming transforming the natural environment into a seriously stressed condition. This in turn has led to accelerated land degradation by erosion, various mass wasting processes and reduction in productivity. Nevertheless the rate of change of land degradation which can be caused by the land use land cover change may or may not match to the recurrence interval of the landslides and other mass wasting processes in the area.

Subsistence of the local community depends on rain fed agriculture and some irrigation scheme. Major crops produced are Barely, Wheat, Sorghum, Teff, Pea and Lentil. Also fire wood and forest products are marketed extensively contributing to the destabilization of the already critically standing steep slope of soil developments and weathered and jointed rocks.

3. Regional Geologic setting of the study area

The geologic setting of the central Ethiopian high lands is characterized by the voluminous tertiary volcanics capping the Mesozoic sediments which are exposed only in the deep incisions by the major rivers. Examples include Abay Gorge, Jemma gorge, Mugher gorge. The tertiary rocks are around 30Ma old. Nevertheless the volcanic outpouring took only around 2 million years (Hoffman et al., 1998; Pik et al., 1998, 1999). The continental flood basalt is the result of impingent of the Afar mantle plume beneath the Ethiopian lithosphere. The longer time of dwelling in some of the temperate like climate along with the quick weathering tendency of the formations which are mostly basic in composition (basalts ad few trachites) quick development of soil and weathering products have lead to potentially high slope instability conditions. Four major volcanic units have been identified in the region. These are the lower Ashange formation, overlain by Aiba basalts which underlies the Alajie formation. The last one is the lowest volcanic unit found in the vicinity of the area. The fourth unit is the latest of the flood basalts with tholeiitic composition. It is called Tarmaber formation which forms one of the prominent landmass shield volcanoes. The change in formation of the volcanic units between the underlying basic and over lying tholeitic basalts have been attributed to the tectonic setting and the decrease magma flux rather than pure compositional variation in the magma source.

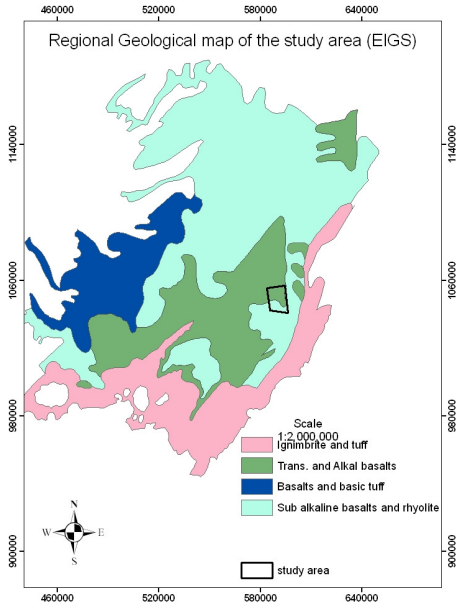


Figure 3. Regional Geologic map of the study area (EIGS, 1996).

4. States of Land sliding processes

4.1. Landslide condition in Ethiopia

Few sites of landslide have been investigated in Ethiopia previously. Their distribution is generally haphazard whereby they have been observed to occur in the three physiographic units of the country. In the north western highlands; Desssie town (Tenalem, 2005; Ayalew 2000; EIGS, 1998); between the towns of Gohatsion and DEjen in Abay gorge (Almaz, 1998;); Debresina area (DPPPRRC *[2], 1972); Butajira area (recent personal acquaintance). In the Ethiopian rift valley areas like "Wondogenet" (Temesgen et, al., 2000) have been observed. The south eastern highlands also experience similar cases of landslide hazard, "Bonga" area (IEGS, 1998). All of the landslide events have been observed to occur following a heavy rainfall along with a number of pre-conditioning attributes. These attributes include proximity to drainage, irrigation channel, infrastructure development (road), which usually have the effect of undercutting, Geologic formations property, and soil condition.

Among the prevalent landslide occurrence areas some have not gained enough attention from investigators despite the great advance that is taking place both on the methodology and the understanding of landslide mechanisms. Nevertheless the problem has persisted affecting peoples' life. One of such cases is the Debresina area (study area). Except one preliminary report on the occurrence of landslide in the area by experts from EIGS, AAU (Addis Ababa University) and DPPPRRC in 1979, there are no relevant primary data on the area at appropriate scale (like fore example geological map of at scales less than 1:2,000,000 scale) covering the study area. Hence in this study most of the data were acquired from satellite image analysis using various enhancement techniques (chapter 7). The image used is Landsat etm+ of May 2000.

4.2. Landslides history of the study Area

Landslides in the study area (Debresina) started with various signs of slope failure some years back. They usually start with fissures in the soil. Landslides have been reported to have occured as early as the Italian invasion (more than sixty years ago) times (eyewitnesses) in Tach Indode and Tid Amba localities. Starting from around 1985, fissures were again

* DPPPRRC: Disaster preparedness Planning Preogramme Rehabilitation and Relief Commision.

observed at Disk Amba and Tid amba, specifically at Ketanit and Bado mesk(EIGS, 1980). Again in 1969, a landslide occurred at the same place and in 1979, additional slope failures followed. From time to time the fissures in the soil widen up developing to slides throughout the year 1979. Other landslides in the area predating to the Italian invasion period have occurred in Majete, Yenat Metoria, Work Amba. Especially the events at Nib Amba and Work Amba have been associated with Seismic events.

instability problems along the road corridor.

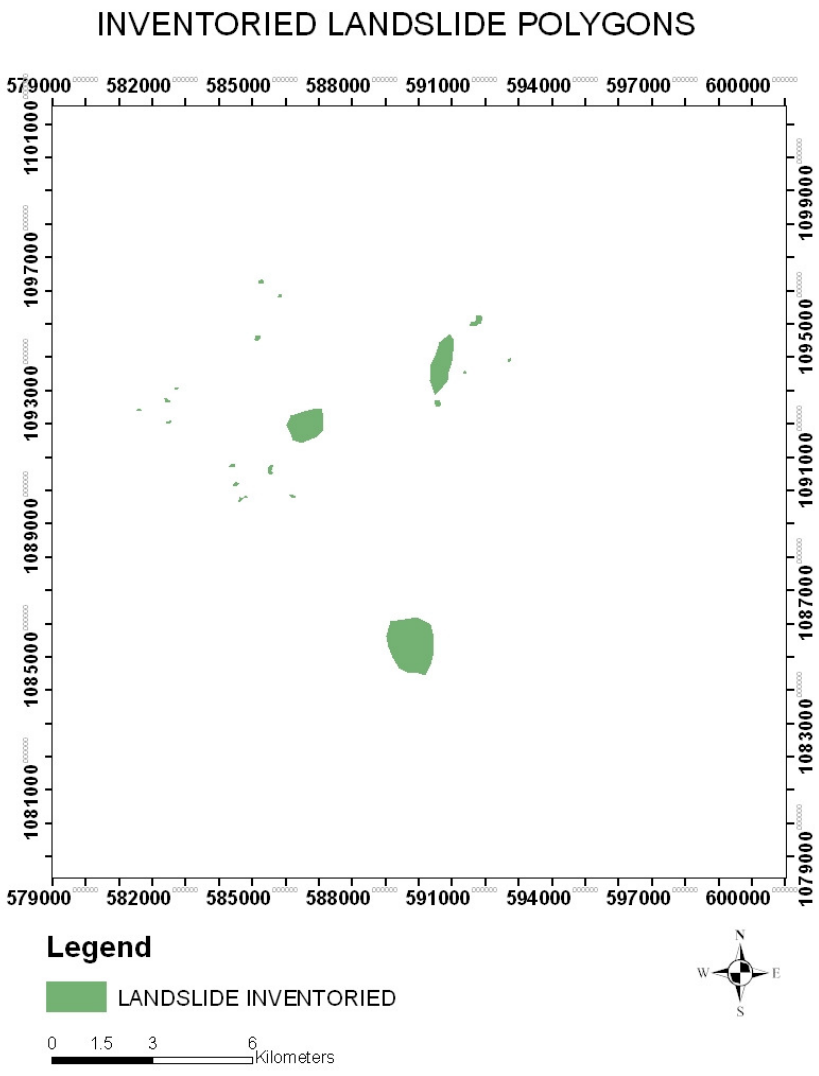


Figure 4.1. Inventoried landslide

Most of the landslides occur at or near 2000 meters above sea level. Most of the streams in the area have two prevalent patterns of NE and oriented head waters and NW oriented major rivers indicating a tectonic control on their course. These stream are perennial owing to the long lasting soaking effect of the surrounding soil formation long after the rainy season. Deforestation for housing, agricultural activities, search for building material and road construction have generally assumed to be the preconditioning agents of slope failure in the area.

A recent mighty landslide has occurred in the area along the Dem Aytmash River. It covers area of ten to eleven square kilometers. This huge slide is still progressing head ward. In addition contractors of the main road passing through the area have reported several slope instability problems along the road corridor.

a.

b.

c.

d.

e.

Figure 4.2 **a** to **e**. Photographs showing the landslide features of the Debresina area

5. Methodology of Study

5.1. Methodology

a. Terminologies:

Factors: The various environmental parameters or themes considered relevant in the slope stability process. e.g., Lithology, upslope area's contribution to flow and Topographic attributes such as planar curvature, profile curvature, slope and aspect.

Classes: The various classifications of the factor maps for implicitly assessing the type of relation between slope instability and the factors and to create discrete representation of continuous data like elevation. Example: the Elevation classified into several intervals.

Weighting: The priority value given to the various factor maps according to their relevance in contribution to the slope instability problem (From Bivariate statistical Analysis, BSA).

Ratings: The priority value given to the varying classes of the factor maps. (From the Logistic regression coefficients and Neural networks classification weights).

b. Tasks

1) Using Remote Sensing data for obtaining input data for GIS

- Prepare landslide inventory map from Aerial photo interpretation of approximately 1:50,000 scale, reported landslide events; field survey of the study area.
- Obtain DEM of the area by vectorizing contours of the 1:50,000 scale topographic map.
- Processing the DEM to generate topographic parameters: Slope, Planar curvature, Profile curvature, Flow accumulation and Aspect.
- Image analysis for interpreting the lithologic map of the study area
- Set the projection of the data sets to a common system.

- Rasterizing and converting the factor maps as well as Landslide inventory maps to ASCII format for GIS analysis.

2) Employing the strategy of using grid cells as units of mapping.
3) Rating the various classes based on BSA method where the number of classes would be decided by the number of classes in the discrete data with the largest number of classes (Lee et. al., 2004, Ayalew et al., 2004). The BSA uses the landslide density of each class (rating irrespective of the relative importance of the various factors to which the classes particularly belong to) to generate weighted ratings (incorporating the relative importance of the factors the classes belong to) for the classes of the factors through pair wise comparison matrix.
4) Applying Logistic regression (LogRec) model to generate the relationship weights between the occurrences of landslide because of the various factors. The merit behind logistic regression rests on the analysis of a problem, in which a result measured with dichotomous variables (dependent) such as 0 and 1 or true and false, is determined from one or more independent factors. Logistic regression involves fitting the dependent variable and independent parameters using an equation of the form, $Y = C_0 + C_1X_1 + C_2X_2 + \Lambda + C_nX_n$ where C_0 is the intercept and C_1, C_2, . . ., Cn are coefficients which, when used as a power to the natural logarithm (e) or in other words as logit transform, measure the contribution of independent parameters (X_1, X_2, . . ., Xn) to the variations in dependent variable (Y).
5) Obtain the weights generated by Artificial neural network (ANN) to classify individual pixels as slide and non-slide based on the value of the attributes which are the seven causative factors
6) Interpret the weights hence obtained, assess, and confirm their accordance with the reality.
7) Produce two Susceptibility maps based on the Weights (ANN and LogReg) and the ratings of the parameters by raster calculation.
8) Validation of the susceptibility map with observed landslide locations.

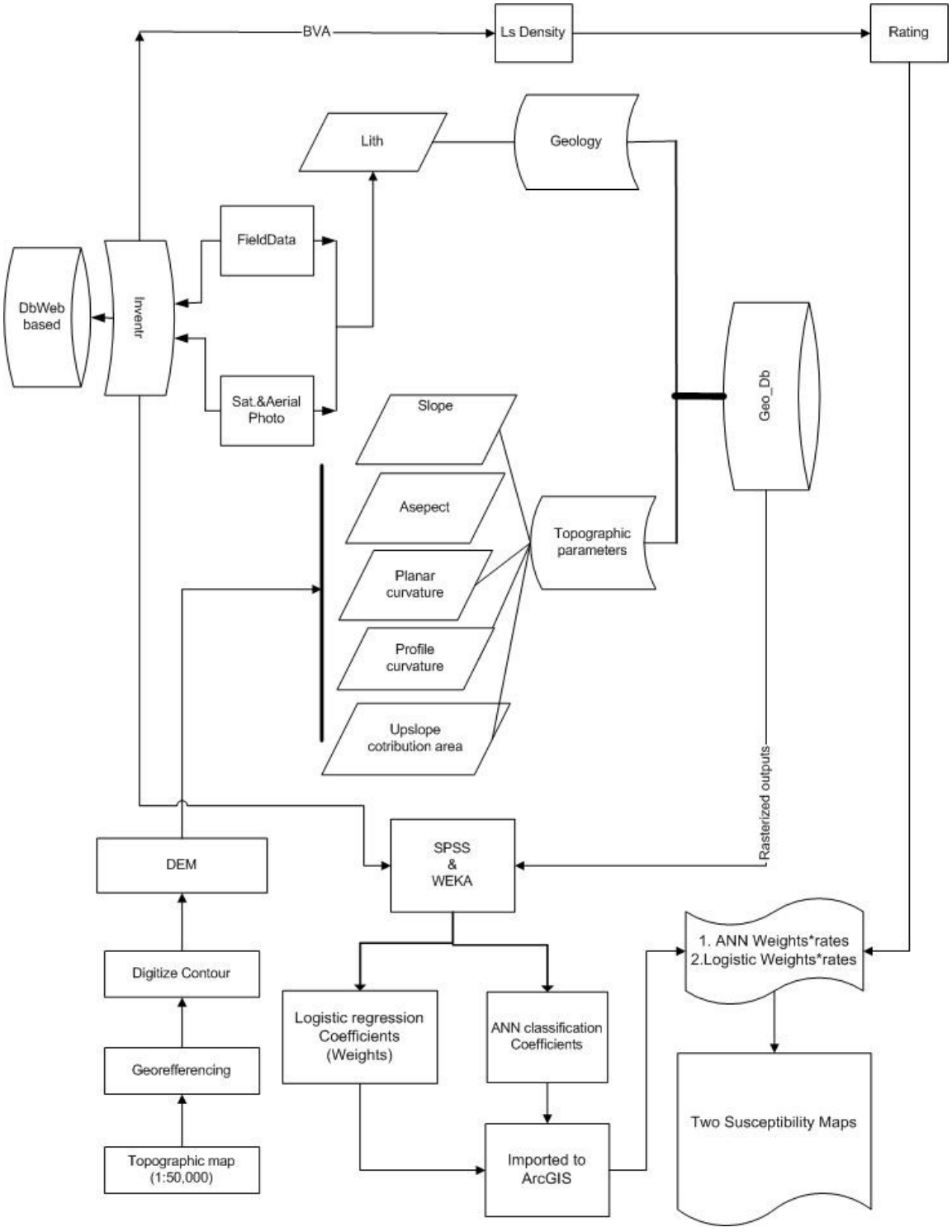


Figure 5. Flow-chart of study methodology.

6. Discussion on the application of the analytical methods in Geologic data analysis.

6.1 Application of the Logistic regression analysis

Logistic regression is a statistical method for classifying elements into one of two populations. In this case, the term elements could more appropriately be replaced by unit of analysis (like fore example a cell or a pixel) and the populations by occurrence/non-occurrence case of landslide. It is a type of multivariate analysis method. This means that the variance of the sets of data of certain populations which are considered as the independent variable are compared and correlated with variation in the data of the dependent variable. Again in this case the dependent variable represents the occurrence/non- occurrence of landslide cases while the independent variables represents the causative factors. In general, stronger relationship between dependent and independent variable in the logistic regression exists when the variation in the independent variable corresponds to matching variation among the dependent variable.

Basically Logistic regression can be used whenever a certain unit of land (a pixel) is to be classified into one of two cases/states, occurrence or non-occurrence of an event. When more than two states are to be used to classify the subjects, what is called polychotoumous or generalized regression analysis can be used. This will be beyond the scope of the present study.

The past applications of the logistic regression have mostly been in areas of medicine. It has been used fore example, to calculate the risk of developing heart disease based on certain personal and behavioral characteristic such as age, weight, blood pressure, cholesterol level and smoking history. Also in social psychology it has been used to predict teenage pregnancy based on sociologic data of family size, religion, and grade point average. In industry operation units have been classified as successful or not based on some objective criteria.

In recent years, the application of logistic regression has been tested in the geosciences fields L. Ayalew (L. Ayalew, et.al, 2005) has used similar methodology to predict the occurrence of landslide based on information on the geology, and topographic parameters and inventoried past landslides of Sado Island in Japan. Similiarly Lee, (Lee, S., et al.,

2004) and L. Ermini (L.Ermini et.al, 2005) have also used the same methodology to predict landslide occurrence in North Korea and Italy. The variations are only the factors taken into consideration which vary depending on the target areas setting.

One of the advantages of logistic regression is that it can be used for classifying elements regardless of the normality of the distribution of the pertinent data. This is a very important use given the usually non-normal distribution of most geologic data. In addition integration of discrete and continuous data can be used for analysis. But the knowledge of the dependent and independent variables of the sample being analyzed is very important where by a scenario with only the predictor variable can be used to generate new dependent variables occurrence.

The logistic regression and discriminant functions are closely related and are similar statistical analytical methods. In discriminant analysis the function has the form

$$Z = a_1X_1 + a_2X2 + \ldots + a_pXp$$

Where the co-efficients $a_1,\ldots,a_p$ represent the degree to which the variables X_n contribute in classifying the object into group Z (occurrence or non- occurrence). Some computations must be done to obtain the values of the discriminant coefficients. Some programs are available that can print out what is called a "classification function" for each group. For each population the coefficients are printed for each variable. The discriminant function coefficients $a_1,\ldots,a_p$ are then obtained by subtraction.

Based on the values of the various attributes (factors) the individual (objects) are classified into either of the two cases of landslide occurrence or non-occurrence. But after performing the discriminant analysis the possibility of making the wrong classification is there and this can be checked with the posterior probability. It is the probability computed for an object whether or not it came from either of the two cases. The formula is:

$$P_z = 1/(1+e^{c-z})$$

This function has a logistic form. It can be used to derive the multiple logistic regression. This implies that multiple regression can be obtained from discriminant function.

This same function can be transformed to produce new interpretations with the following definitions of odds as:

$$\text{Odds} = P_z/1-P_z$$

Odds, represents the ratio of probability of occurrence of an event (a landslide) to the probability of absence of an event (non-landslide occurrence).

The above function has symmetrical problem if treated with linear sense. Hence remedy can be taken by considering Logarithmic scale. For example the variation of P_z (probability) from 0 to 1 corresponds to variation of the odds from 0 to infinity. The asymmetry is that while the odds ratio vary from 0 to 1 for P_z values from 0 to 0.5, the odds value differs to be from 1 to infinity for same interval of difference for P_z, i.e.
0.5 to1. Up on considering the logarithmic scale (transformation), the Logit can be defined as ln(odds) where $P_z = 0$ corresponds to ln (odds) = -(infinity), $P_z = 0.5$ corresponds to ln(odds) 0.0 and $P_z = 1$ corresponds to ln(odds) = +(infinity).
Hence the previous equation is transformed to ln $(P_z/1-P_z) = -C + Z$ After some algebraic manipulation. The bottom line is that the logits of the odds is linear in the discriminant function Z. This implies that

$$\ln(\text{odds}) = a_1 + b_1X_1 + b_2X_2 + \ldots + b_pX_p$$

which is of similar form to that of the linear regression except that this one has been transformed to be linear from the original (previous) non-linear equation form by using logarithmic scale. Hence the equation is named multiple logistic regression equation, and the coefficients in the equation can be interpreted as regression coefficients.
The fundamental assumption in logistic regression analysis is that ln(odds) is linearly related to the independent variables. No assumptions are made regarding the distributions of the x-variables. In terms of the probability of belonging to either of the cases population can be rewritten as:

$$\text{Prob. of belonging to case I} = 1/\,1 + \exp(+ b_1X_1 + b_2X_2 + \ldots + b_pX_p)$$

which is called the logistic regression equation. It is equivalent to the posterior probability of the discriminant function. The coefficients a_1 up to a_p can be estimated by linear discriminant analysis or maximum likelihood method, the previous is preferred for normal distribution of variables with smaller sample size otherwise the later can be used for any of the variable distribution. The procedure is generally an iterative one.

6.1.1. Variable Selection

In case of exploratory situations the variable selection, for logistic regression analysis, is crucial for the sake of computational simplicity. Screening out of predictor variables can be done by using the common chi-square test of association for discrete predictor variables and test of equal group means for continuous predictor variables.

A fairly large significant level is usually chosen for screening so that useful predictors will not be missed (alpha = 0.15) are kept for further analysis.

6.1.2. Checking the fit of the model

Several approaches have been proposed for testing how well the fitted logistic regression model fits the data. Most approaches rely on the idea of comparing an observed number of individuals with the number expected if the fitted model were valid. These observed and expected numbers are combined to form a chi-square statistic called the goodness of fit. Large values of the test statistics, P, indicate a poor fit of the model (equivalently a low P value indicates a good fit).

A number of approaches are in practice for implementing the statistic. One of this is the one developed by Lemeshow and Hosmer (1982). In this approach the probability of belonging to population I (say a landslide occurrence case) is calculated for every individual in the sample, and the resulting numbers are arranged in increasing order. The range of probability value is then divided into ten groups (deciles). For each deciles the observed number of individuals in population I (O) is computed. Also the expected number (E) is calculated by adding the logistic probabilities for all individuals in each decile. Then the goodness of fit statistics is calculated from the pearson chi-square statistics as goodness of fit;

$$\lambda^2 = \Sigma\ [(O=E)/E]$$

where the summation extends over the two populations and the ten deciles. The goodness of fit can also be used to check the linearity of the logits with respect to independent variables.

6.2. Artificial neural networks

Artificial neural networks are computational models of the biological brain. There are many types of neural networks resembling the biological nervous system operation with varying degree of sophistication.

Types of Neural Networks

Neural networks generally consist of a number of interconnected processing elements widely known as neurons. The two important elements of neural networks are the various types of neural interconnections arrangements available and the types of algorithms used to set the strength of the connections.

The structural arrangement means that out puts of one node (neuron) are fed to another neuron in the succeeding layer or to the same neuron. Such arrangements which may or may not allow the outputs of one node to be feedback to other nodes in the preceeding layer or to the same node. Hence two types of structural arrangement classifications can be recognized: Feed-forward networks (unidirectional flow of signals /outputs of each node) and recurrent networks (bidirectional /back and forth/ flow of signals). Examples of the first type include Multi-layer perceptron (MLP), the Learning vector quantization (LVQ) are well known. The later types of networks include Hopefield networks, Elman networks, and Jordan networks.

In addition to the structural arrangements the various types of algorithms distinguish a number of neural networks. These algorithms are used to determine the strength (weights) of inter-node connections. The main types are Supervised learning and Unsupervised learning algorithms. A multitude of both types of algorithms have been developed. Delta rule, Generalized delta rule or back propagation algorithm and LVQ algorithm (Pham, D.T., 1997) correspond to few of the supervised algorithms.

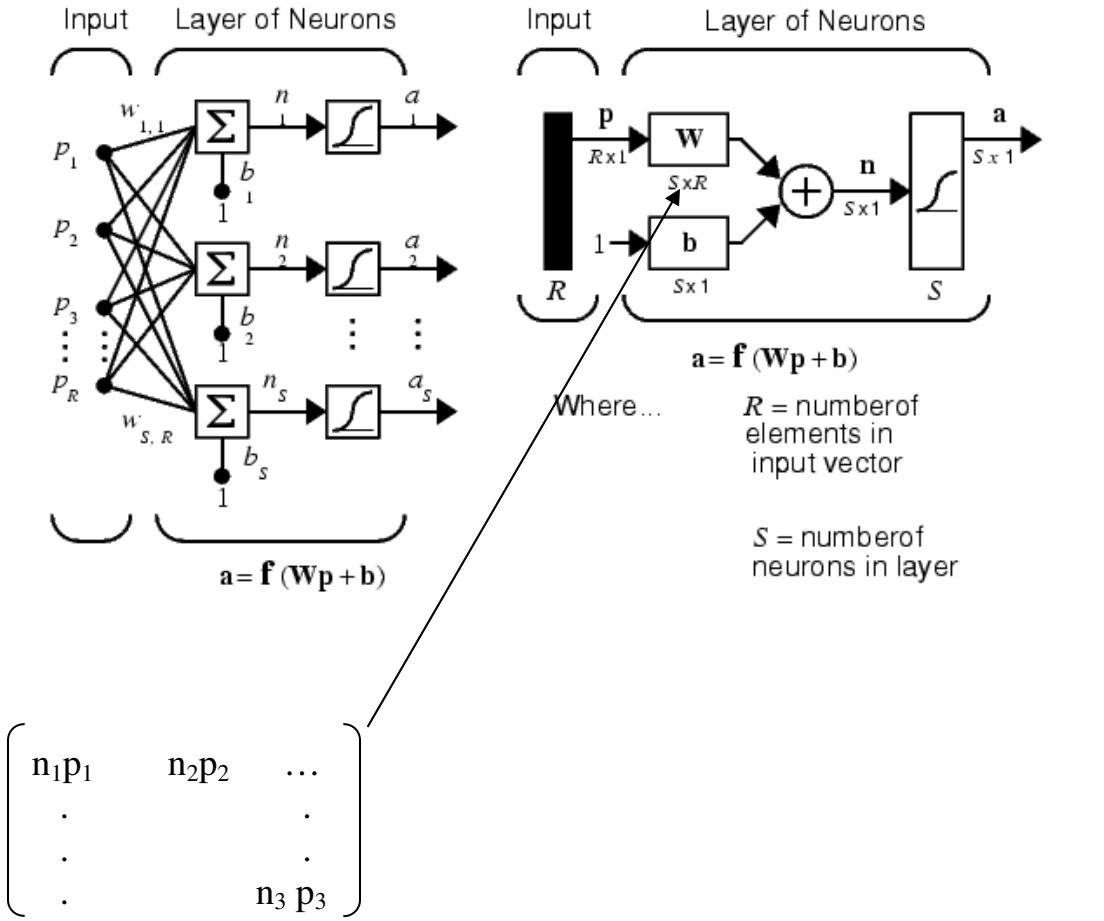
On the other hand Kohonen and Carpenter Grossberg Adaptive Resonance Theory (ART) (Carpenter Grossberg, 1988) are some of the varieties of unsupervised algorithms in use.

The principle behind supervised learning is adjustment of the strength (weight) of the interneuron connections according to the difference between desired (from inventory data) and the actual network outputs. Such desired outputs are not required in unsupervised learning rules. During training only input patterns (factors) are presented to

the neural networks which automatically adapts the weights of its connections to cluster the input patterns into groups with similar features.

For the purpose of this study the multi layer perceptron (MLP) are employed. Discussion on further details of the type of MLP used follows.

6.2.1. Multi Layer Perceptron



$$\begin{pmatrix} n_1p_1 & n_2p_2 & \dots \\ . & & . \\ . & & . \\ . & & n_3\ p_3 \end{pmatrix}$$

$\mathbf{n_1}$ up to $\mathbf{n_s}$ are the neurons while P_1 up to P_R are the R input vectors.

Figure 6.1. A Multi-layer Perceptron structure

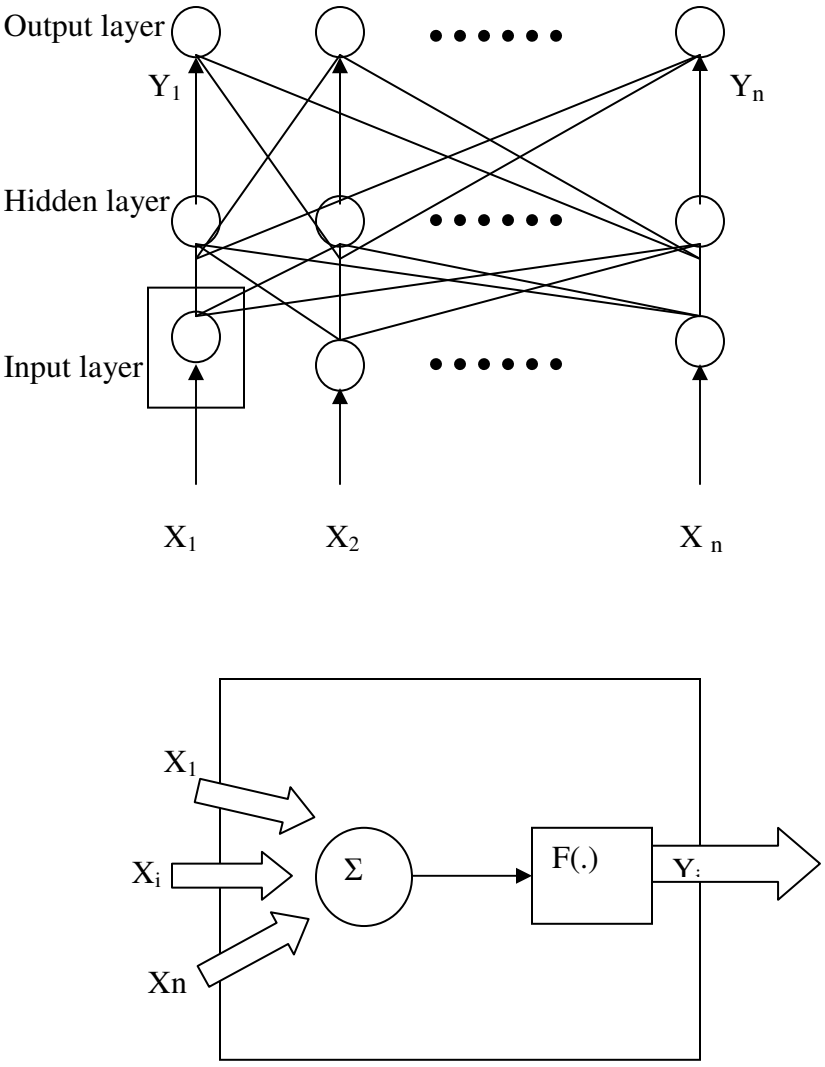


Figure 6.2. A close-up look of the structure of a neural node. Xi represent the input weight matrix, resulting in an output weight through the activation function F.

With reference to the above figure, three layers are observed namely input, intermediate (hidden), and out put layer each made up of neurons. The neurons are only connected between layers not with in a layer. The purpose of the input layer is only to buffer between raw data input and the hidden (intermediate layer) for distributing the input signals X_i to neurons in the hidden layer. All the Neurons, j, in the hidden layer sum up their input signals X_i from all the input layer neurons after weighting them with the strengths of the respective connections W_{ji} from the input layer and computes its output Y_j as a function of the sum:

$$Y_j = f\,[\Sigma W_{ji} X_i]$$

There are various types of functions for computing the output Y_j depending on the preference of the subject being addressed. These are summarized in the following table. The functions are more commonly called activation functions

Type of functions	**Functions**
Linear	$F(s) = s$
Threshold	$F(s) = +$, if $s > s_t$ otherwise -
Sigmoid	$F(s) = 1/(1+e^{-s})$
Hyperbolic tangent	$F(s) = 1- e^{(-s)} / (1+e^{(2s)}$
Radial basis function	$F(s) = e^{(-s/B)(s/B)}$

Table 6.1. Most common activation functions of neural networks architecture.

The output layer nodes are in turn similar in their function as the previously discussed neurons in the middle layer.

ANN implementation proceeds with two epochs system, one for training and the other for generalization. Back propagation (BP) algorithm which is a gradient decent algorithm is most commonly applied for MLP training. Gradient decent means upon computing the difference between the desired (network output) and the actual (inventory of real data, landslide) outputs of the model and propagate the signal backwards to the neurons either to increase or decrease the new weights depending on the nature of the difference (i.e. positive difference leads to decreasing of the weights and negative difference leads to increasing of the weights) to certain threshold level.

There are some parameters with which such control mechanisms are seeded into the ANN model. These are represented in the formula:

$$\Delta W_{ji} = \eta \, \delta_j \, X_i$$

Where η is called the learning rate, δ_j is a factor depending on whether the j^{th} neuron is an output layer neuron or a hidden layer neuron.

- For an output layer neuron the differential equation becomes

$$\delta_j = [\sigma f / \sigma net_j] (Y_j^{(t)} - Y_j),$$

where net_j is the total weighted sum of input signals to neuron j and $Y_j^{(t)}$ is target output for neuron j.

For the hidden layer neuron the previous differential equation is modified as

$$\delta_j = [\sigma f / \sigma net_j] (\Sigma W_{qj} \delta_q)$$

In this formula there are target outputs, the difference between the target and the actual output of the hidden layer neuron , j, is replaced by the weighted sum of the δ_q term already obtained for neurons q connected to the output of j neuron layers.

Thus, iteratively beginning with output layer, the δ term is computed for neurons in all layers and weight updates determined for all connections. The weight updating process can also take place after the presentation of the whole set of training patterns (batch training). In either case, a training epoch is sad to have been completed when all training patterns have been presented once to the MLP.

For all but the most trivial problem, several epochs are required for the MLP to be properly trained. A commonly adopted method to speed up the training is to add a "momentum" term to the first back-propagation equation which effectively lets the previous weight changes influence the new weight outputs (modifications).

$$\Delta \underline{W_{ji}\ (k+1)} = \eta\ \delta_j\ X_i + \mu\ \Delta \underline{W_{ji}(k)}$$

Where the underlined terms are the weight changes in epochs k+1 and K respectively, η is the learning rate and μ is the momentum coefficient. Setting the learning rate can be tricky as higher magnitude would result in oscillation state of the weight output along a pivot (because of large increase or decrease in the modification of the outputs upon back propagating signal without reaching the threshold value). On the other hand setting a very low value would lead to extremely longer computational time for the PC to handle. The usually recommended interval is between 0 and 1.

7. Procedures of the remote sensing and GIS analysis implementation

As per the flow chart in the methodology section (5.1.2), the analysis of landslide susceptibility started with acquisition of the most basic data, satellite image. Enhancement of ETM+ image was done and used for interpretation.

In addition field inventory of actual landslide sites was undertaken during the field work from December 27 to January 4, 2006. During the field work ground truthing of the satellite images and confirmation of the types of lithologic units identified during satellite image interpretations was done. Finally, Image classification was done to enhance the delineation of the lithologic units.

For generation of the DEM of the area the 1:50,000 scale topographic map was used to obtain the contour map by capturing the scanned raster of the topographic map and saving the contours as a vector using the R2V (Raster to Vector) software.

7.1. Generation of Digital Elevation Model (DEM) of the area

The 1:50,000 scale topographic map of "Debresina" and part of "Sela Dingay" sub sheets were first obtained as positive print out on a transparent film and scanned. The raster is then imported to R2V software (a specialized software for automatic vectorization of raster data sets). The contours were then vectorized and imported to ArcGIS 9.0 and edited for topology. Following this, the DEM at 20m cell resolution of the area was created from which the Aspect, Slope, Elevation, Horizontal and Profile curvature, and the upslope pixel flow accumulation layer was created for incorporation into analysis (Figure 7.2a - g).

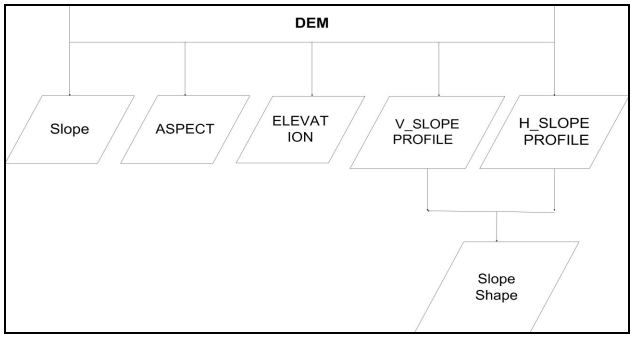


Figure 7.1 Parameters obtained from DEM.

7.2. Image Analysis

Image analysis has been done for geologic map interpretation. LANDSAT, ETM+ images have been processed with various enhancement techniques in order to get a good contrast between features. The image is that acquired on May, 2000. The analysis was carried out following the scheme of first enhancing the original image followed by supervised image classification. Finally the classified image has been evaluated and units from both the classified image and visual interpretation were incorporated to prepare the final lithologic map. During visual interpretation a high DN value in band 5 and band 3, and a low DN value in band 4 was utilized to visualize the image for lithologic discrimination. Various band ratios (band ratios 5/1, 5/4, 3/4) (Figure 7.3) were also performed subsequently to eliminate the effect of vegetation and slope interference with reflectance from exposed rocks or bare lands.

7.2.1. Image enhancement

The originally acquired image had relatively low contrast between features. In order to get clear display of features for visual interpretation Gaussian stretch has been applied on the image after trying all other stretching types following ENVI 4.2 routines. Specially, the low bands (band 1-3) were originally very much compressed. The original and contrast stretched images histogram are shown in Table 7.1 and Table 7.2.

Basic Stats	Min	Max	Mean	Stdev	Num	Eigenvalue
Band 1	53	181	90.914792	8.822364	1	1492.887131
Band 2	36	174	84.371737	11.083140	2	189.101591
Band 3	30	210	102.874232	18.278905	3	83.384698
Band 4	18	146	71.592874	12.067463	4	12.218300
Band 5	16	255	114.090678	26.268464	5	9.947011
Band 7	11	255	87.359830	20.492127	6	2.832887

Table 7.1 Table showing the original band statistics

Basic Stats	Min	Max	Mean	Stdev	Num	Eigenvalue
Band 1	0	255.000000	126.022918	56.502757	1	14722.654347
Band 2	0	255.000000	126.936421	56.422206	2	2862.510929
Band 3	0	255.000000	127.529670	56.382931	3	1179.595242
Band 4	0	255.000000	126.967149	56.645842	4	222.956173
Band 5	0	255.000000	127.815036	56.455773	5	84.099401
Band 6	0	255.000000	127.432706	56.485235	6	69.833273

Correlation	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6
Band 1	1.000000	0.942727	0.906649	0.373197	0.573404	0.690439
Band 2	0.942727	1.000000	0.964723	0.520703	0.693361	0.770903
Band 3	0.906649	0.964723	1.000000	0.501491	0.730341	0.817989
Band 4	0.373197	0.520703	0.501491	1.000000	0.780455	0.547554
Band 5	0.573404	0.693361	0.730341	0.780455	1.000000	0.914581
Band 6	0.690439	0.770903	0.817989	0.547554	0.914581	1.000000

Eigenvector	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6
Band 1	0.403798	0.437488	0.440323	0.320187	0.412721	0.422858
Band 2	0.462171	0.292563	0.264202	-0.666253	-0.419595	-0.105117
Band 3	-0.237526	-0.234706	-0.047047	-0.597703	0.332043	0.647132
Band 4	-0.694607	0.200723	0.670674	-0.027087	-0.137862	-0.087676
Band 5	-0.193277	0.443969	-0.373015	0.185450	-0.563829	0.523547
Band 6	-0.217005	0.656145	-0.380993	-0.247412	0.454357	-0.331019

Table 7.2. The enhanced images statistics of eigen vectors, eigen values and band correltons.

One of the problems in using remotely sensed data in the EMR (electromagnetic radiation) is the effect of variability in reflectance of the same feature with respect to the position with respect to solar radiation. This means that although there is the general notion that a surface must receive the same amount of energy irrespective of its orientation to the sun, topographic variation plays a significant role in suppressing and lighting different sides of the same material. Various combinations of band ratios: 5/1,

5/4, 3/4, 5/3, 7/4 have been tried but showed only slight improvement of features contrast in the image.

The fact that it is possible to display a combination of three bands or channels at a time, necessitates dictates the need of a method to choose most informative bands. These can be achieved by principal components analysis. It is a method of improving the spread of data by redistributing them about another set of axes in multi-dimensional space which can not exceed the number of original bands in number. It involves definition of a new axis along the largest spread of the data in such a way that it maximizes the variance of the values of the data plotted with new values projected along the new axes. Two parameters needed for this purpose are the eigen vectors that define the direction of the successively decreasing spread of the data set and the eigen values which define the magnitude/variance of the directions defined earlier. This procedure can continue with specified number of times (which equals to the number of components that is to be obtained) orthogonal to the previous axes successively. This would result in transformation of originally sub equal variance bands to new bands/components of large variances with dramatic decrease in consecutive components.

In this study three principal components were used to redistribute the data from the six original ETM+ bands 1 through 7 with the exception of the 6^{th} band which was the thermal band. The transformation incorporates 98.14% of the total data from the all the bands while the remaining of mainly noisy data has been discarded (Table 7.6).

Basic Stats	**Min**	**Max**	**Mean**	**Stdev**
Band 1	-307.428925	314.101654	0.000000	121.336946
Band 2	-210.871750	205.065598	-0.000000	53.502439
Band 3	-143.976685	216.388809	0.000000	34.345236

PC	**Eigen Value**	**percentage of loading**
1	14722.654348	78.46%
2	2862.510930	93.71%
3	1179.595242	98.21%

Covariance	**Band 1**	**Band 2**	**Band 3**
Band 1	14722.654348	0.000000	-0.000000
Band 2	0.000000	2862.510930	0.000000
Band 3	-0.000000	0.000000	1179.595242

Table 7.3 Table showing the Principal components statistics

7.2.2 Image interpretation for Geologic unit identification

The enhanced image has been classified using maximum likelihood algorithm in ENVI 4.2. The signatures for the classification were created by using regions of interest. The regions of interest for training the classification come from a geologic map adjacent to the study area (EGS, 2005) and field survey ground truths. Specific formations have been included in the training sites (regions of interest) where exposures have been identified using various band combinations corresponding to the geologic maps units of the area identified by the maps. The image has been classified with regional spatial extent including the study area and the mapped areas and interpreted later on by comparing with the local context and field observations.

Seven training classes were utilized for the classification based on the predominant lithologic types in the area. Also two land cover units vegetation and water bodies have been included in the training class. The classification was performed for the band ratio image of 5/1, 5/4, 3/4 combination (Figure 7.3a). Accuracy assessment of the classified image was done by constructing confusion matrix by ROI's (regions of interest), in ENVI 4.2 (Table). It has showed 70% accuracy which is a bit less than the recommenced threshold, 80%. Nevertheless, given the highly interfering reflectance from vegetation and cultivated land it should be acceptable. Furthermore, upon the information obtained from the regionally classified image and field association new sets of training sites were selected and used for further classification (Figure 7.3a). The area previously classified primarily as Aphanitic basalt was segregated into two; one the aphanitic basalt itself and the other cultivated land. Nevertheless what is interesting is that the area under intense cultivation is primarily underlain by the Aphanitic basalt. This can help to infer indirectly the boundary of this formation. Other classification schemes have resulted in much less reliable accuracy. Also changing the bands used for the classification did not bring significant change and even with less accuracy (PC-1, PC-2, PC-3 with 68% accuracy, Figure 7.3b). The interpretation of the lithologic units were then delineated based on the information from the classified image and the visual association of the ground truths during the field visit and the images of band combinations 3, 1, 2; 7, 4, 1 and 5, 3, 4. It has been possible to delineate the upper and lower basalts and ignimbrite units by morphological association where the later ones are found at a topographic low position

and with characteristically dissected nature unlike their counter parts of the upper basalt and ignimbrite units. Nevertheless separation of the units was not possible as their occurrence is abruptly interlayer.

The lithologic units starting from the north western part of the study area are described below. The cliff forming western boundary of the study area is mapped as basalt dominated unit. During field survey it has been witnessed to have well developed vertical joint. It is relatively stable except at few places. It could be part of what has been mapped as Ashange formation in the regional geologic map, which has transitional to Tholietic composition. It marks the western boundary of the study area forming an east facing cliff.

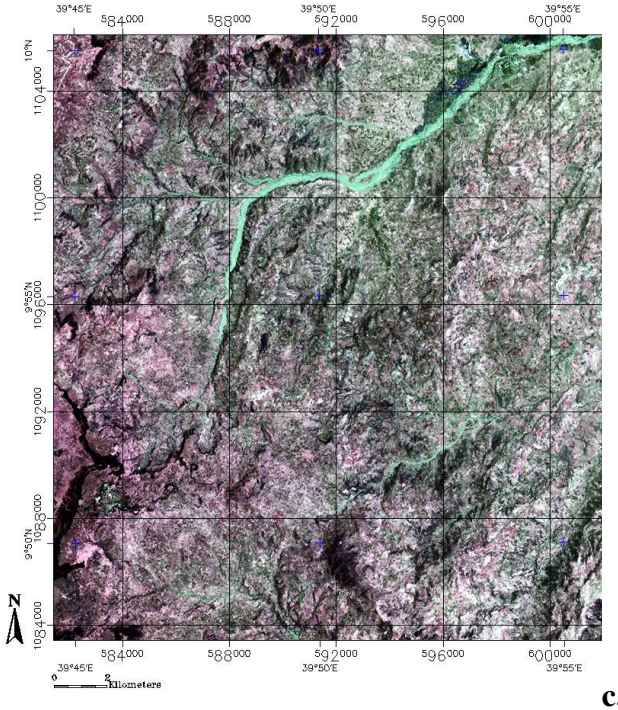
The unit lying right at the foot of the Cliff to the east from Ashange basalt is very similar to it but with numerous intercalations of ignimbrite. This unit is less stable than the previous one due to the low joint spacing and higher weathering going on. It has to be noted that the area is extremely wet, with temperate like climate which has resulted in extreme weathering of rocks which are often jointed.

A dome like feature appears around midway where the head waters of "Dem Aytimash" river flow around. It has been interpreted as Rhyolitic unit from the classified image.

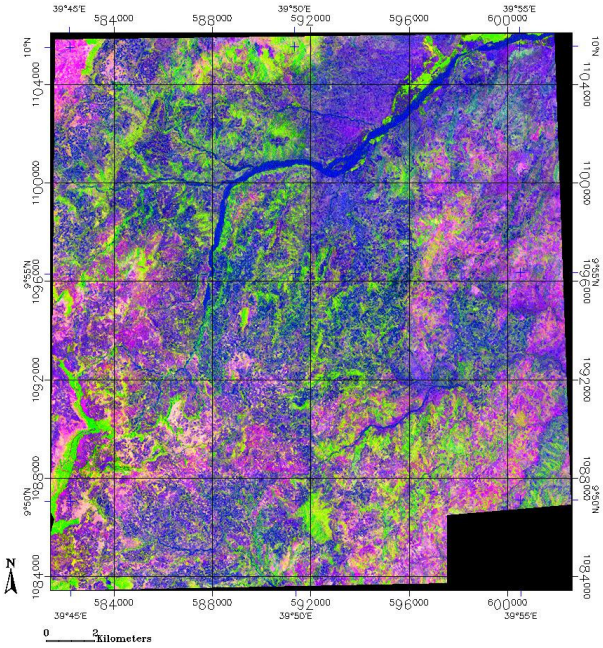
Mid way along the length of the Basalt and Ignimbrite unit and lying below the rhyolithic unit is a localized pyroclastic dominated Ignimbrite where old landslides are observed.

Lying at the foot of the Ignimbrite and basalt unit is the lowest lying unit in the area, the Aphanitic basalt, forming the undulating surface on large scale. South eastern section of the study area is dominated by the extensive Rhyolitic units (Figure 7.2b). Apart from the tertiary volcanics tertiary sediments of poorly consolidated sandstones are observed indicating an interruption in the volcanism (Figure 7.2) along with numerous dyke swarms of basaltic composition. Most of the joint sets in the area are hazardously oriented with gentler or parallel in dip than that of the topographic slope.

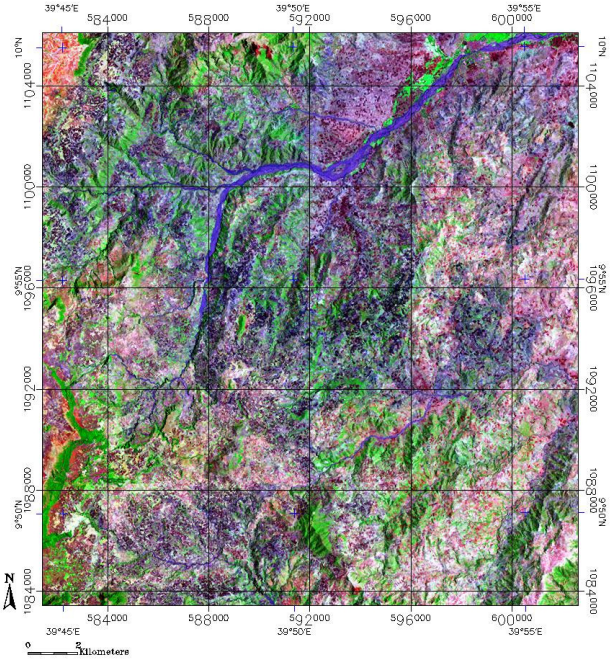
Figure 7.2. Tertiary sediment (sandstone) with cross-cutting dyke in the middle of the photograph. Found in the Aphanitic basalt unit.



a.



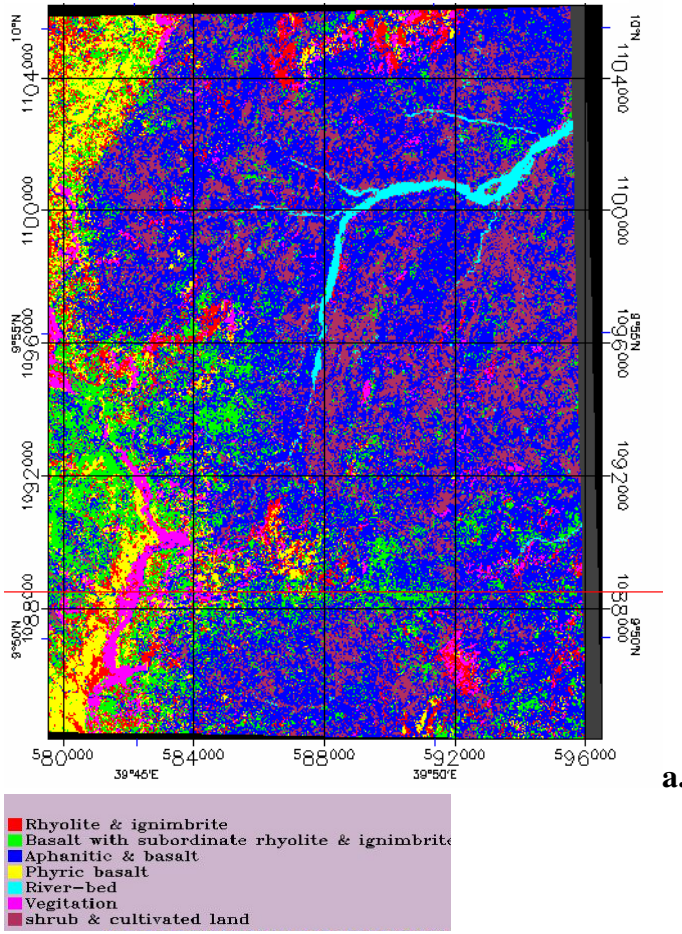
b.



c.

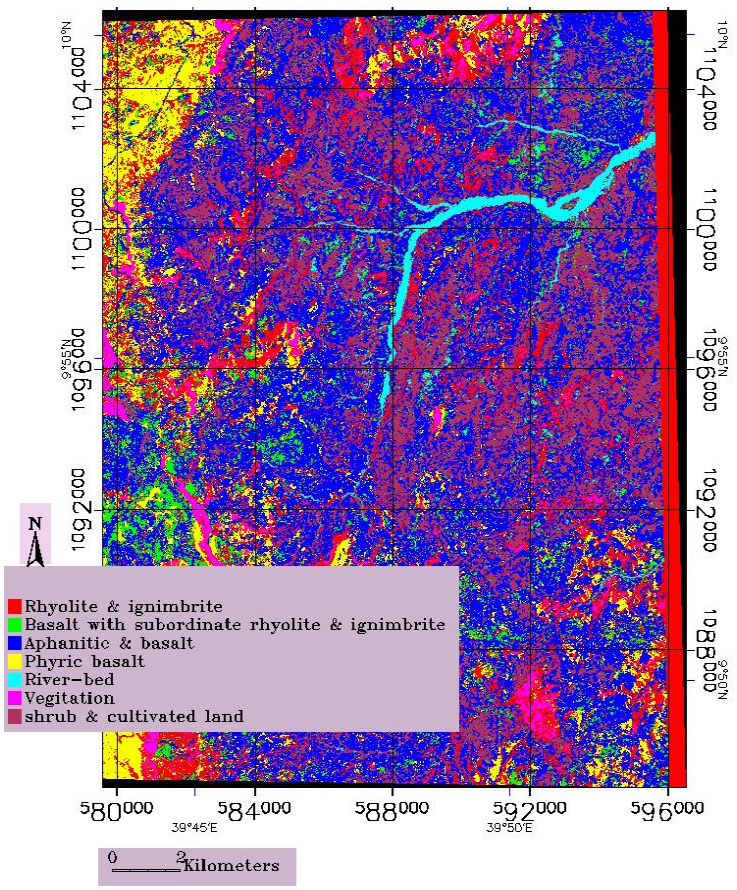
Figure 7.3 (a-c). Traditional band combinations of 312 in RGB order (a) indicating the clear topographic variation in occurrence of the lower lying aphanitic basalts and cliff forming lower basalts. Also visible are the south eastern and eastern part of the area indicating the rhyolithic units distribution (figure b,). Figure c (742 in RGB order) showing the land cover types.

Supervised image classification on /1,5/4,3/4 band ratio images.



a.

Supervised image classification on the first three principal components of the image



b.

Figure7.4a-b. Supervised classification of the indicated band ratio image (**a**) clipped from the regionally classified image and That of first three principal components (**b**)

	No. of ground truth pixels							
	Rhyolite & ignimbrite	Basalt & rhyolite	Aphanitic & basalt	Phyric basalt	River-bed	Vegitation	Shrub & Cultivation land	Total
Unclassified	0	0	0	0	0	0	0	0
Rhyolite & ignimbrite	1233	252	212	124	0	2	0	1823
Basalt & rhyolite	193	549	398	76	0	4	15	1235
Aphanitic & basalt	180	397	2521	13	3	6	202	3322
Phyric basalt	631	302	52	756	0	0	0	1741
River-bed	0	0	0	0	711	0	10	721
Vegitation	8	1	1	3	0	717	0	730
Shrub & Cultivation land	11	105	292	0	8	0	2836	3252
Total	2256	1606	3476	972	722	729	3063	12824

Overall accuracy = 9323/12824, 72.69%

Kappa Coefficient = 0.6652

Table 7.4. Maximumlikelihood classification statistics of 5/7, 5/4, 34 band ratio images

7.3. Preparing the data for GIS analysis.

Before weighting the factors, rating values for the various classes' susceptibility to landslide was obtained. This has been done by obtaining the ratio of area of each class of the parameters to the area of landslide inventoried falling in those classes. This ratio was used to rate the susceptibility degree at class of parameter (factor) level (Table 7.5a to 7.5f). This figure is reclassified to intervals for consideration in the analysis. It is used as a rating value of the relative importance of the individual classes for landslide occurrence. What remained afterwards were the weighting values which are obtained from the Logistic regression analysis and neural networks classification.

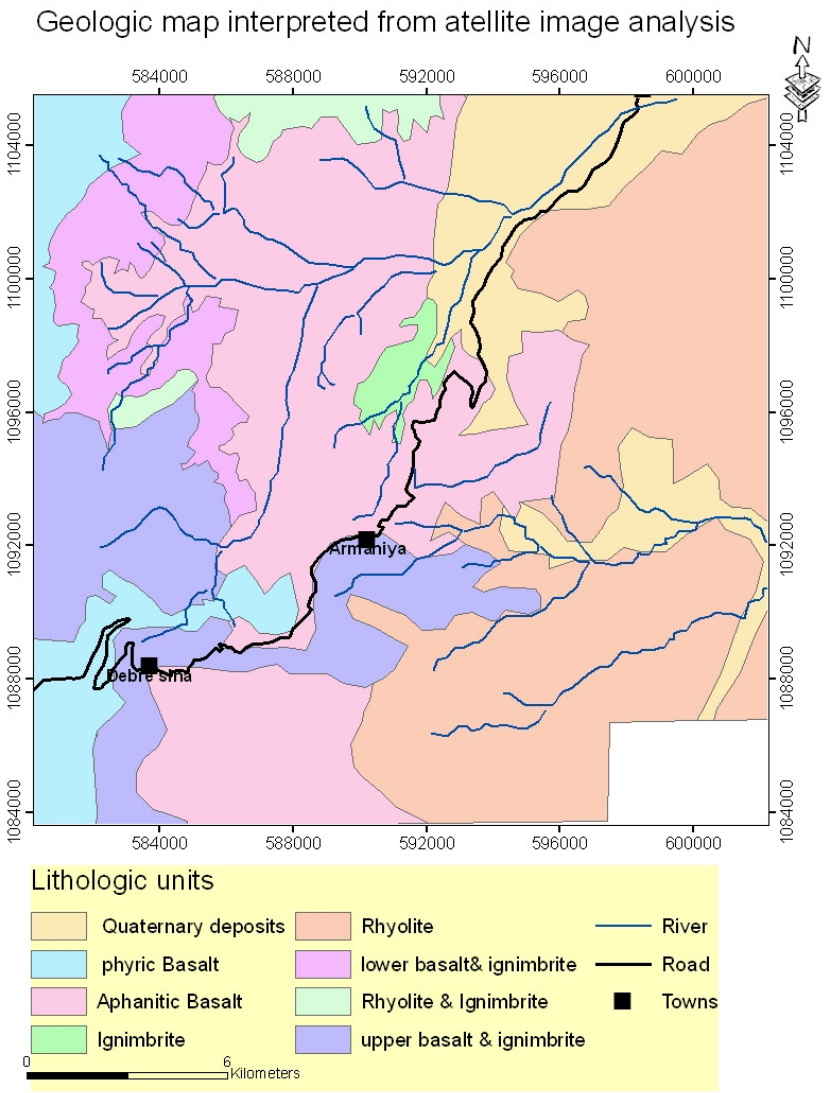

Geologic map interpreted from atellite image analysis
584000
588000
592000
596000
600000
1104000
1100000
1096000
1092000
1088000
1084000
N
Armaniya
Debre sina
Lithologic units
Quaternary deposits
phyric Basalt
Aphanitic Basalt
Ignimbrite
Rhyolite
lower basalt& ignimbrite
Rhyolite & Ignimbrite
upper basalt & ignimbrite
River
Road
Towns
0
6
Kilometers

Figure7.5a

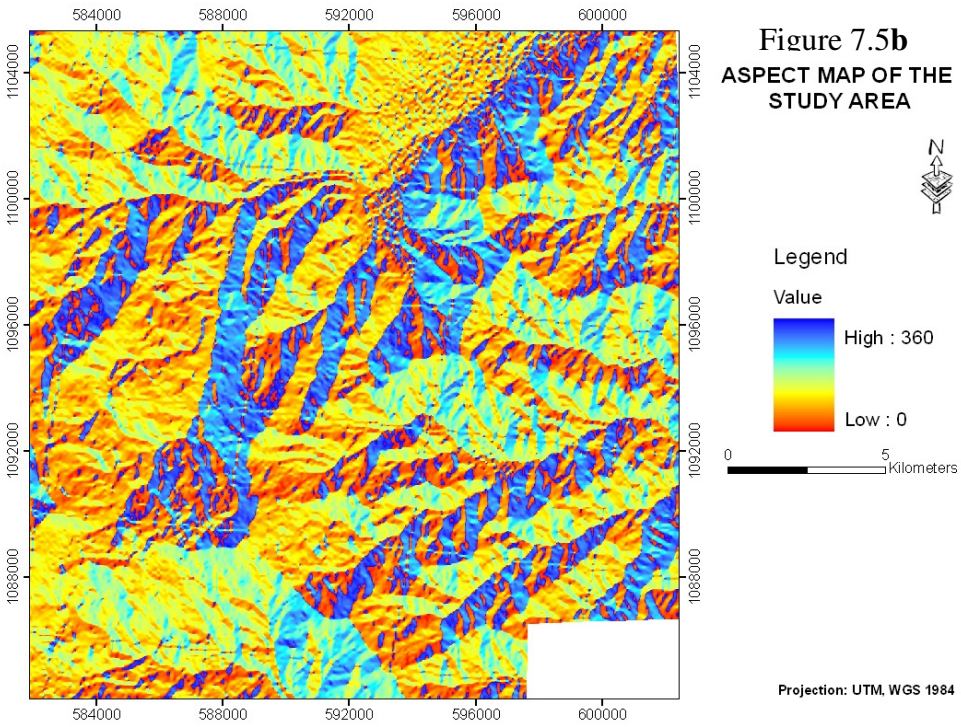


Figure 7.5**b**

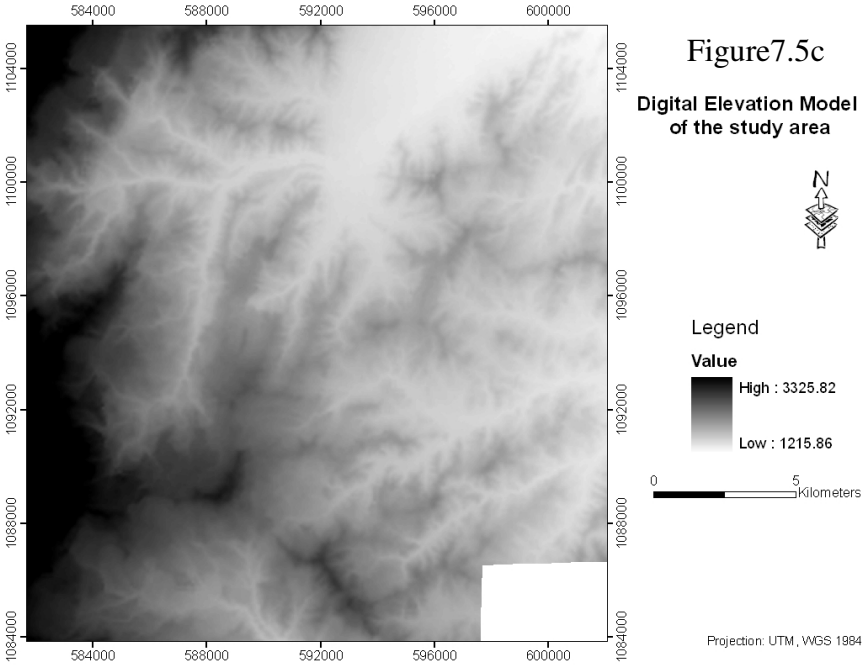


Figure7.5c

Figure 7.5d.

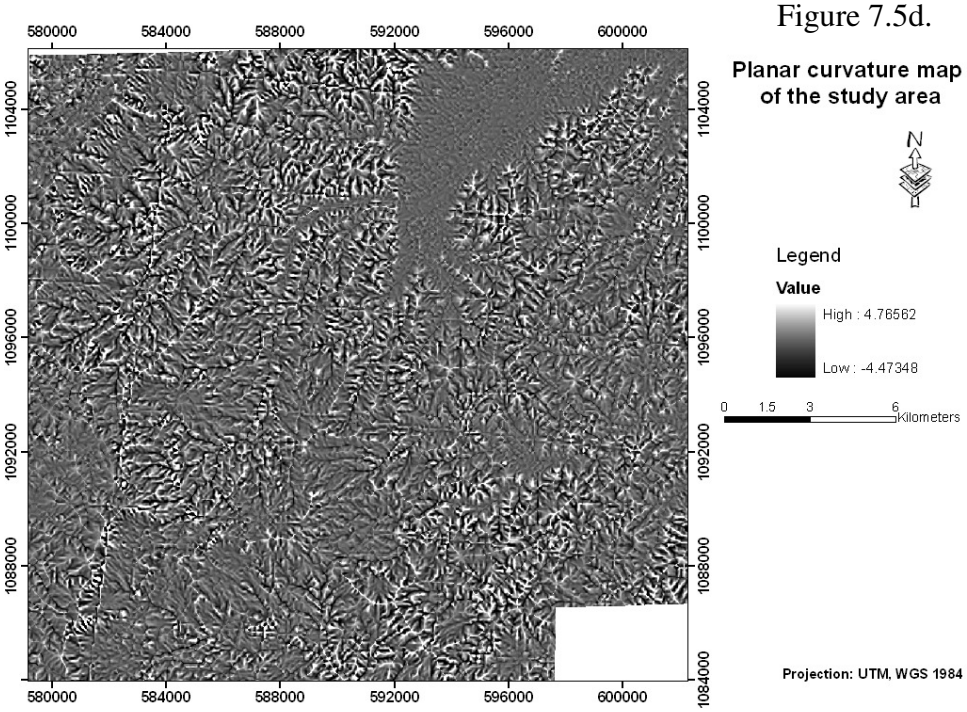
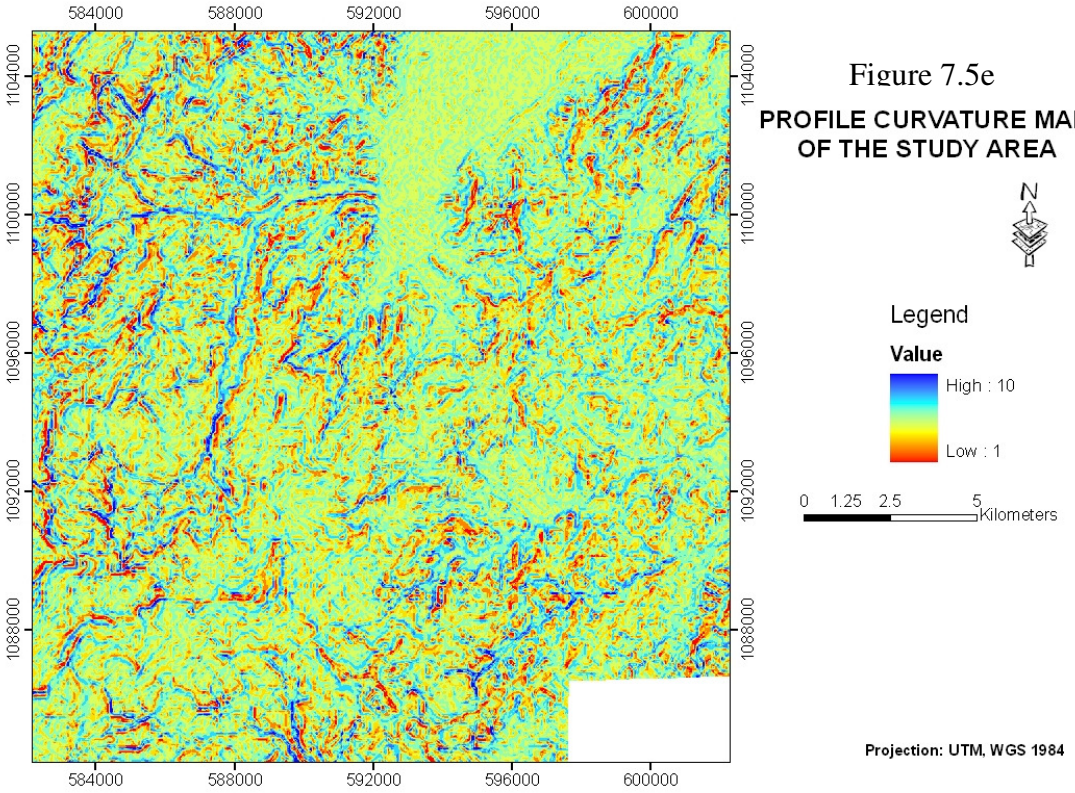


Figure 7.5e



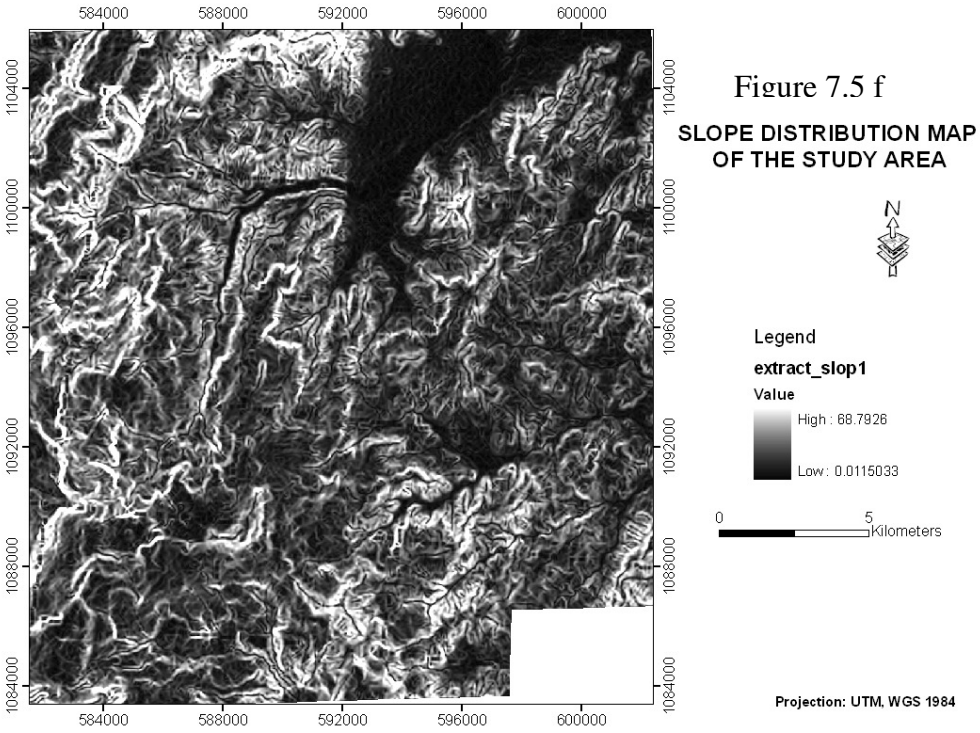


Figure 7.5 f

Figure 7.2 g.

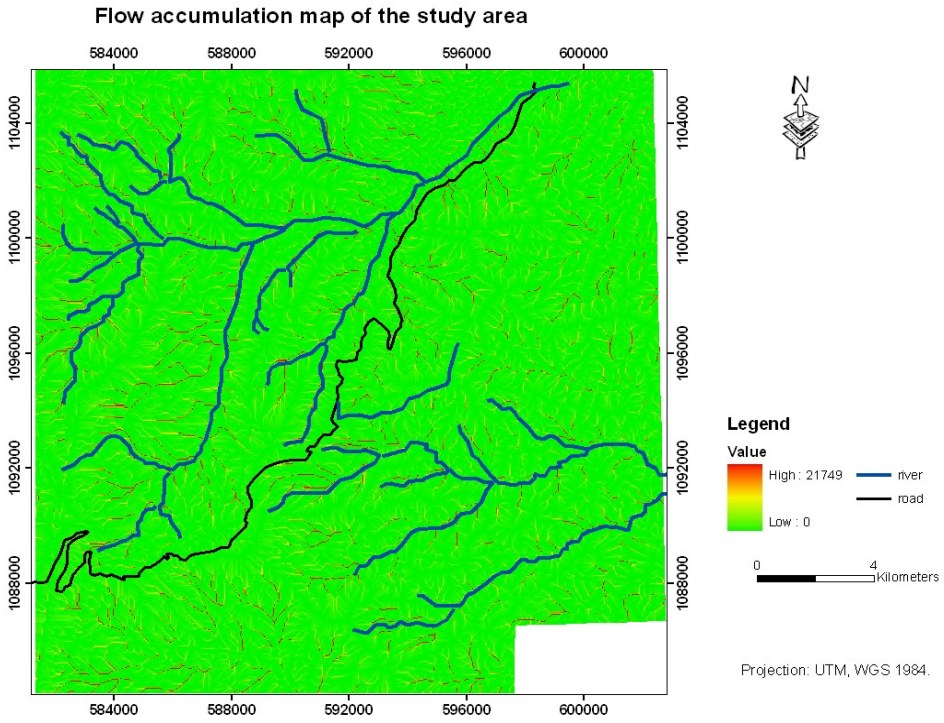


Figure 7.2 **a** to **g**. parameters obtained from DEM.

Aspect (degrees in clockwise from north)	No. of slide Pixels	Total no. of pixels	Index	Rank
0- 59.99	1859	191083	0.009729	3
59.99-119.99	2439	286248	0.008521	2
119.99-179.99	935	223788	0.004178	1
179.99-239.99	1414	111315	0.012703	5
239.99-299.99	1411	101740	0.013869	6
299.99-359.99	1683	159326	0.010563	4

Table 7.5 a. BSA of Aspect parameter for class wise rating.

Slope (in degree)	No. of slide Pixels	Total No.of pixels	Index	Rank
0.01-7.05	2341	415996	0.005627	1
7.05 - 13.32	4212	413808	0.010179	2
13.32 - 19.58	2365	190910	0.012388	3
19.58 - 26.36	684	47312	0.014457	4
26.36- 34.97	132	5156	0.025601	6
34.97 - 66.54	7	317	0.022082	5

Table 7.5 b. BSA of Slope parameter for class wise rating.

Planar Curvature	No. of slide Pixels	Total No. of pixels	Index	Rank
-4.47 - -0.67	0	8	0	1
-0.67 - -0.25	2	386	0.005181	3
-0.25 - -0.02	1529	105510	0.014492	6
-0.02 - 0.20	8138	962381	0.008456	4
0.20 - 0.56	72	5181	0.013897	5
0.56 - 3.81	0	33	0	1

Table 7.5 c. BSA of planar curvature parameter for class wise rating.

Lithologic units	No. of slide pixels	Total no. of pixels	Index	Rank
Ignimbrite	0		0	1
Basalt with rhyolite & ignimbrite	0		0	1
Phyric basalt	84	35327	0.00238	3
lower basalt & ignimbrite	209	68349	0.00306	4
upper basalt & ignimbrite	254	125563	0.00202	2
Rhyolite	1666	329356	0.00506	5
Aphanitic Basalt	7528	337919	0.02228	6
Quaternary deposits	0	139106	0	1

Table 7.5 d. BSA of Lithologic parameters for class wise rating of susceptibility.

No of pixels contributing to flow	No. of slide Pixels	Total Area	Index	Rank
0 – 426	9726	1071595	0.009076	5
426- 1,791	14	1447	0.009675	6
1,791 - 4,179	1	228	0.004386	3
4,179 - 8,102	0	122	0	1
8,102 - 2,793	0	92	0	1
12,793 - 1,749	0	16	0	1

Table 7.5 e. BSA of flow accumulation parameters for class wise rating of susceptibility

Planar curvature	No. of slide Pixels	Total Area	Index	Rank
-4.51 - -0.61	0	16	0	1
-0.61 - -0.23	8	667	0.011994	4
-0.23 - 0.003	3216	254321	0.012645	5
0.003 - 0.24	6430	815939	0.00788	3
0.24 - 0.65	87	2512	0.034634	6
0.65 - 4.21	0	44	0	1

Table 7.5 f. BSA Planar curvature for class wise rating of susceptibility

8. Analysis of Results

8.1. Logistic regression analysis on the data sets

Logistic regression analysis has been performed using the standard statistical software package of SPSS. For that purpose summery statistics of all the data sets was performed and stored as spread sheet file. The model out puts include model chi-square, improvement chi-square, classification table, correlations between variables, observed groups and predicted probabilities chart, residual chi-square for the analysis while coefficient (B), standard error of B, Wald statistic, estimated odds ratio (exp(B)), confidence interval for exp(B), log-likelihood if term removed from model for each of the variable considered in the regression is given.

Analysis of variables using logistic regression requires sample cases with both known dependent and independent variables value. These are used to generate the best fit model statistically. As the SPSS is not tailored for spatial analysis, transferring the data to appropriate form is required. For this purpose, using the spatial analysts extraction tool by sample is used (the algorithm annexed), domain of analysis samples were created. It proceeded by first creating the inventoried and vectorized as polygon features landslide and non-landslide boundaries. These are then set as the dictating features for selecting attributes from all the layers considered in the analysis (which have already been rasterized at 20m cell resolution). Hence two samples slide and non-slide cases were obtained. The outputs are displayed only as a database file of the attributes from all the layers. These can be easily imported to the SPSS package with ODBC or exported to spread sheet program for further filtering of the data and then directly loaded to the SPSS environment. The second option has been favored as it allows to mix the slide and non-slide cases together so that both cases are proportionally represented in the sample. In addition unwanted parameters that can be mistakenly interpreted as a variable can be filtered out with the later option. This resulted in a large sample size of over 24,000 cases of both slide and non-slide cases. The samples need to be split into two one set for fitting the model and the other for validation purpose in logistic regression. This has been left out as there happens to be a good validation of the cases by using the extremely wide landslide that occurred in March, 2006.

The weights hence generated are then multiplied with each of their respective class ratings obtained earlier in the raster calculation tools of ArcGIS. The outputs are then reclassified into intervals of degree of susceptibility.

Model outputs of the logistic regression analysis in SPSS software using the conditional step wise consideration of the variables is given in annex 3.

The coefficients of the Logistic regression output are:

Y = -7.245 + .408slope + .397* Lothology + .004) * Aspect + .001* flow accumulation slope flow contribution + .023* Profile curv. - 0.164* Planar curv.

,where Y stands for the occurrence or non occurrence (nominal) of a landslide based on the values of the independent variables in the right side of the equation.

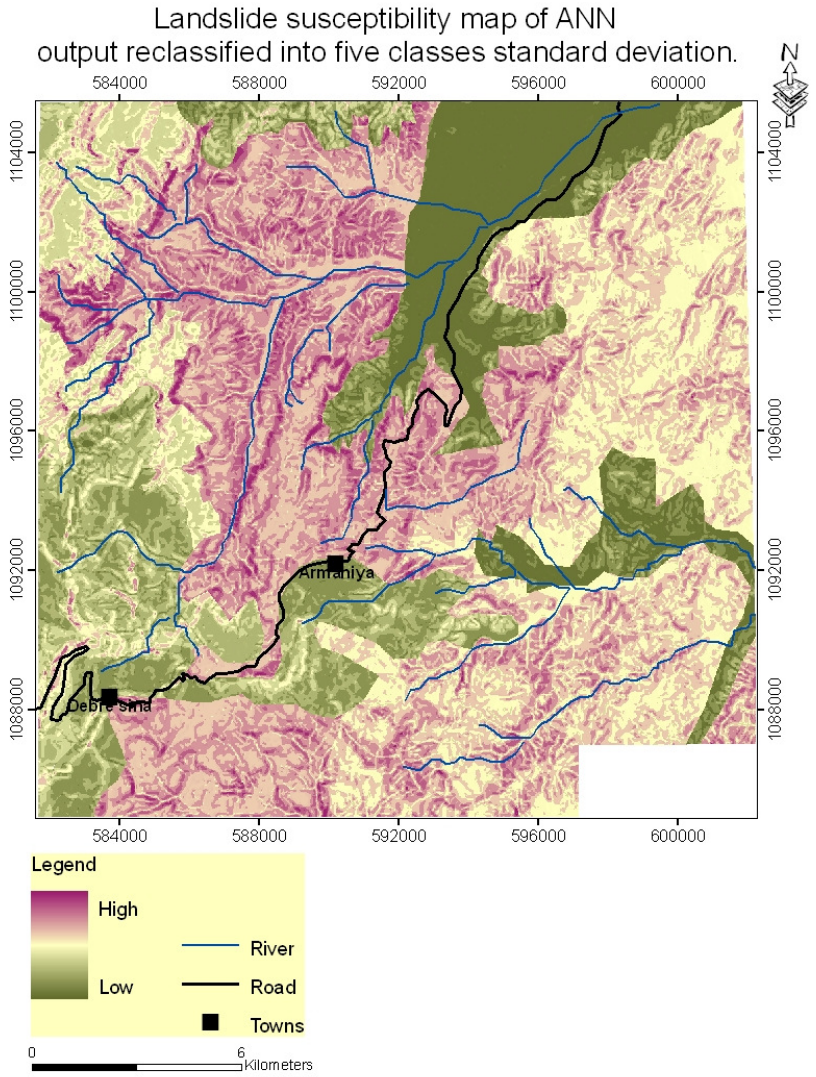


Figure 8.1. Susceptibility output using the Logistic regressions coefficient.

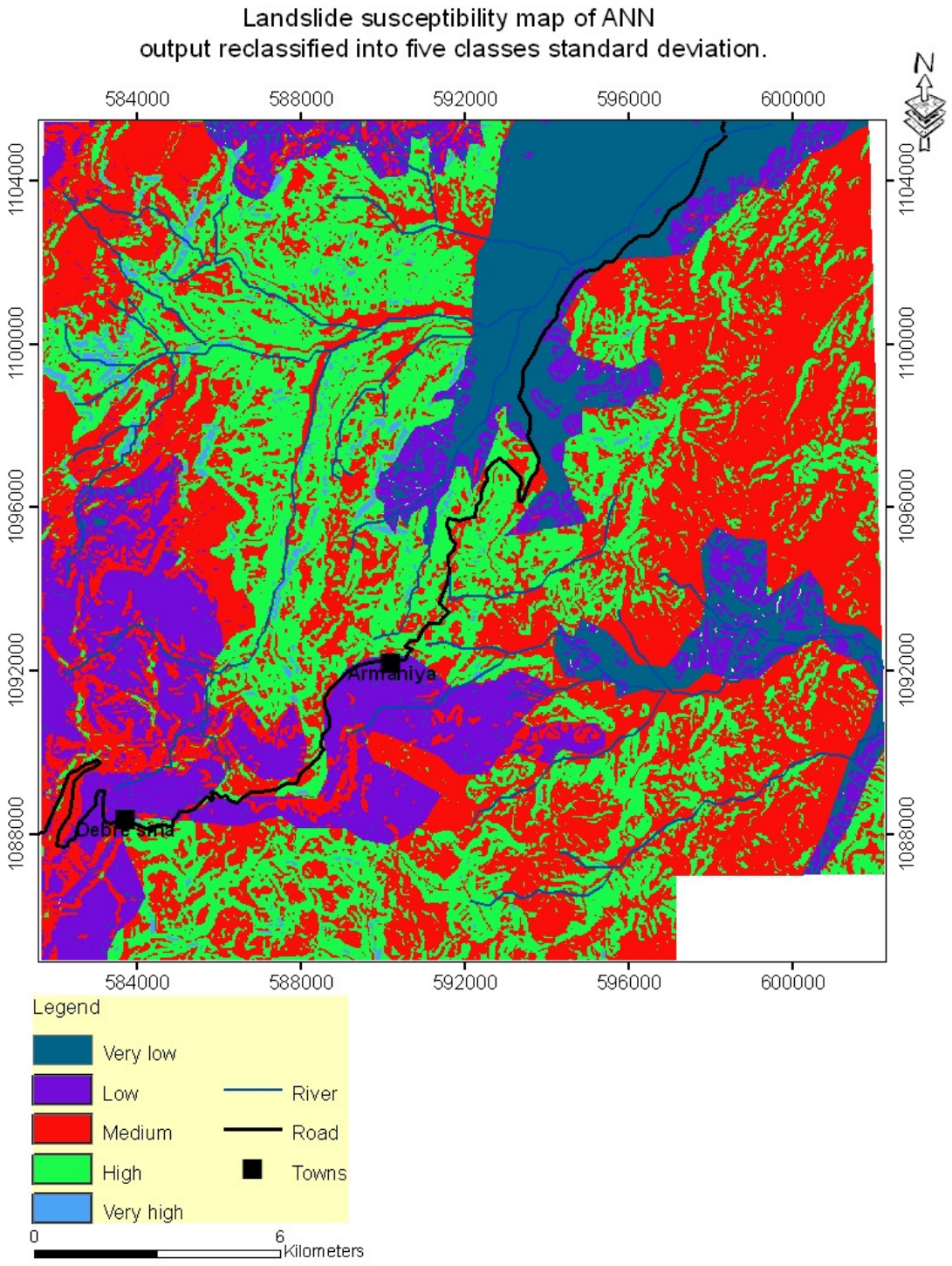


Figure 8.2. Reclassified Susceptibilty output to different levels of susceptibility degree.

8.2. Artificial Neural Networks Analysis (ANN)

For the neural networks analysis, the same sampled data on pixel by pixel base from the seven layers of causative parameters was used for sub-equal slide and non-slide cases. Hence a database of more

than 22,000 entities has been used to train the network using a multilayer preceptron. The first move was to use the Matlab software to construct the network, import the database, and split it into training and adapting phases. But the easiness of using an alternative simpler tool of data mining software was used, which was sufficient as long as we are interested in the weights it generate for the statistical data we imported to the application. Hence the Weka soft ware (Waikato Environment for Knowledge Analysis: version 3.4.10, 2007) was used to classify the entities with the following parameters.

```
=== Run information ===

Evaluator:    weka.attributeSelection.CfsSubsetEval
Search:       weka.attributeSelection.BestFirst -D 1 -N 5
Relation:     landslidedsina
Instances:    22187
Attributes:   7
              slope_recl
              lith_final
              aspec_recl
              upsloprecl
              prof_recla
              plan_recla
              status
Evaluation mode:    10-fold cross-validation

=== Attribute selection 10 fold cross-validation (stratified), seed: 1 ===

number of folds (%)  attribute
          10(100 %)   1 aspect
          10(100 %)   2 Pofile curv
          10(100 %)   3 Slope
           0(  0 %)   4 Lithology
          10(100 %)   5 plan curv
          10(100 %)   6 Flow acc
```

The outputs also include attribute selectors, whereby ranked list of the output is given, but it has been preferred to keep the rest of predictor parameters in the model and use the ranking only for interpretation purpose. Accordingly the lithology and slope parameters have been rated as most decisive factors in predicting landslide in the area.

Based on the analysis the susceptibility map was divided into four classes of quartiles which represent the different susceptibility degree of the areas.

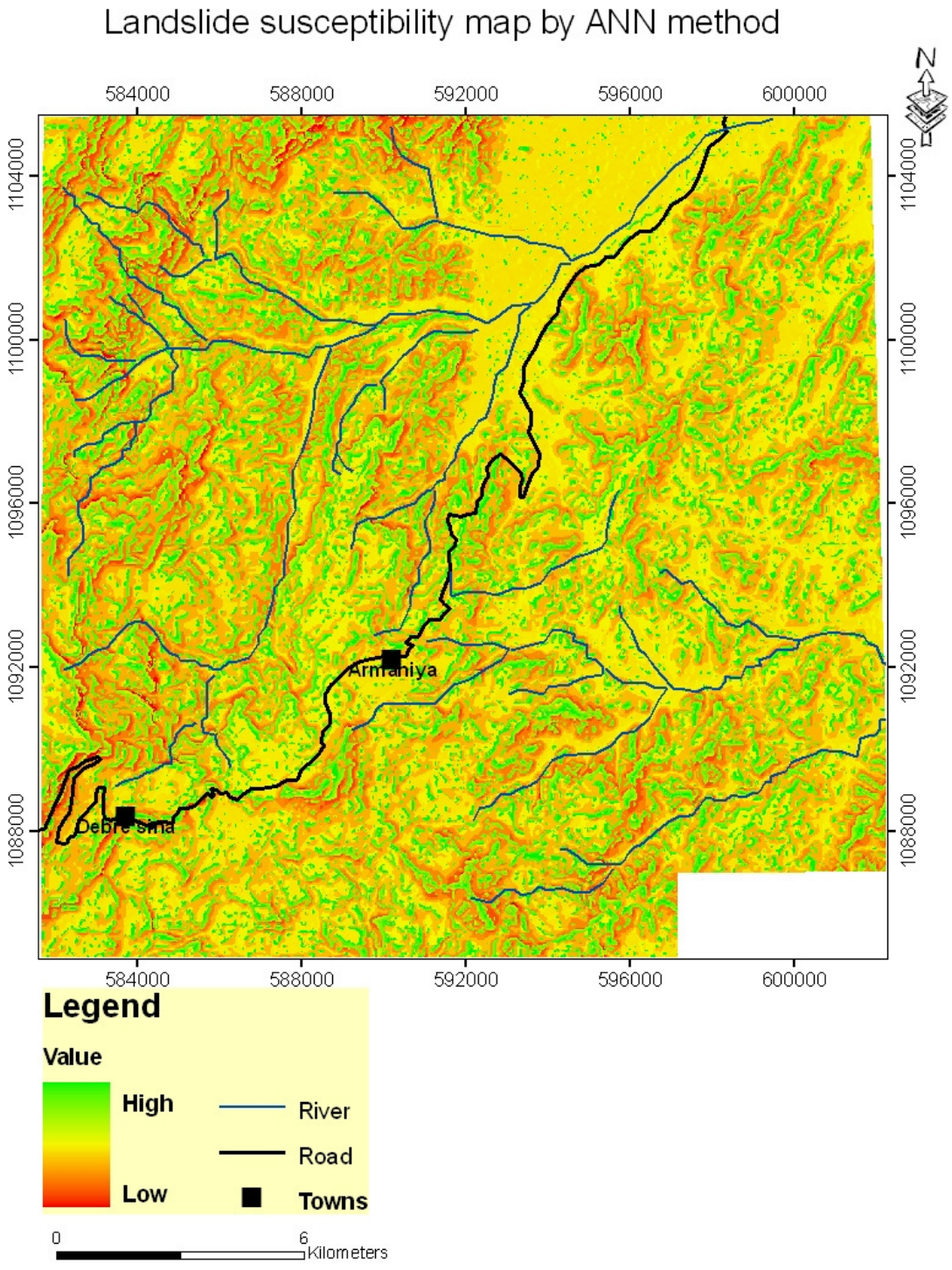


Figure 8.3. Susceptibility output using the ANN classification weights.

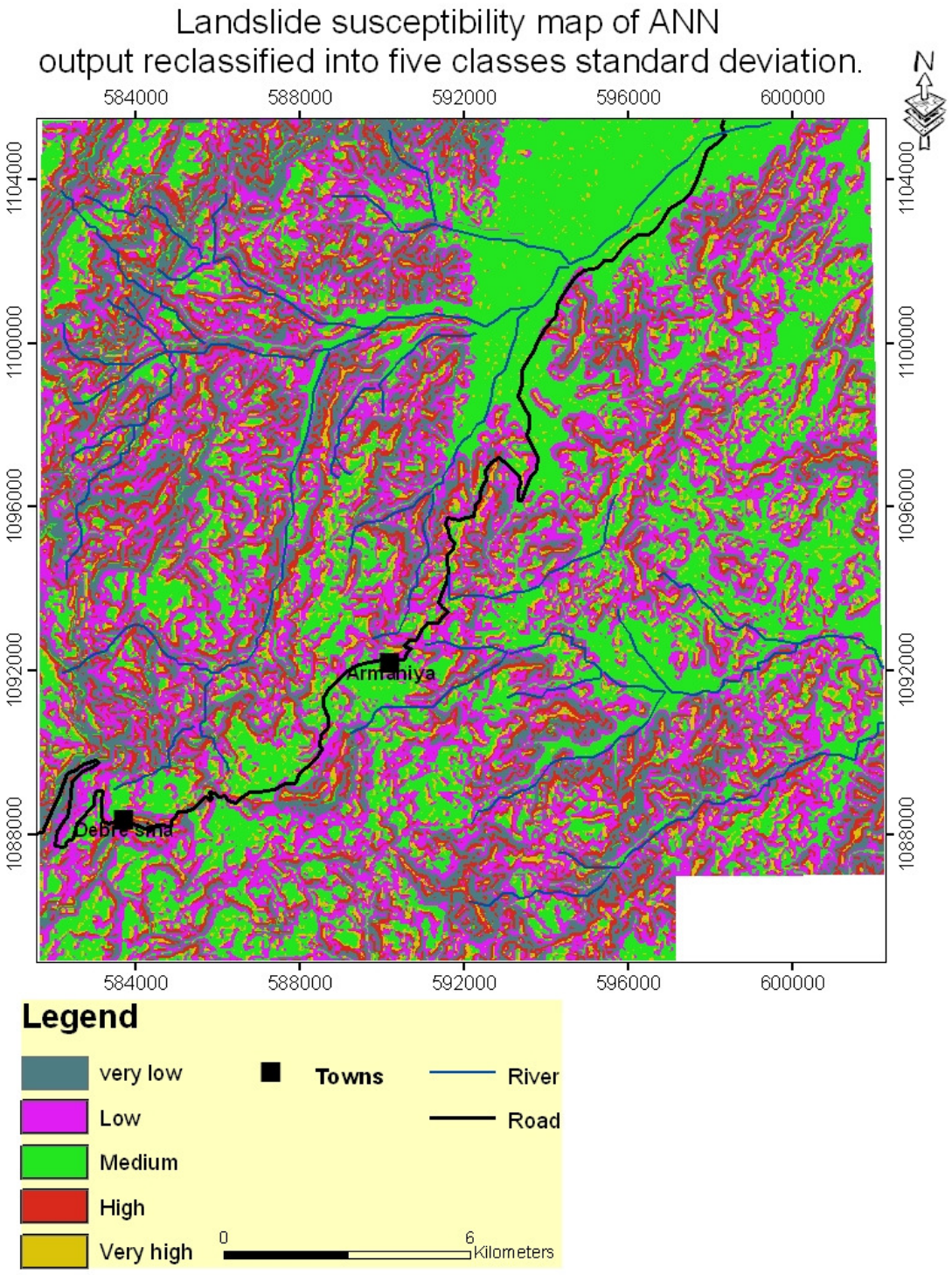


Figure 8.4. Reclassified Susceptibilty output to different levels of susceptibility degree.

8.3. Verification of outputs

In order to verify the practicality of the results, a comparison was made between the two susceptibility maps and a landslide activity map of the same time period including the major massive landslide that occurred in the area.

Accordingly, the massive landslide crowns have been fairly predicted but the foot of the landslide which is 4 to 5 Km away falls in the unit predicted to be more stable due to the fact that the immediate unit has a relatively flat slope which is the most prominent dete ıtive in predicting the landslide susceptibility classes.Outputs of ANN imply a stronger control of landslide susceptibility by aspect, profile curvature and slope and primarily.

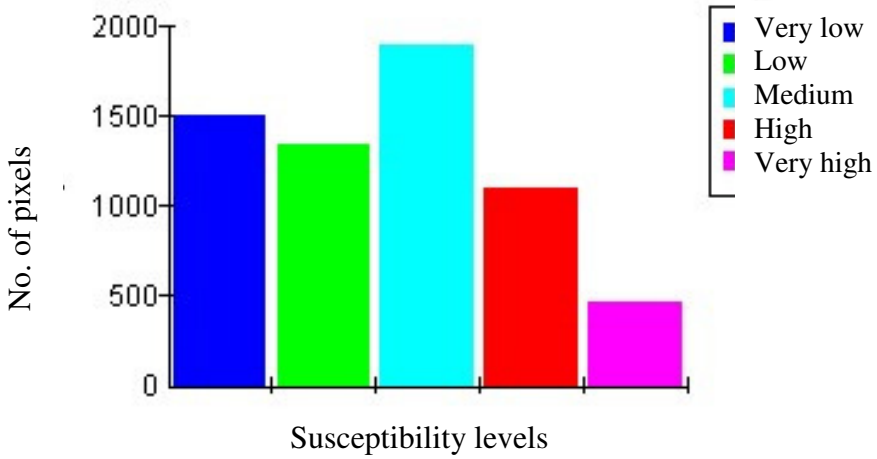


Figure 8.5 a. A histogram of evaluation landslide inventory versus susceptibility classes of the ANN output.

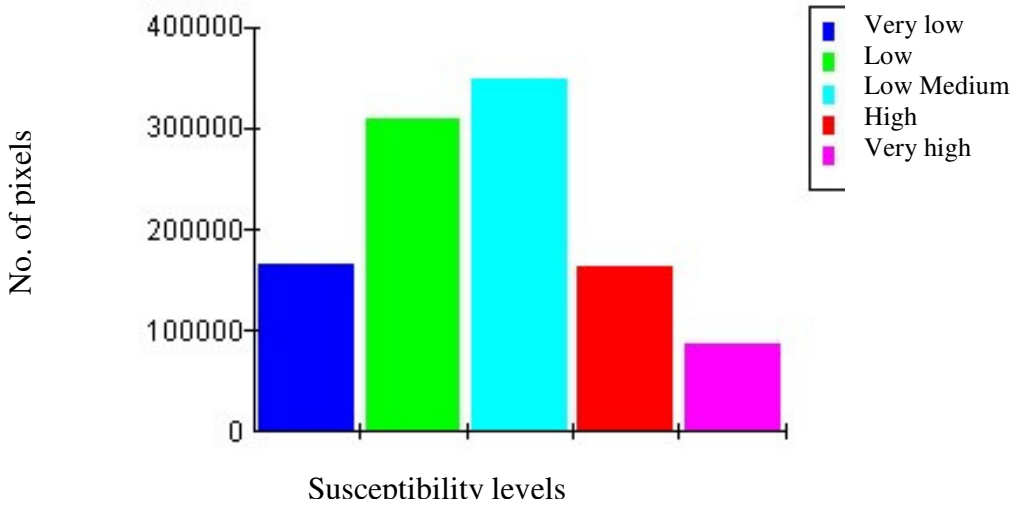


Figure 8.5 b. Histogram of total landslide susceptibility distribution of the Logistic regression analysis. Similar trend by both the evaluation and total susceptibility classes indicate close relationship.

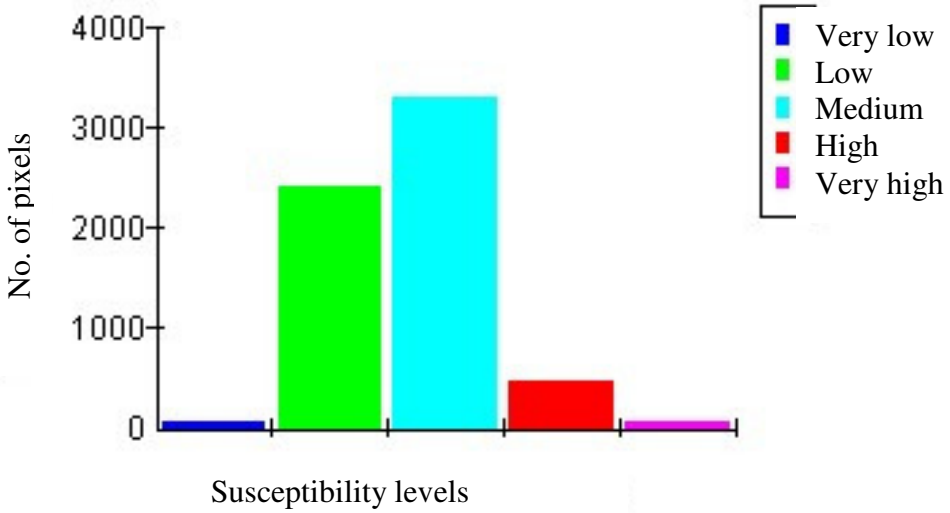


Figure 8.5 a. Histogram of evaluation Landslide inventory events versus susceptibility classes of the Logistic regression method.

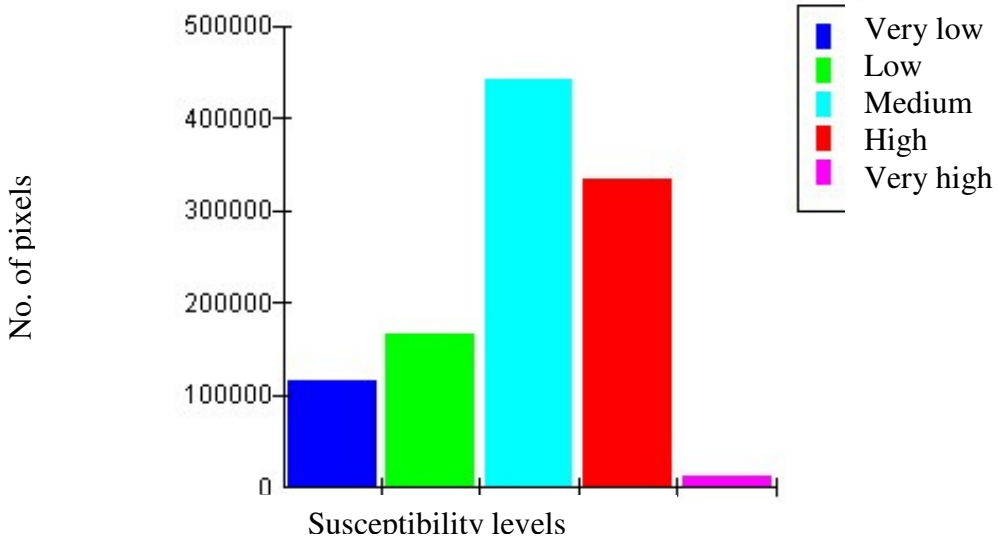


Figure 8.5 b. Histogram of total landslide susceptibility distribution of the Logistic regression analysis. Similar trend in the evaluation and the total susceptibility shows good estimation of susceptibility by this method for the area.

9. Conclusion and Recommendation

9.1. Conclusion

Understanding the processes that lead to land sliding is quite a complex task asking for a very long and exhaustive investigation of an area. It could even get worse if lack of detail data is there. But efforts for susceptibility mapping of a particular area do have some answers to that end of the problem. They can give the professionals invaluable aid in understanding and even predicting landslide processes. The overall effect would then be provision of fundamental knowledge about the evolution of landscapes and setting up the foundation for hazard management and safety measures. Despite their powerful ability to provide such utility as mentioned, most researches seem to focus/prefer specific and deterministic site investigation and try to generalize wrongly the slope stability condition of the region. Nevertheless, efforts are seen here and there resulting in more realistic and representative method of modeling and predicting landslides with increasing flexibility and accuracy.

Two of the most recently applied methods: Logistic Regression and ANN have been tried to be applied simultaneously in this particular study. Keeping the limitations in time, ability to know the applications systems and rarity of available data; it has been used to predict the landslides in the study area, “Debresina”.

In this study care has been taken not to mix up the planar and linear, triggering and conditioning, natural and artificial factors during variable selection. With data on the geologic formation and the geomorphologic parameters landslide susceptibility has been modeled with integration of GIS and other applications of analysis. In fact, the morphologic parameters have accounted significantly for conditioning the area towards landslide occurrence. Hence Slope, Aspect, Flow accumulation, Planar curvature, Profile curvature have been included from the topographic parameters while lithologic parameter has been incorporated from satellite image interpretation. The contribution of upslope flow contribution area has been used in a way to represent the effect of rainfall on landslide occurrence as a preparatory effect rather than triggering.

Each of the methodologies used were preceded by a common preliminary evaluation of the various classes of the parameters (factors) by BSA. Then the weights generated by logistic regression coefficients and ANN classification weights were used separately to obtain two types of susceptibility classes.

The outputs of the analysis indicated that Aspect, flow accumulation and slope as most conditioning factors of landslide in the area while the outputs of the logistic regression indicated significance of Lithology and slope as most controlling factor of landslides. The Logistic regression outputs have been found to be better predictors of the landslides as compared to the artificial neural networks prediction.
The clustering of the predicted landslide sites in certain geologic formations can be explained by the tectonically jointed nature of the rocks and the extreme weathering of the basic rocks. The fact that the area is in general one of the wettest and humid region of the country clearly implies an accelerated mechanical weathering of the rocks that fail carrying along the overlying soil deposits when they other suitable conditions prevail like steep slope and stream under cuttings.

9.2. Recommendation

Application of the same methodology in the area on a wider scale can be undertaken to delineate other potential sites for landslide occurrence to aid in planning and management of land use in the area.
The Susceptibility map can be utilized to come up with a prognostic hazard assessment from risk of infrastructures point of view, where the main high way passes through this highly susceptible zone. Possible realignment may be considered as a solution as there exists a wider susceptible zone for landslide occurrence.

Monitoring mechanisms of the progress of active landslide activity and recording of new landslide occurrence should be undertaken for strengthening further investigation.

Land use/cover change frequency and recurrence of slope failure for a longer period of time with landslides inventory fairly representing the various epochs corresponding to the land use/cover of that epoch has been considered as an interesting problem to be investigated in later study

Reference

1. Ayalew, L., 1999. Causes and mechanisms of slope instability in Dessie town, Ethiopia. Slope stability engineering.
2. Ayalew, L., 2000. Factors affecting slope stability in the Blue Nile Basin. Landslide in research, theory and practice, Thomas Telford, London.
3. Ayalew, L., Yamagishi, H., 2004. Slope failures in the Blue Nile Basin, as seen from landscape evolution perspective. Elsevier. Geomorphology 57 (2004) 95-116.
4. Ayalew, L., Yamagishi, H., Marui, H., Kanno, T., 2005. Landslides in Sado Island of Japan: Part II. GIS based susceptibility mapping with comparisons of results from two methods and verifications. Elsevier. Engineering geology 81- 432-445.
5. Abay River Basin Integrated Development Master PlanProject: climate,Section 2, Volume 3, part 1, phase 2, Ministry of Water resources. pp 1-22.
6. Bonham-Carter, G.F., 1995. Geographic Information Systems for Geoscientists: Modeling with GIS.
7. Brenning, A., 2005. Spatial prediction models for Landslide hazards: review, comparison and evaluation. Natural hazards and Earth System Sciences, 5, 853-862.
8. Capolongo, D., Schattarellla, M., 2005. Implication of earthquake-induced and tectonic evolution. An example from the southern Appennines. Geophysical research abstracts, vol. 7, 05517.
9. Chang, K.T., Liu, J.K., 2000. Landslide features interpreted by neural network using a high resolution satellite image and digital topographic data. Commission VII TS WG VII/5.
10. EGS, 1999. Integrated engineering geological and geophysical investigation for Landslides study in Bonga town and its surrounding. EGS. Ethiopia. Unpublished report.
11. EIGS, 1974. Engineering Geophysical investigation of the Gohatsion to Degen re-route. EIGS
12. EIGS, 1979. A report on the survey of landslides in Mafud Woreda, Yifat and Timuga Awraja Shewa Administrative Region. Disaster preparedness planning program relief and rehabilitation commission. 26 pp.
13. EIGS, 1994. Engineering geophysical investigation of the Blue Nile basin for rerouting of the main road. EIGS. Unpublished report.

14. EIGS, 1996. An integrated geophysical investigation of the Abay basin master plan. Unpublished report.
15. EGS, Unpub. Regional geological map of Debrebirhan and Sela dingay sheets at 1:50,000 scale. EGS.
16. EIGS, 1996. Geologic map of Ethiopia at 1: 2,000,000 million scale.
17. EIGS, 1998. Investigation of slope instability problem in the Blue Nile gorge. EIGS.
18. Ermini, L., Catani, F., Casagli, N., 2005. Artificial neural networks applied to landslide susceptibility assessment. Geomorphology 66 327-343.
19. Foumelis, M., Lekkas, E., Parcharidis, I., 2004. Landslide Susceptibility mapping by GIS based weighting procedure in Corinth area. Bulletin of Geological Society of Greece Vol XXXVI, Proceedings of the 10th International Congress, Thessaloniki.
20. Gezahegn, A., Desse, T., 1996. Engineering Geological Mapping of Blue Nile gorge along the road from Dejen to Gohatsion. EIGS. Unpubl.
21. Gorsevski, P.V, Gessler, P.E., Jankowski P., 2003. Journal of Geographic Systems 5:223-251
22. Hofmann C., Courtillot V., Fe´raud G., Rochette P., Yirgu G., Ketefo E.,and Pik R., 1997. Timing of the Ethiopian flood basalt event and implications for plume birth and global change. Nature 389, 838–841.
23. Hopefield, J.J., 1982. Neural networks and physical systems with emergent collective computational abilities, proceedings of the national academy of Sciences, 79, 2554-2557.
24. Keith, T., and Robert, L., 1996. Landslides: Investigation and Mitigation. Transportation Research Board, National Research Council. 900 pp.
25. Kohnen, T., 1989. Self-organization and Associative Neural networks. NY: Springer 346 pp.
26. Lee, S., Ryu, J., Won, J., Park, H., 2004. Determination and application of the weights for landslide mapping using an artificial neural network. Engineering geology 71 289-302. Special report 247. Transportation Research Board, National Research Council. 500-534.
27. Lemeshow, S., and Hosmer, D.W., 1982. A review of goodness of fit statistics for use in the development of logistic regression models. American Journal of Epidemiology. In Affifi, A. and Clark, V., 1996. Computer Aided Multivariate Analysis.CHAPMAN & Hall. 454 pp.

28. Mackean, J., Roering, J., 2004. Objective landslide detection and surface morphology mapping using high resolution airborne laser altimetry. Geomorphology 57 331-351.
29. Malczewski, J., 1999. GIS and Multi criteria decision analysis. John Wiley and son, Inc. Canada. 420 pp.
30. Pham, D.T., Liu, X., 1997. Neural Networks for Identification, Prediction and Control. Springer publ. 238 pp.
31. Pik. R., Deniel, C., Coulon, C., Yirgu, G., and Marty B., 1999. Isotopic and trace element signatures of Ethiopian flood basalts: Evidence for plume-lithosphere interactions. Geochem. Cosmochem. Act 63, 2263–2279.
32. Rumelhart, D. E. and McClelland, J. L., 1986. Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volumes 1 & 2.
33. Santacana, N., Baeza, B., Corominas, J., Paz, A., Marturia, J., 2003. A GIS based Multivariate statistical analysis for shallow landslide susceptibility mapping in La Pobla De Lillet area (Eastern Pyrenees, Spain). Natural Hazards. Kluwer Academic Publishers. 281-295.
34. Soeters, R., Van westen, C.J.,b., 1996. Slope instability recognition, analysis and zonation. In: Turner, K.A., Schuster, R.L. (Eds.), Landslides: Investigation and Mitigation. Transport Research Booard Special Report 247, 129-177.
35. Thurston, J., (2002). GIS and Artificial neural networks: Does your GIS think? URL address: GISCafe.com. Unpub.
36. Wieczorek, G.F., Mandrone, G., Decolla, L., 1997. The influence of hill-slope shape on debris flow initiation. In: Chen, C.L., (Eds.), Debris flowHAzard Mitigation: Mechaniks, Prediction and Assessment. American Society of Civil Engineers. New york, 1269-1272.
37. Xiaogang, Y., 2005. Implementation of Neural Network Interpolation in ArcGIS: case study for spatial temporal interpolation of Temperature. Unpub. University of Texas at Dallas.
38. Xie, M., Esaki, T., Zhou, G., (2003). GIS based probabilistic mapping of landslide hazard using a three-dimensional deterministic model. Natural Hazards 00: 1-18. Kluwer Academic Publishers.
39. Zhou, G., Esaki, T., Mitani, Y., Xie, M., Mori, J., 2003. Spatial probabilistic modeling of slope failure using an integrated GIS Monte Carlo simulation approach. Elsevier. Engineering Geology 68 373-386.

Annex

Annex. 1

Script used for sampling failed and non failed (stable) units of the ground from the whole raster dataset.

```
# Sampleattrib_sample.py
# Description: Writes a sample of cell values from a group of rasters to a table.
# Requirements: None
# Author: ESRI
# Date: 12/01/03

# Import system modules
import sys, string, os, win32com.client

# Create the Geoprocessor object
gp = win32com.client.Dispatch("esriGeoprocessing.GpDispatch.1")

try:
# Set the workspace
   gp.Workspace = "C:/landslide_proj/data/samples"

# Set local variables
   InRasters = "slope_recl; Lith_final; aspec_recl; upsloperecl; prof_recla; plan_recla"
   InLocationRaster = "slide_invent"
   OutTable = "sampleslide_populn.dbf"

# Check out Spatial Analyst extension license
   gp.CheckOutExtension("Spatial")

# Process: Sampleattrib
   gp.Sample_sa(InRasters, InLocationRaster, OutTable, "BILINEAR")

except:
# Print error message if an error occurs
   gp.GetMessages()
```

```
# Sampleattrib_sample.py
# Description: Writes a sample of cell values from a group of rasters to a table.
# Requirements: None
# Author: ESRI
# Date: 12/01/03

# Import system modules
import sys, string, os, win32com.client

# Create the Geoprocessor object
gp = win32com.client.Dispatch("esriGeoprocessing.GpDispatch.1")

try:
# Set the workspace
    gp.Workspace = "C:/landslide_proj/data/samples"

# Set local variables
    InRasters = "slope_recl; Lith_final; aspec_recl; upsloperecl; prof_recla; plan_recla"
    InLocationRaster = "nonslide_invent"
    OutTable = "samplenoslide_populn.dbf"

# Check out Spatial Analyst extension license
    gp.CheckOutExtension("Spatial")

# Process: Sampleattrib
    gp.Sample_sa(InRasters, InLocationRaster, OutTable, "BILINEAR")

except:
# Print error message if an error occurs
    gp.GetMessages()
```

Annex 2.
Logistic Regression

Notes

Output Created		24-MAR-2007 21:29:47
Comments		
Input	Data	C:\leta_doc\thesis\processed_data\raw_data\After_thesis\total_sample.txt
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	4534
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing
Syntax		LOGISTIC REGRESSION Status /METHOD = ENTER Aspect /METHOD = ENTER FlowAcc /METHOD = ENTER Lithology /METHOD = ENTER Plancurv /METHOD = ENTER Profilecurv /METHOD = ENTER Slope /CRITERIA = PIN(.05) POUT(.10) ITERATE(20) CUT(.5) .
Resources	Elapsed Time	0:00:00.38

Case Processing Summary

Unweighted Cases(a)		N	Percent
Selected Cases	Included in Analysis	4534	100.0
	Missing Cases	0	.0
	Total	4534	100.0
Unselected Cases		0	.0
Total		4534	100.0

a If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
0	0
1	1

Block 0: Beginning Block

Classification Table(a,b)

			Observed		Predicted
			Status		Percentage Correct
			0	1	
Step 0	Status	0	2619	0	100.0
		1	1915	0	.0
	Overall Percentage				57.8

a Constant is included in the model.
b The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-.313	.030	108.422	1	.000	.731

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Aspect	289.596	1	.000
	Overall Statistics		289.596	1	.000

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

			Chi-square	df	Sig.
Step 1	Step	291.669	1	.000	
	Block	291.669	1	.000	
	Model	291.669	1	.000	

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	5884.036(a)	.062	.084

a Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

Classification Table(a)

			Observed		Predicted
			Status		Percentage Correct
			0	1	
Step 1	Status	0	2287	332	87.3
		1	1104	811	42.3
	Overall Percentage				68.3

a The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	Aspect	.006	.000	273.680	1	.000	1.006
	Constant	-1.116	.058	373.103	1	.000	.327

a Variable(s) entered on step 1: Aspect.

Block 2: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	7.506	1	.006
	Block	7.506	1	.006
	Model	299.175	2	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	5876.529(a)	.064	.086

a Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

Classification Table(a)

			Observed		Predicted
			Status		Percentage Correct
			0	1	
Step 1	Status	0	2276	343	86.9
		1	1097	818	42.7
	Overall Percentage				68.2

a The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	Aspect	.006	.000	272.096	1	.000	1.006
	FlowAcc	.000	.000	6.833	1	.009	1.000
	Constant	-1.131	.058	377.995	1	.000	.323

a Variable(s) entered on step 1: FlowAcc.

Block 3: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	16.320	1	.000
	Block	16.320	1	.000
	Model	315.495	3	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	5860.209(a)	.067	.090

a Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

Classification Table(a)

			Observed		Predicted
			Status		
			0	1	Percentage Correct
Step 1	Status	0	2261	358	86.3
		1	1090	825	43.1
	Overall Percentage				68.1

a The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	Aspect	.006	.000	271.472	1	.000	1.006
	FlowAcc	.000	.000	6.429	1	.011	1.000
	Lithology	-.104	.026	16.340	1	.000	.901
	Constant	-.428	.182	5.523	1	.019	.651

a Variable(s) entered on step 1: Lithology.

Block 4: Method = Enter

Omnibus Tests of Model Coefficients

			Chi-square	df	Sig.
Step 1	Step	.098	1	.754	
	Block	.098	1	.754	
	Model	315.593	4	.000	

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	5860.111(a)	.067	.090

a Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

Classification Table(a)

			Observed		Predicted
			Status		
			0	1	Percentage Correct
Step 1	Status	0	2262	357	86.4
		1	1091	824	43.0
	Overall Percentage				68.1

a The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	Aspect	.006	.000	269.058	1	.000	1.006
	FlowAcc	.000	.000	6.058	1	.014	1.000
	Lithology	-.104	.026	16.310	1	.000	.901
	Plancurv	-.039	.124	.098	1	.754	.962
	Constant	-.430	.182	5.567	1	.018	.650

a Variable(s) entered on step 1: Plancurv.

Block 5: Method = Enter

Omnibus Tests of Model Coefficients

			Chi-square	df	Sig.
Step 1	Step	5.801	1	.016	
	Block	5.801	1	.016	
	Model	321.394	5	.000	

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	5854.310(a)	.068	.092

a Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

Classification Table(a)

			Observed		Predicted
			Status		Percentage Correct
			0	1	
Step 1	Status	0	2264	355	86.4
		1	1108	807	42.1
	Overall Percentage				67.7

a The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	Aspect	.006	.000	272.117	1	.000	1.006
	FlowAcc	.000	.000	5.397	1	.020	1.000
	Lithology	-.106	.026	16.972	1	.000	.899
	Plancurv	.036	.129	.080	1	.778	1.037
	Profilecurv	.275	.114	5.786	1	.016	1.316
	Constant	-.424	.182	5.398	1	.020	.654

a Variable(s) entered on step 1: Profilecurv.

Block 6: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	3148.067	1	.000
	Block	3148.067	1	.000
	Model	3469.460	6	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	2706.244(a)	.535	.719

a Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

Classification Table(a)

			Observed		Predicted
			Status		Percentage Correct
			0	1	
Step 1	Status	0	2440	179	93.2
		1	363	1552	81.0
	Overall Percentage				88.0

a The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	Aspect	.004	.001	52.359	1	.000	1.004
	FlowAcc	.001	.000	34.787	1	.000	1.001
	Lithology	.397	.045	78.094	1	.000	1.487
	Plancurv	-.164	.248	.437	1	.509	.849
	Profilecurv	.023	.216	.011	1	.915	1.023
	Slope	.408	.013	997.191	1	.000	1.503
	Constant	-7.245	.360	404.374	1	.000	.001

a Variable(s) entered on step 1: Slope.

Annex 3.
Out put parameters of the ANN model of WEKA 3.10.4 version.

=== Run information ===

Scheme: weka.classifiers.functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a
Relation: lasttry2
Instances: 22696
Attributes: 7
aspect
flow acc
lith
plan curv
profile
slope
status
Test mode: split 66% train, remainder test

=== Classifier model (full training set) ===

Linear Node 0
Inputs Weights
Threshold -0.7193469652217681
Node 1 1.7752457595130031
Node 2 0.8877583472453727
Node 3 1.9112178383478038
Sigmoid Node 1
Inputs Weights
Threshold 16.955310090405327
Attrib aspect 2.07224942497695
Attrib flow acc 3.736104052998569
Attrib lith 3.193420064125631
Attrib plan curv -4.726846027675766
Attrib profile -16.820958023158774
Attrib slope 25.8972219705234
Sigmoid Node 2
Inputs Weights
Threshold -11.573720274431697
Attrib aspect -13.928706461258479
Attrib flow acc 2.470821748094568
Attrib lithology 0.07584980550641542
Attrib plan curv -3.593633185915278
Attrib profile curv 3.6602520036013018
Attrib slope -0.4808630777951031
Sigmoid Node 3

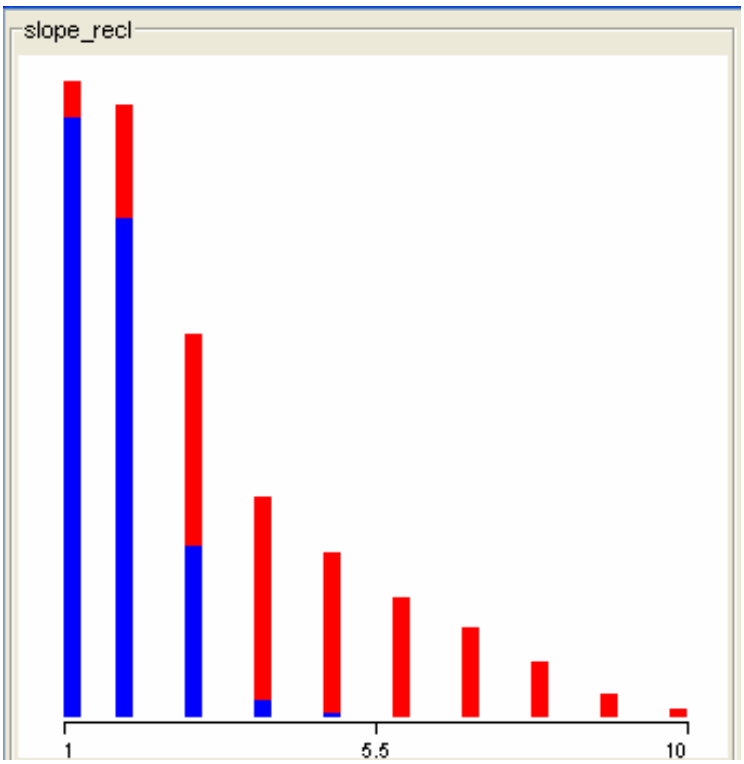
```
   Inputs    Weights
   Threshold    -13.477104274370115
   Attrib aspect    1.1747923846210313
   Attrib flow acc    4.413461965622397
   Attrib lith    -0.2864531149324411
   Attrib plan curv    -0.5965894298858868
   Attrib profile    43.27132740184121
   Attrib slope    -19.952059758633673
Class
   Input
   Node 0
Time taken to build model: 155.83 seconds
=== Evaluation on test split ===
=== Summary ===
Correlation coefficient                  0.8257
Mean absolute error                      0.1433
Root mean squared error                  0.2893
Relative absolute error                 29.2259 %
Root relative squared error             58.3934 %
Total Number of Instances            7717
Evaluator:    weka.attributeSelection.CfsSubsetEval
Search:       weka.attributeSelection.BestFirst -D 1 -N 5
Relation:     lasttry2
Instances:    22696
Attributes:   7
              aspect
              flow acc
              lith
              plan curv
              profile
              slope
              status
Evaluation mode:    evaluate on all training data
=== Attribute Selection on all input data ===
Search Method:
          Best first.
          Start set: no attributes
          Search direction: forward
          Stale search after 5 node expansions
          Total number of subsets evaluated: 26
          Merit of best subset found:    0.705
Attribute Subset Evaluator (supervised, Class (numeric): 7 status):
          CFS Subset Evaluator
          Including locally predictive attributes
Selected attributes: 1,5,6 : 3
                    Aspect, Proifle, Slope.
```

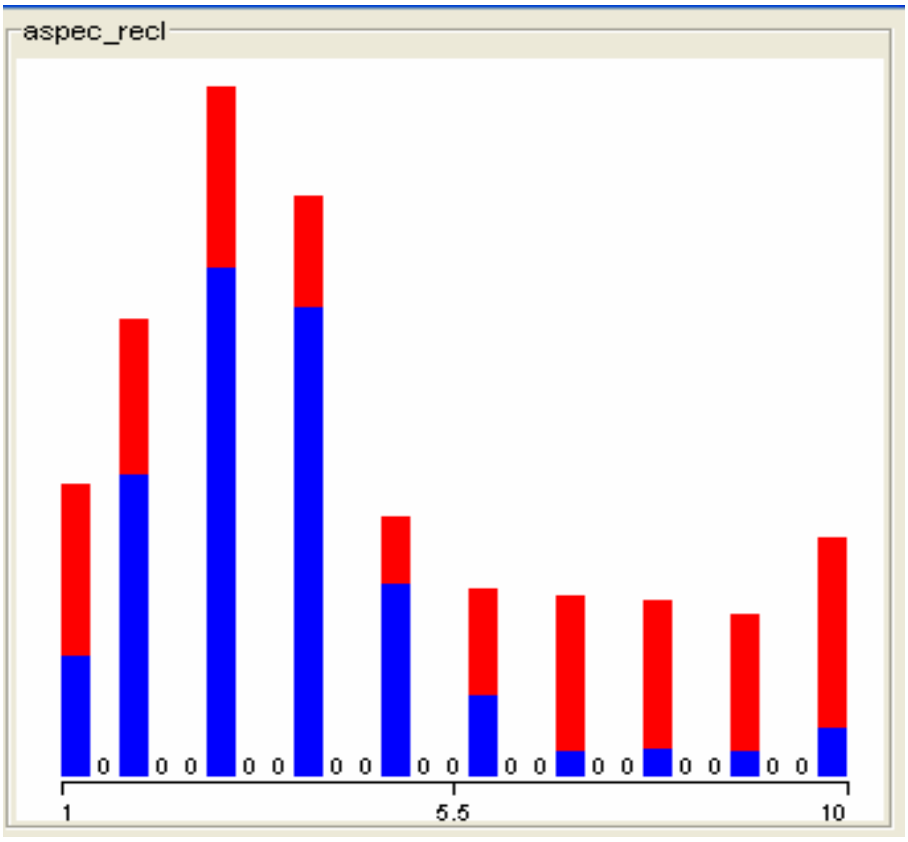
Annex 4.

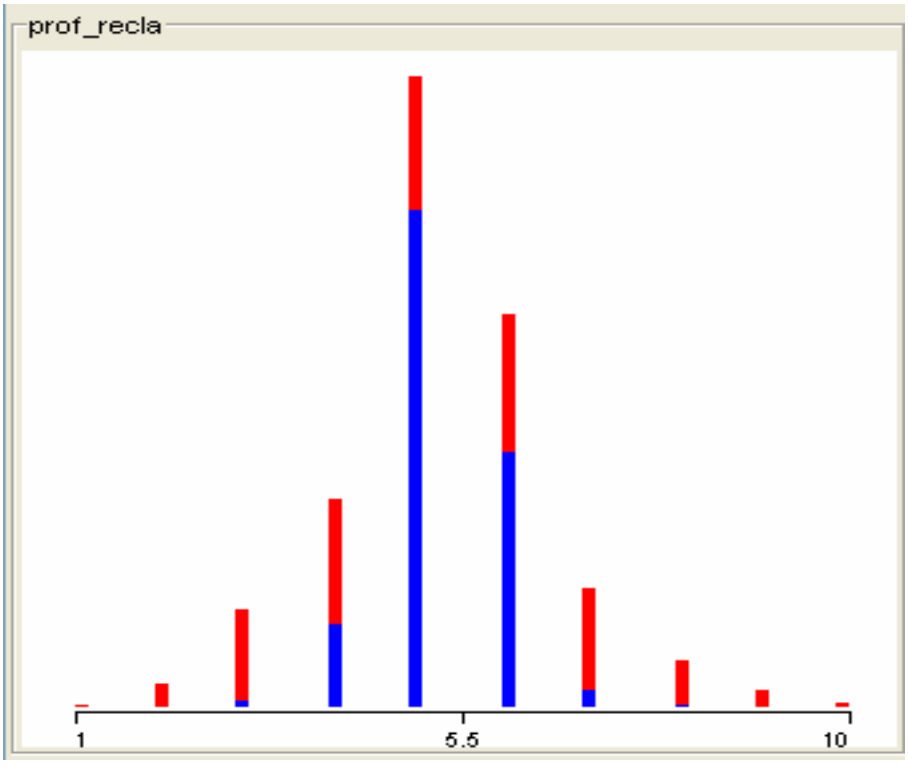
Plots of cross comparison of the dependent variables

Annex 5.

Comparison histogram of the dependent variables (causative parameters with that of landslide and non landslide cases)







Red blocks are non slide portions
Blue blocks are slide portions.

Annex 6. Regional geologic map of Debrebirhan and Seladingay sheets at 1: 50,000 scale.