Landslide susceptibility prediction based on GIS and Artificial Neural Network

C K Samarasinghe

2020

Landslide susceptibility prediction based on GIS and Artificial Neural Network

A dissertation submitted for the Degree of Master of Computer Science

C K Samarasinghe
University of Colombo School of Computing
2020

UCSC

DECLARATION

The thesis is my original work and has not been submitted previously for a degree at this or any other university/institute.

To the best of my knowledge it does not contain any material published or written by another person, except as acknowledged in the text.

Student Name : C K Samarasinghe

Registration Number : 2015/MCS/067

Index Number : 15440675

Signature: Date:

This is to certify that this thesis is based on the work of

Mr. C K Samarasinghe

under my supervision. The thesis has been prepared according to the format stipulated and is of acceptable standard.

Certified by:

Supervisor Name: Dr. M I E Wikramasinghe

Signature: Date:

ABSTRACT

Landslide is one of the hazardous events which cause lives and property damages. Many factors, such as soil type, slope of the terrain, precipitation and manmade activities incorporate for landslides. Frequent of occurring landslide is increasing in Sri Lanka due to the erratic rainfall pattern with climate change especially in the monsoon seasons. Therefore, to prevent loss of lives and damage to property, proper observation and analysis of unstable slope behaviour is crucial.

This research inspects the success rate and the effectiveness of using the Artificial Neural Network (ANN) which is a widely used machine learning technique for pattern recognition and prediction purposes and using modern GIS tools and technologies. Badulla district was selected as the study area of the study which is a protruding geographical area of the country when considering landslide susceptibility due to uneven sloppy shapes of the terrain.

Seven landslide causing factors are selected after performing statistical and GIS analysis. The factors consist precipitation, surface slope, aspect, soil density, land use and the distance to water bodies. The extracted data of these factors are then used for the ANN network training process. The factors were identified by interpreting satellite images, topographical maps and factor layer maps obtained from National Building Research Organisation and Survey Department. ArcMap for desktop and several other GIS tools were used for GIS analysis and factor data extraction.

A comprehensive geospatial database was created using the above collected factor data. This geospatial database contains data related to the terrain, surface as the varying factors. Precipitation was identified as the most prominent and triggering factor for landslides. The above geospatial database was then used for landslide susceptibility prediction using Artificial Neural Network. MATLAB scripting language was used for ANN network creation and for manage training process.

After training and several neuron weight adjustments, successful landslide prediction results could be obtained by the network. These results were further processed and then used for the evaluation processes. The Evaluation was done comparing several previously occurred known landslide events and the prediction results returned by the above trained ANN network. Several geoprocessing tools of ArcMap were used for the comparing process. Further, the performance measures of ANN was determined by calculating accuracy, precision, recall and confusion matrix. The network returned 71.3% prediction accuracy according to the performance calculations.

ACKNOWLEDGEMENT

I would like to express my deep and sincere gratitude to my research supervisor, Dr. M I E Wikramasinghe, University of Colombo School of Computing, for excellent guidance throughout this research. His dynamism, vision, sincerity and motivation have deeply inspired me. It was a great privilege and honour to work and study under his guidance. I am extremely grateful for what he has offered me.

My sincere thanks goes to Dr. B H R Pushpananda, project coordinator(MCS) for whose guidance and comprehensive online support toward us, for the fulfilment of this task. Furthermore, I would like to extend my thanks to all academic and non-academic staff of the University of Colombo School of Computing for their valuable support and acceptance throughout the project.

I thank to National Building Research Organization, Survey Department of Sri Lanka and Department of Meteorology for providing data, maps and directions which were essential for the accomplishment of the research study. And I especially thankful to Mr. Dayan Munasinghe (Scientist) of National Building Research Organisation, for giving me enormous support and the contribution to the study.

Finally, I owe my deepest gratitude to my parents for their dedication and the many years of support during my postgraduate studies that provided the foundation for this work. Last, but not least, I would like to thank my friends who were giving encouragement at the end what made this possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ANN	-	Artificial Neural Network
GIS	-	Geographic Information Systems
GPS	-	Global Positioning System
NBRO	-	National Building Research Organisation
R&D	-	Research and Development
HZM	-	Hazard Zonation Map
DEM	-	Digital Elevation Model
AHP	-	Analytical Hierarchical Process
ESRI	-	Environmental Systems Research Institute
WGS	-	World Geodetic System
SLSD	-	Sri Lanka Survey Department
DMET	-	Department of Meteorology

CHAPTER 01

INTRODUCTION

CHAPTER 01

INTRODUCTION

1.1 Background

Landslides are amongst the most dangerous and damaging natural disaster in the mountainous terrain of a country. Every year, hundreds of people all over the world lose their lives due to landslides; furthermore, there are large impacts on the local, regional and global economy from these events. Over the past few years, many governments and international research institutions across the world have invested considerable resources in assessing landslide susceptibilities and in attempting to produce maps portraying their spatial distribution. [1]

Figure 1:1: Aranayake Landslide- 2016 May

1.1.1 Landslide Prone areas and Susceptibility in Sri Lanka

Sri Lanka is an island in the northern Indian Ocean just south of southernmost part of India and extends in latitude form approximately 060 N to 100 E and in longitude from approximately 800 N to 820 E with an extent of about 65,000 km2. Adams (Adams, 1929) was the first to draw attention to the existences of three well-marked plains of erosion cut in the precam brian rocks of Sri Lanka. These "three terraces" present three successive stages of denudation brought about by successive uplift of Island as a whole. [2]

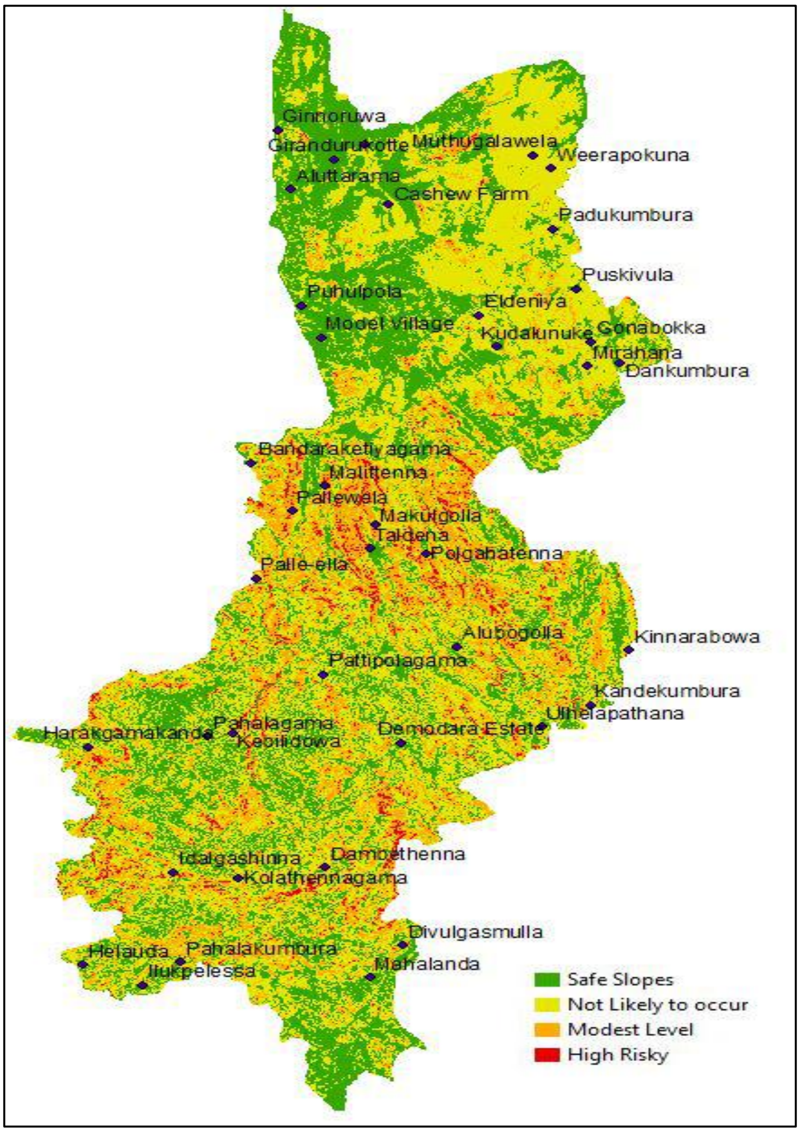


Figure 1:2: Landslide susceptibility areas in Badulla district: Data source- NBRO

> "Landslide is defined as the movement of a mass of rock, debris or earth down the slope, when the shear stress exceeds the shear strength of the material." [3]

The recent detailed structural and tectonic mapping of central highlands indicates that in addition to the vertical epirogenic movements of the southerly drifting manipulate of Sri Lanka, there are horizontal thrust developed regionally and a series of strike slip faults along mega-lineaments.

Some of the lineaments appear to be active and some of the older highly weathered lineaments are commonly associated with large destructive recurrent landslides. Since there are no simple mechanisms to foresee instabilities, monitoring is the most appropriate mechanism to understand their behaviour.

NBRO (Sri Lanka) was aroused the need of early warning systems and it came up with atmospheric models that forecast the temporal and spatial distribution of rainfall which is a useful alternative source of rainfall information. [2] This is now increasingly being used in hydrological applications though it mainly focuses on rainfall information. Figure 1:2 shows Landslide susceptibility map in Badulla district.

Landslide disaster risk reduction demands integration of a number of disciplines associated with various aspects of physical and hydro-meteorological characteristics of a region as well as social and cultural dimensions. Therefore, landslide disaster mitigation requires collective and corporative efforts of all relevant R&D institutions lead by strong executing organizations of the country.

1.2 Motivation

Sri Lanka is a country which gradually becoming landslide are one of major hazard influences over past few years. Heavy precipitation is the triggering factor for most landslide incidents in Sri Lanka. (Figure 1.1 shows Aranayake Landslide event -in 2016 in Kegalle district which caused 120 residents have been reported missing and more than 17 deaths) [4]

As per the statistical data by National Building Research Organisation, lots of lives buried under landslides in 2016, 2017, 2018 and in 2019. Figure 1.3 shows past landslide events in 1947-2007 in Sri Lanka.

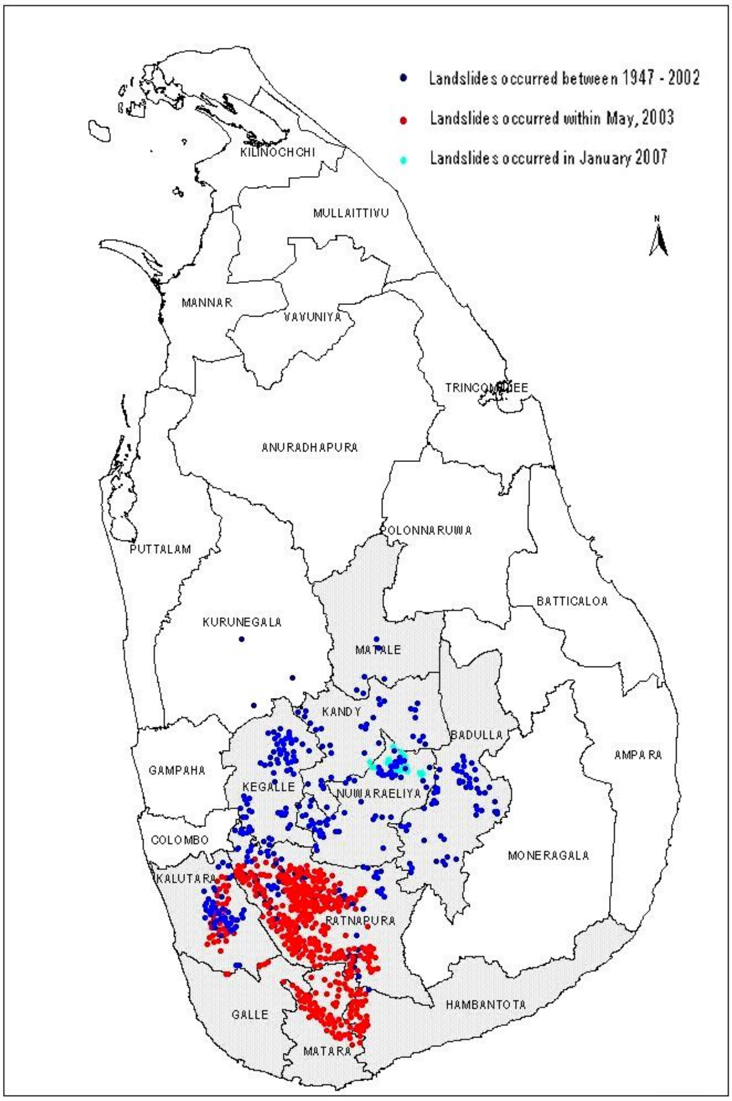


Figure 1:3: Distribution of past Landslide events in Sri Lanka

Even though there have been number of serios landslide events in different areas of the country, still people in these risk areas are not properly aware of what actions to be taken or where to evacuate to a safe place in an event of landslide take place. The major reason for this much of deaths and severe damages to properties is, still there is no proper method of notify the people in high risky areas about the susceptibility of landslides. Then many lives would have been saved since people can immediately evacuate the area or take precautions when they are pre-notified about propension of landslide occurrence.

Currently there are some GIS techniques used to predict landslide susceptibility. But the above statistical data depicts that the success rate of these techniques are still in very low value. This research attempts to improve the success rate of landslide predictions by using GIS and Artificial Neural Networks.

1.3 Aims and Objectives

The main objective of this research is to predict landslide susceptibility (as maps) caused using the GIS analysis and modelling it to an Artificial Neural Network with a considerable success rate. Landslide occurrence study area is selected based on interpretations of GIS data. (A spatial database considering certain geographical factors). The above factors and related GIS data will be used to generate a landslide susceptibility map.

This research applied two statistical analysis methods in a GIS and compared their results. Based on the findings, we were able to derive a more effective method for analysing landslide susceptibility. The spatial prediction of landslide susceptibility applying artificial neural network.

The resultant data (Predicted data) and landslide susceptibility map can be used to analyse and forecast other geographical hazard which relates to selected factors.

Apart from the above-mentioned objectives, the following specific objectives are formulated:

- To create the geospatial database for landslide susceptibility study using GIS data.
- To delineate a Landslide Hazard Zonation mapping with high accuracy by employing Artificial Neural Networks and GIS integrated model.

1.4 The Problem

Sri Lanka is a tropical island contains mountains in the central region. Every year, Sri Lanka experiences landslides, resulting in property damage and casualties during winter months due to two monsoons. The combination of the geological formation, improper land use practices and heavy rainfall has caused for irregular landslides throughout the hill country. Approximately 20% of the area from 65,000 km2 of total area in Sri Lanka was identified as the landslide susceptible areas by the National Building Research Organization (NBRO) as the nodal governmental agency engaged in landslide studies in Sri Lanka. These Landslides threaten lives and property especially throughout the up country.

National Building Research institute together with Survey Department of Sri Lanka conduct a hazard prediction system to predict and warn sophisticated hazard event in several areas. But up to now no existing proper early warning system for landslide susceptibility areas in Sri Lanka.

The above situation exposes a conclusive requirement to a proper prediction system for landslide occurrences especially to the up country of Sri Lanka. This research is to build a landslide prediction model based on Artificial Neural Networks. The prediction model then can be applied together with modern technology to focus more sophisticated events and accuracy level.

GIS data in Badulla district are analyzed in areas of people living in the landslide locations and identify the most causal factors for past landslide events. These factors are also used for preparation of landslide hazard maps. Landslide locations are identified by interpreting the GIS data and field survey data, and a spatial database of the topography, soil, forest, and land use. Then the landslide assessment factors are extracted from the spatial database. These factors are then used with an artificial neural network to analyze landslide susceptibility. Each factor's weight is determined by the back-propagation training method. Numerous training sets will be identified and applied to analyze and verify the effect of training. The landslide susceptibility index will be calculated by back propagation method and the susceptibility map will be created by GIS program. The results of the landslide susceptibility analyze are verified using landslide location data.

1.5 Scope of the Research

This research mainly focusses on, analyzing, mapping and processing of GIS and other related data of Badulla district by using several GIS tools and ANN to compute susceptible data for the preparation of Hazard Zonation Maps. However past landslide data from other districts are also considered as training data set of ANN.

As described above there are several environmental factors which are caused to trigger a landslide. Rainfall is the highest fluctuating factor for this event. This research has considered only the major factors (Hydrology, Land use, Land form) for conducting the study.

In this research GIS is used to analysis the considerable amount of data very efficiently and an Artificial Neural Network to be an effective tool to maintain precision and accuracy. Finally, the artificial neural network will prove it’s an effective tool for analyzing landslide susceptibility compared to the conventional method of landslide mapping.

This integrated research on landslide hazard zonation has produced a categorized (low, moderate, medium, high, very high) landslide hazard zonation map with higher accuracy. This methodology can be adopted for similar problematic areas and terrain conditions to assess the landslide susceptibility on large scale.

1.6 Structure of the Thesis

This thesis is formed in six chapters as stated below.

Chapter 01 – Introduction

This chapter encompasses the background of the research identifying the subject area of study, goals and objectives of carrying out the research, in light of giving the reader to gain an idea of the substance of the research project.

Chapter 02 – Literature Review

This chapter presents the literature study of related research works in the field of landslide prediction by using GIS and ANN. A comprehensive comparison of related studies, their advantages and drawbacks are discussed in this chapter.

Chapter 03 – Analysis and Methodology

This chapter includes analysis of data and the research methodology of the dissertation. This part describes research strategy, methods and the designed approach to the solution to the identified problem. It also describes data collection, the selection of the samples and identification of relevant factors.

Chapter 04 – Implementation

Provides implementation details and successful approach to the methodology and data analysis which is identified under Analysis and Methodology, chapter 03.

Chapter 05 – Evaluation

Provides details of evaluation approach to the results and outcomes of the research work.

Chapter 06 – Conclusion

This chapter provides conclusion for the research work, suggestions and identified future work to be performed.

CHAPTER 02

LITERATURE REVIEW

CHAPTER 02

LITERATURE REVIEW

The main goal of this research is to predict landslide susceptibility caused using the GIS analysis and modelling it to an Artificial Neural Network. Landslide occurrence study area is selected based on interpretations of GIS data. (A spatial database considering certain geographical factors). The above factors and related GIS data will be used to generate a landslide susceptibility map.

This research applied two statistical analysis methods in a GIS and compared their results. Based on the findings, we were able to derive a more effective method for analysing landslide susceptibility. The spatial prediction of landslide susceptibility applying artificial neural network.

This section explains about the existing similar systems in the world including the countries which are currently developing and countries which have currently been developed. A complete survey of literature was carried out in order to fulfil this requirement. Few researches have been selected among the others as they seemed to be the most relevant to this topic. Most of them have not considered antecedent rainfall or even rainfall as a factor when doing their researches. All the factors have been considered according to the environmental conditions of the particular countries which differ from country to country. For example: earthquakes, snowing.

2.2 Landslide Risk Factors (Geographical)

Identification and prediction of natural disaster lead to shut off a tragedy and to ensure the safety of lives and properties. For that purpose, collecting and analysing of accurate data is very important. Every so often manually collected data under very hard circumstances may not get as much of accurate and it directly affects to the final results hence the decisions making.

National Building Research Organization (NBRO) of Sri Lanka has identified six causative factors inducing landslides as [2] ,

(I) Bedrock geology
(II) Hydrology & drainage
(III) Surface overburden
(IV) Slope angle range
(V) Land use
(VI) Land forms

Hydrology & drainage and land use can be identified as major controllable and changing factors among above factors. These factors should be obtained time by time to generate the landslide susceptibility potential map. Thus, the necessity of a methodology to collect land use data by a remote sensing way is emerged. This paper introduces a methodology to generate a landslide susceptibility potential map with the aid of land use data collected by using remote sensing and GIS techniques.

Each causative factor is weighted according to the weight system used by NBRO. In this research, satellite imaginary is used as the main data resource and it is analysed by using GIS software to extract land use patterns. Those land use patterns are categorized and weighted under three categories as low, medium and high hazardous. Intercepting weighted land use shape file and six weighted shape files from other causative factors, landslide susceptibility potential map is generated. The effect of clouds to the analytical results is eliminated by observing the true colour satellite image.

2.2 Landslide Causal Factors

In every slope there are forces which tend to promote downslope movement and opposing forces which tend to resist movement.

A general definition of the factor of safety, F, of a slope results from comparing the downslope shear stress with the shear strength of the soil, along an assumed or known rupture surface. Starting from this general definition, divided landslide causes into external causes which result in an increase of the shearing stress (e.g. geometrical changes, unloading the slope toe, loading the slope crest, shocks and vibrations, drawdown, changes in water regime) and internal causes which result in a decrease of the shearing resistance (e.g. progressive failure, weathering, seepage erosion).

Mihail et al [5] illustrates major reasons which affects landslide formation.

- Ground conditions
- Geomorphological processes
- Physical processes
- Man-Made processes

However, Varnes [6] pointed out there are a number of external or internal causes which may be operating either to reduce the shearing resistance or to increase the shearing stress. There are also causes affecting simultaneously both terms of the factor of safety ratio. The great variety of slope movements reflects the diversity of conditions that cause the slope to become unstable and the processes that trigger the movement. Figure 2:1 shows factor of safety with changes of above factors and time.

It is more appropriate to discuss causal factors (including both “conditions” and “processes”) than “causes” per se alone. Thus, ground conditions (weak strength, sensitive fabric, degree of weathering and fracturing) are influential criteria but are not causes. They are part of the conditions necessary for an unstable slope to develop, to which must be added the environmental criteria of stress, pore water pressure and temperature. It does not matter if the ground is weak as such - failure will only occur as a result if there is an effective causal process which acts as well.

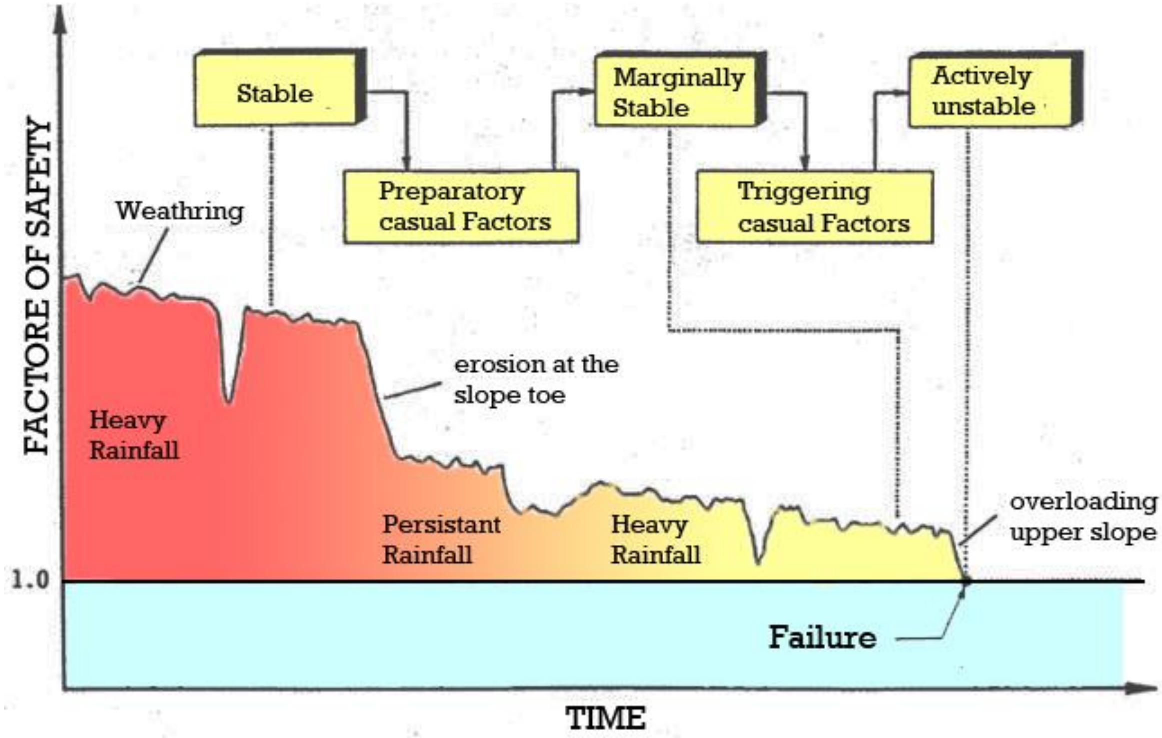


Figure 2:1: An example of changes in the factor of safety with time

This demonstrates that seldom, if ever, can a landslide be attributed to a single causal factor. The process leading to the development of the slide has its beginning with the formation of the rock itself, when its basic properties are determined and includes all the subsequent events of crustal movement, erosion and weathering.

The computed value of the factor of safety is a clear and simple distinction between stable and unstable slopes. However, from the physical point of view, it is better to visualize slopes existing in one of the following three stages: stable, marginally stable and actively unstable. [7]

Stable slopes are those where the margin of stability is sufficiently high to withstand all destabilizing forces. Marginally stable slopes are those which will fail at some time in response to the destabilizing forces attaining a certain level of activity. Finally, actively unstable slopes are those in which destabilizing forces produce continuous or intermittent movement.

2.3 Artificial Neural Network

An Artificial Neural Network (ANN) is a computational technique that was inspired by attempts to model the human central nervous system using human reasoning and problem-solving abilities. The design of this computational technique is to emulate biological neural networks in a similar way as the human brain learns tasks that are difficult to simulate with other logical and/or analytical techniques.

An ANN differs from other forms of computer intelligence because it is not rule based and doesn't need a predefined knowledge base. ANN is trained to learn associative patterns to recognize and generalize the relationship between a set of inputs and outputs. A well-trained ANN in some instances could be highly effective in discerning patterns that are difficult to detect otherwise. The ANN has been used in a number of applications in [8] such as,

- Image processing
- Pattern recognition
- Function approximation
- Optimization
- Forecasting
- Data retrieval
- Predicting purposes

The typical architecture of an ANN model consists of topology, a learning paradigm, and a learning algorithm. Network topology refers to the organization, the connection of the nodes, and the flow of data and error information between the layers, while the learning paradigm often handles two broad categories of supervised and unsupervised network learning using different learning algorithms.

The multilayered topology is the most common algorithm, where all processing units or the nodes are organized in three essential layers: input, hidden, and output. The input layer is used to pass predictive attribute data forward to the hidden layer which is connected to an output layer. Within this framework, the produced output is recursively compared with the desired output until error signal is minimized and computed output resembles the desired output.

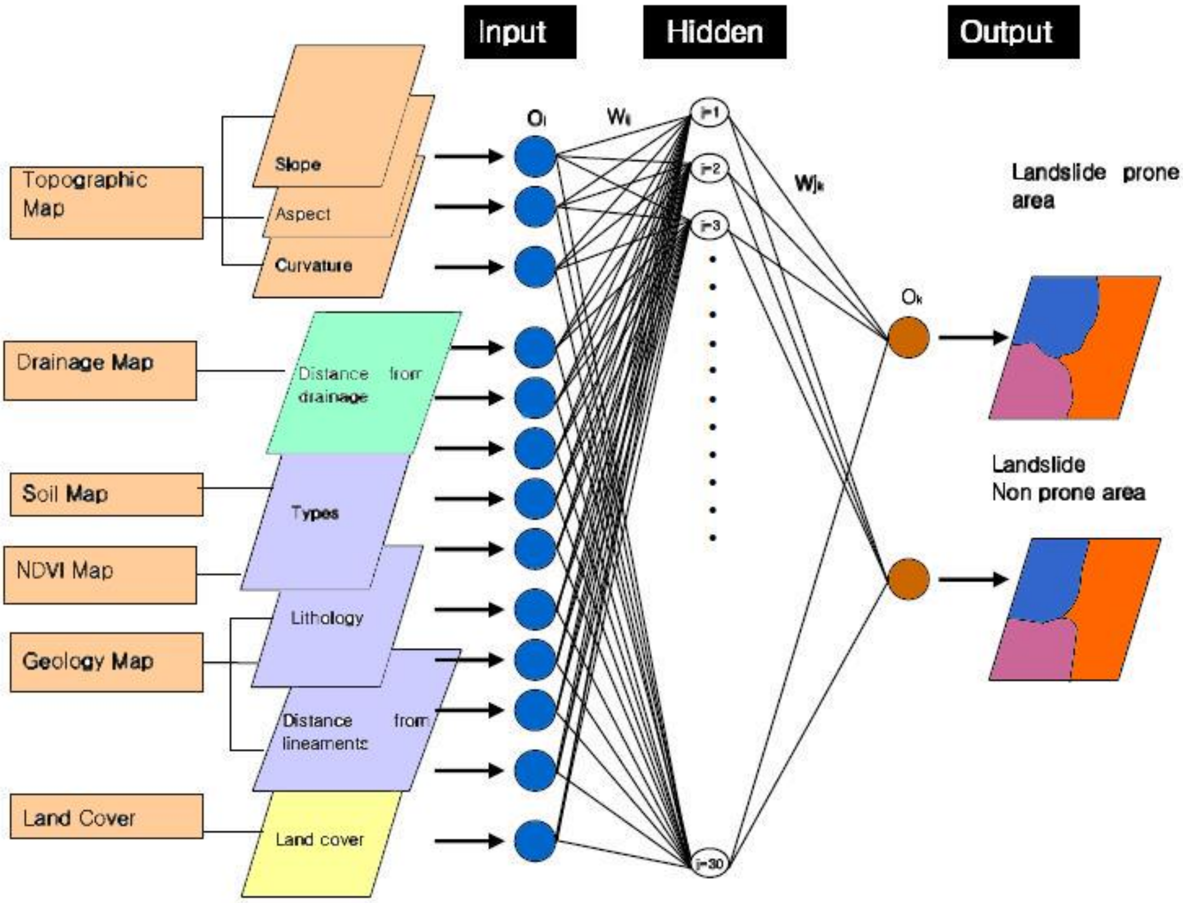


Figure 2:2: Structure of a landslide prediction Artificial Neural Network

Figure 2:2 shows the structure of the an ANN model for landslide prediction . The most common feed-forward neural networks are constructed by multilayer perceptron (MLP) and radial basis functions that use the back propagation (BP) algorithm Lippman et al [9]. In back propagation neural networks the data passes forward from input layers to output layers via the hidden layer(s). The initial output is produced for the input data and randomly assigned weights. The resulting outcome is compared with the desired output and error discrepancies are propagated backward through the network from the output to the input nodes. This is an iterative backward propagation intended to adjust the synaptic strength of weights to ensure similarity between the computed and desired output.

The purpose of an artificial neural network is to build a model of the data-generating process, so that the network can generalize and predict outputs from inputs that it has not previously seen. This learning algorithm is a multi-layered neural network which consists of an input layer, hidden layers, and an output layer. The hidden and output layer neurons process their inputs by multiplying each input by a corresponding weight, summing the product, and then processing the sum using a nonlinear transfer function to produce a result. An artificial neural network “learns” by adjusting the weights between the neurons in response to the errors between the actual output values and the target output values. At the end of this training phase, the neural network provides a model that should be able to predict a target value from a given input value. There are two stages involved in using neural networks for multi-source classification: the training stage, in which the internal weights are adjusted; and the classifying stage.

Typically, the back-propagation algorithm trains the network until some targeted minimal error is achieved between the desired and actual output values of the network. Once the training is complete, the network is used as a feed-forward structure to produce a classification for the entire data [10]. A neural network consists of a number of interconnected nodes. Each node is a simple processing element that responds to the weighted inputs it receives from other nodes.

2.4 Application of Artificial Neural Network for landslide disaster prediction

As described in the section ‘Artificial neural network’, the phases involved in the development of the MLP network are various and need to be carefully defined in order to build up the most successful model. Such steps are identified as follows:

(1) Data selection for the network training

(2) MLP design (input, hidden and output layers, and number of nodes)

(3) Network training (choosing activation and transfer functions)

(4) Classification.

(01) Data selection for the network training

The selection of a proper dataset for the network training is far from being obvious and represents one of the most critical steps, although not many works on landslide susceptibility have noted this. It is common to select a random subset of cells corresponding to a portion of the entire database, which includes both landslide and no landslide cells. However, in most studies, the selected subset rarely maintains the actual ratio between landslide and no landslide areas, and the percentage of landslide cells in the subset is often increased with respect to the total area. In particular, randomly selected a fixed number of cells from each of the two classes (landslide and no landslide), setting a ratio between landslide and no landslide cells equal to 1:1, despite the recorded ratio relative to the entire area being equal to 1:56. This study [11] analysed three different sizes of cell number, 200, 400, and 600 pixels/class, respectively, which corresponded to percentages of 1.7, 3.4, and 5.1% of the landslide cells and 0.03, 0.06, and 0.09% of the no landslide cells of each factor map. The study claimed that the number of training locations had little influence on the analysis work, the training phase was executed on a subset corresponding to one-third of the entire database, randomly selected from the whole dataset.

In this application, three different subsets were defined in order to pursue the best configuration. Both the methodological (one subset) and the random (two subsets) criteria have been applied. In the case of the methodological criterion, 10 kernels of 15 × 15 pixels were placed over the entire basin, selecting the pixels (landslide and no landslide) falling therein.

(02) MLP design

The definition of the MLP structure requires the definition of input, hidden and output layers, and number of nodes for each layer. The structure of the input vector depends on

(i) Number and type (continuous or categorical) of triggering factors

(ii) Methodology used in representing the data

Here the methodology presented in has been used, representing each variable as a sequence of binary numbers. First of all, the approach requires categorization of each landslide factor in classes; in the case of categorical factor (such as land use, pedology, etc.), the original classes were kept; in the case of continuous variable (e.g., slope, distance from road, etc.), a quantile-based classification was used. The number of classes for each variable is shown in Table 1, for a total number of 73 classes. At each computational cell (i.e., each cell of the basin), the input vector is represented by a string of binary values indicating whether the cell belongs (1) or not (0) to each of the 73 classes. Such a method, although increasing considerably the number of computational nodes, is capable of providing an efficient and objective approach.

3) Network training

Among all the back-propagation algorithms available in the literature, two of the most suitable to treat a large amount of data are the GDM (gradient descend with momentum) and the SCG (scaled conjugate gradient) algorithms. Both the algorithms were used in the simulations according to the configurations described in Table 1, in order to compare the performances of both the models. The chosen activation function is a sigmoid function (sgm), which returns values ranging from 0 to 1.

(4) Classification

In order to estimate the landslide susceptibility, all the landslide-inducing factors are fed into the designed MLP network. The network returns the susceptibility values at each cell grid on the basis of the weights found during the training phase. For each point, the relative position in the grid structure is recorded and used to reconstruct the susceptibility map.

2.5 Artificial Neural Network methods used for Landslide susceptibility prediction

Landslide locations were identified in the study area from the interpretation of aerial photographs, field survey data, a spatial database of the topography, soil type, timber cover, geology and land use.

The landslide-related factors (slope, aspect, curvature, topographic type, soil texture, soil material, soil drainage, soil effective thickness, timber type, timber age, and, timber diameter, timber density, geology and land use) were extracted from the database. Using those factors, landslide susceptibility was analysed by artificial neural network methods. Maps constructed in a vector format spatial database using the GIS software ARC/INFO were used for the application of ANN methods.

2.5.1 Landslide Deformation Prediction Based on Recurrent Neural Network

Landslide deformation prediction has significant practical value that can provide guidance for preventing the disaster and guarantee the safety of people's life and property.

In this paper [12], a method based on recurrent neural network (RNN) for landslide prediction is presented. Genetic algorithm is used to optimize the initial weights and biases of the network. The results show that the prediction accuracy of RNN model is much higher than the feedforward neural network model for Baishuihehe landslide. Therefore, the RNN model is an effective and feasible method to further improve accuracy for landslide displacement prediction.

The Elman's architecture is chosen for the time series prediction. The network consists of a context layer, an input layer, a hidden layer and an output layer. It has been shown theoretically that an Elman net with all feedback connections from the hidden layer to the context layer set to 1 can represent an arbitrary nth order system. Modified Elman net has been adopted which introduces self-feedback connections for the context units. This neural network also has four layers, with the main feedback connections taken from the hidden layer to the context layer. The modified Elman net has better dynamic capabilities than the original Elman network

The geology of the Cameron Highlands consists of mostly two types of litho types: igneous and metamorphic rocks.. Several field observations have been carried out in the month of April/ June/ September 2006 and 2007 for collecting ground data. The landuse in the study area is mainly peat swamp forest, plantation forest, inland forest, scrub, grassland and ex-mining area. The slope angle of the area ranges from 0 degrees to as much as 86 degrees. The relief of the study area varies between 860- 2110 msl.

Advantages & Limitations of this method

- Single layer can be used only for simple problems. However, its computation time is very fast.
- Recurrent ANN are used for dynamic systems mainly and for control systems since it is time-based and information might flow in any direction.

2.5.2 Landslide Susceptibility Analysis using an Artificial Neural Network (ANN) Model: A Case Study In Yushan National Park, Taiwan

This research [13] aims to perform landslide susceptibility analysis for the Yushan National Park (YNP) in central Taiwan based on an Artificial Neural Network (ANN) Model using Remote Sensing data and Geographical Information System (GIS). In recent years, Machine Learning (Artificial Intelligence) and Data Mining techniques have been introduced as efficient tools in hazard and susceptibility analysis. ANN is one of the commonly used not only because it can deal with complex and non-linear relationships between slope stability and conditioning factors, but also minimize subjectivity. To perform ANN analysis, besides the static (predisposing) factors of landslide occurrence including topographic slope, aspect, curvature, elevation, topographic index, distance to geological lineament, some dynamic (triggering) ones have been selected in this study such as vegetation index (NDVI) and precipitation (rainfall).

All factors are analysed with back – propagation training method to generate the landslide susceptibility map for the YNP. A landslide inventory map available for this study is used to validate the model. The results show where landslide is more likely to occur and highlight important factors that can explain the slope stability in YNP. Finally, this work can be used as a reference to assist slope failure, slope management and tourism planning considering landslide susceptibility in YNP.

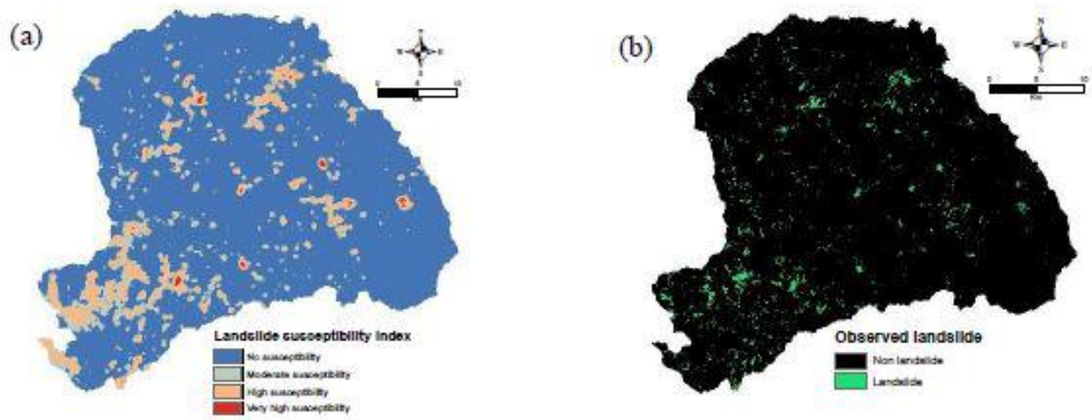


Figure 2:3: ANN modelling result: (a) Landslide susceptibility map (b) Observed Landslide

Figure 2:3 shows a visual comparison between a landslide susceptibility map generated from an ANN model and actual observed landslide map. Landslide susceptibility maps play a vital role in assisting and managing hazards for land use planning and risk mitigation. LSM provide information on the likelihood of landslides occurring in an area given the local terrain conditions. Using GIS, various methods for landslide susceptibility mapping have been proposed in the past. These methods can be grouped into qualitative and quantitative, based on the properties.

Advantages & Limitations of this method

- The selected factors are sufficient to qualitatively and quantitatively model the relationship between landslide occurrences.
- Major variations in some areas of Landslide susceptibility map and observed landslide map. Therefore, recommend to modify some calculated factors.

2.5.3 An integrated artificial neural network model for the landslide susceptibility assessment of Osado Island, Japan

The objective of this study [14] is to select the maximum number of correlated factors with landslide occurrence for slope-instability mapping and assess landslide susceptibility on Osado Island, Niigata Prefecture, Central Japan, integrating two techniques, namely certainty factor (CF) and artificial neural network (ANN), in a geographic information system (GIS) environment.

The landslide inventory data of the National Research Institute for Earth Science and Disaster Prevention (NIED) and a 10-m digital elevation model (DEM) from the Geographical Survey of Institute, Japan, were being analysed. This study identified fourteen possible landslide-conditioning factors. Considering the spatial autocorrelation and factor redundancy, applied the CF approach to optimize these set of variables. The study hypothesize that if the thematic factor layers of the CF values are positive, it implies that these conditioning factors have a correlation with the landslide occurrence.

The applied approach indicates that the values of the AUC at optimized and non-optimized BPNN were 0.82 and 0.73, respectively. Hence, it is concluded that the optimized factor model can provide superior accuracy in the prediction of landslide susceptibility in the study area. In this context, this research paper propose a method to select the factors with landslide occurrence. This work is fundamental for further study of the landslide susceptibility evaluation and prediction.

Advantages & Limitations of this method

- More geographical factors including the slope aspect from DEM, are considered and therefore the resultant data is capable of presenting the direction of landslide.

2.5.4 A neural network model applied to landslide susceptibility analysis (Capitanejo, Colombia)

In this research [15] , the analysis of the susceptibility is generated by grouping the 14 variables created from the attributes of geology and geomorphology; this process is done by integrating all the inputs in a calculation matrix under a processing module, where, on the one hand, the variables are generated and on the other, there is data for learning.

Analysis is carried out using the neural network (ANN), being an information processing system that allows for the solution of complex problems of classification, functional estimation and optimization through the estimation of a linear or nonlinear relationship between the input and output data.

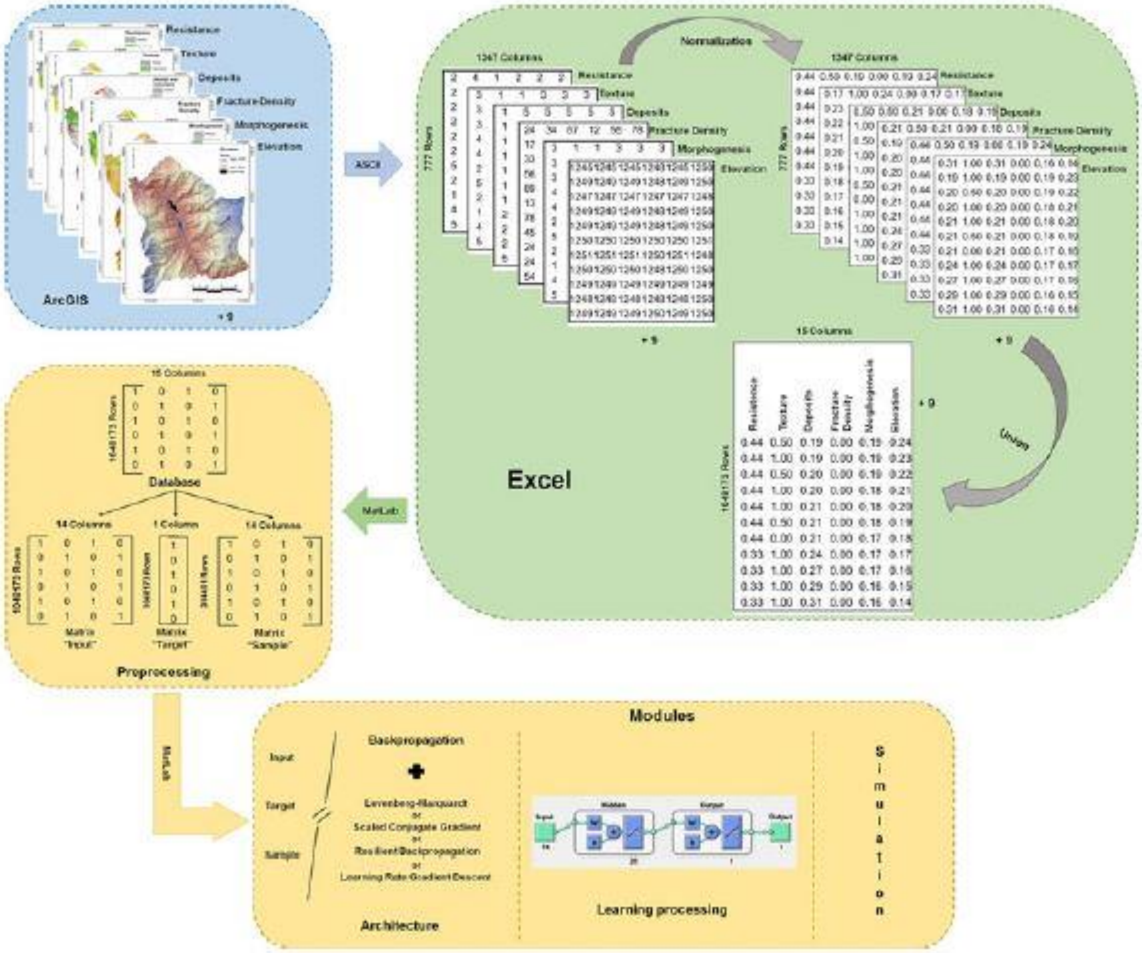


Figure 2:4: Processing structure of the ANN

This analysis requires three steps. Firstly, there is the generation of a database with the created factors of geology and geomorphology. Secondly, there is the creation of the matrices (input, target and sample) and the lastly, there is the establishment of training parameters for the simulation within the ANN. The database is composed of the following variables: resistance, texture, surface formations, fracture density, morphogenetic, elevation, aspect, curvature, distance to rivers, distance to roads, slope, rugosity, SPI, TWI and landslide (50% for training). These data were taken randomly so as not to establish a predisposition in the calculation performed by the ANN as shown in Figure 2:4.

Advantages & Limitations of this method

- This model predicted 92.86% of the data that were not entered in the learning module, which represents 50% of the landslides mapped in the study region

2.6 Other methods used for Landslide prediction

Landslide prediction methods can be classified into three types: Image analysis, machine learning, and mathematical evaluation models. Table 2:1 shows a comparison among these types of methods. First, image analysis uses Geographic Information Systems (GIS), which can collect, store, manage, and analyse geographical data. By analysing disaster data, such as history of landslides and data on land development for agriculture, the risk of landslides can be predicted.

The probability of landslides is variable, as it is based on the number of layers of data used for analysis. Second, machine learning techniques, such as Bayesian networks [16], neural networks, or genetic algorithm, use computational intelligence to calculate the probability of landslides. These methods incorporate different factors that might cause landslides to evaluate the probability of landslide occurrence. This is not real-time, because they require huge computational times for prediction. Finally, mathematical evaluation models use a single evaluate equation, such as Factor of safety. A hazard model is combined with the physical concepts of mechanics and hydrographic data for the stability of slopes. It is easy for simulation and fits a wide range of environments, but it is difficult to obtain the whole hydrographic data as groundwater elevation is difficult to measure.

Types	Methods	Advantages	Accuracies
Image Analysis	Geographical Information System(4)	Suitable for large Area	Accuracy based on number of layers
Machine Learning	Bayesian Network(2)	Simple Network	75%
	Neural Network(3)	Simple Network	67%
	Genetic Algorithm(4)	Optimal Solution	90%
Mathematical Evaluation	SHASLTAB(6)	High Accuracy	>90%

Table 2:1: Landslide prediction method types

Landslide Susceptibility Evaluation and Factor Effect Analysis

Landslide Susceptibility Evaluation and Factor Effect Analysis were done by using Probabilistic-Frequency Ratio Model [17]. Landslide location map has been generated on the basis of image elements interpretation from aerial photos, satellite data and field observations. Display, manipulate and analysis have been carried out to evaluate layers such as geology, geomorphology, slope, soil, land use, distance from roads and drainages. The area under the prediction rate curve, evaluates how well the method predicts landslides. The results showed satisfactory agreement between prepared susceptibility map and existing data on landslide locations (92.59%).

Probabilistic model Landslides Hazard Mapping

Chung and Fabbri [17] introduced a probabilistic model Landslides Hazard Mapping. The model A use a spatial database for predictive modelling (i.e., with all the landslide characteristics, including topographic, geotechnical, geological, infrastructural, and temporal settings) must be built so that each information layer clearly contributes to the characterization of the typical setting of one event to be predicted. been performed, the statistical results, showing the frequency distribution of the occurrences of the past landslides with respect to the supporting pieces of evidence, should be reviewed by experts. For the least-squares estimation of the regression coefficients in Equations 17 and 21, we have also studied several different training data sets ranging from about 1,000 pixels to the whole study area (43,7019 pixels) using the weighted least-squares estimation. The prediction rates with respect to the size of the training data set appear to be robust and the study on the effect will be a subject to a future contribution.

Landslide Prediction using Wireless Sensor Networks

Kalyana et al [11] came up with a Routing Protocols for Landslide Prediction using Wireless Sensor Networks. Landslide prediction and early warning system is an important application where sensor networks can be deployed to minimize loss of life and property. Due to the dense deployment of sensors in landslide prone areas, clustering is an efficient approach to reduce redundant communication from co-located sensors. In this research they propose two distributed clustering and multi-hop routing protocols, CAMP and HBVR, for this problem. While CAMP is a new clustering and routing protocol, HBVR is an enhancement of BVR with HEED. We further enhance CAMP and HBVR with TEEN, a threshold based event driven protocol. TEEN is most suitable protocol for this application since different rock types can have different thresholds for stress values. Simulation results show that CAMP- TEEN gives the best performance with respect to network life time and energy consumption.

Out of these methods the Artificial Neural Network model was selected since a very high uncertainty is involved in predicting landslides, a Neural Network is a better solution as it handles uncertainty to a very high degree for the prediction of Landslide Disasters as a dynamic prediction model.

Since a very high uncertainty is involved in predicting landslides, an ANN is a better solution as it handles uncertainty to a very high degree to prediction of Landslide Disasters as a dynamic prediction model. This is an effective solution when it is difficult to build the relationship. Neural Network can process simultaneously qualitative and quantitative data. Therefore, a Neural Network is a good candidate to establish these models. The benefits of integrating GIS and artificial neural network are efficiency and ease of management, input, display and analysis of spatial data for landslide susceptibility analysis.

CHAPTER 03

ANALYSIS AND METHODOLOGY

CHAPTER 03

ANALYSIS AND METHODOLOGY

3.1 Research Methodology

There have been various studies conducted on landslide mitigation, which is one of the threatening problems in developing nations like Sri Lanka where the population explosion requires more judicious practices for landslide management.

In this context, a review on works of researchers has been discussed in the literature review chapter. The use of GIS/remote sensing data and its applications in identifying the content of hazard zones on regional scale using GIS is widely established over the past years.

Research methodology was proposed to achieve the objectives of the research in the best possible results. The methodology and flow path of process of the research is designed based on related techniques and methods which are identified in literature study described in chapter 02.

Following steps should be followed for the purpose of data analysis development of susceptibility maps. The eventual target is to build a successful landslide susceptibility model as HZM using Artificial Neural Network with use of GIS technologies.

1. **Collecting data from field survey (where necessary) and related organisations.**
   This involves collecting data for to perform necessary tasks to identify factors and create GIS database of the selected study area (Badulla District)

2. **Investigate and find the main factors associate with landslides.**

   This process is identification of appropriate factors and conditions data collected in previous step.

3. **Acquire factors which are identified from existing maps and using past data.**

   This step includes collecting, analysing and grouping data values of the identified factors, specially from historical data. Special data analyse tools and software s are used to perform this task.

4. **Analyse data using GIS and preparation of Landslides Susceptibility Index.**

   GIS data (Ground Elevations, DEM, Point value collections) are analysed and structured in to a spatial database. ESRI ArcGIS and QGIS softwares are used to build the database.

Figure 3:1 flow diagram represents main processes and flow of data throughout the research. Some process steps are done during the implementation phase (See Chapter 4) of the research. The Figure 3:1 depicts only the main steps of the process and there may be some alteration at the time of implementation phase.

Research Methodology

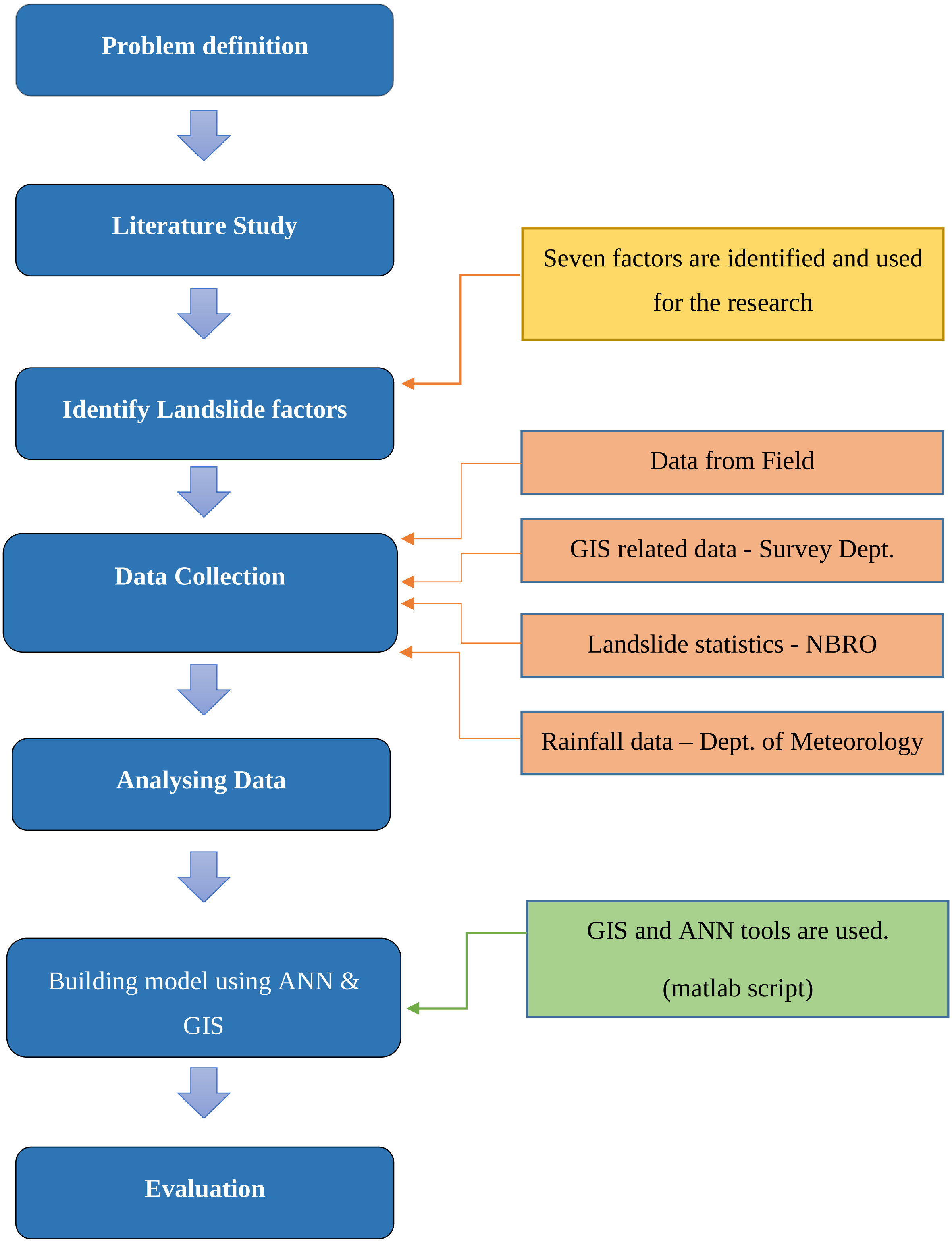


Figure 3:1: Outline of research methodology

3.2 Study Area

Badulla district, which has suffered much landslide damages following heavy rains [18], was selected as a suitable pilot area as study area to evaluate landslide risk analysis.

Badulla district (Figure 3:2) is located within latitudes of 7° and Longitudes 81.25°. The entire land area of the Badulla district is 2,861 km² and total population is 837,000. The district is bounded by the districts of Monaragala and Rathnapura on the East & South, by Ampara and Kandy districts on the North and by Nuwara Eliya and Matale on the West. Mainly the economy of the district is based on agricultural farming and livestock. [19]

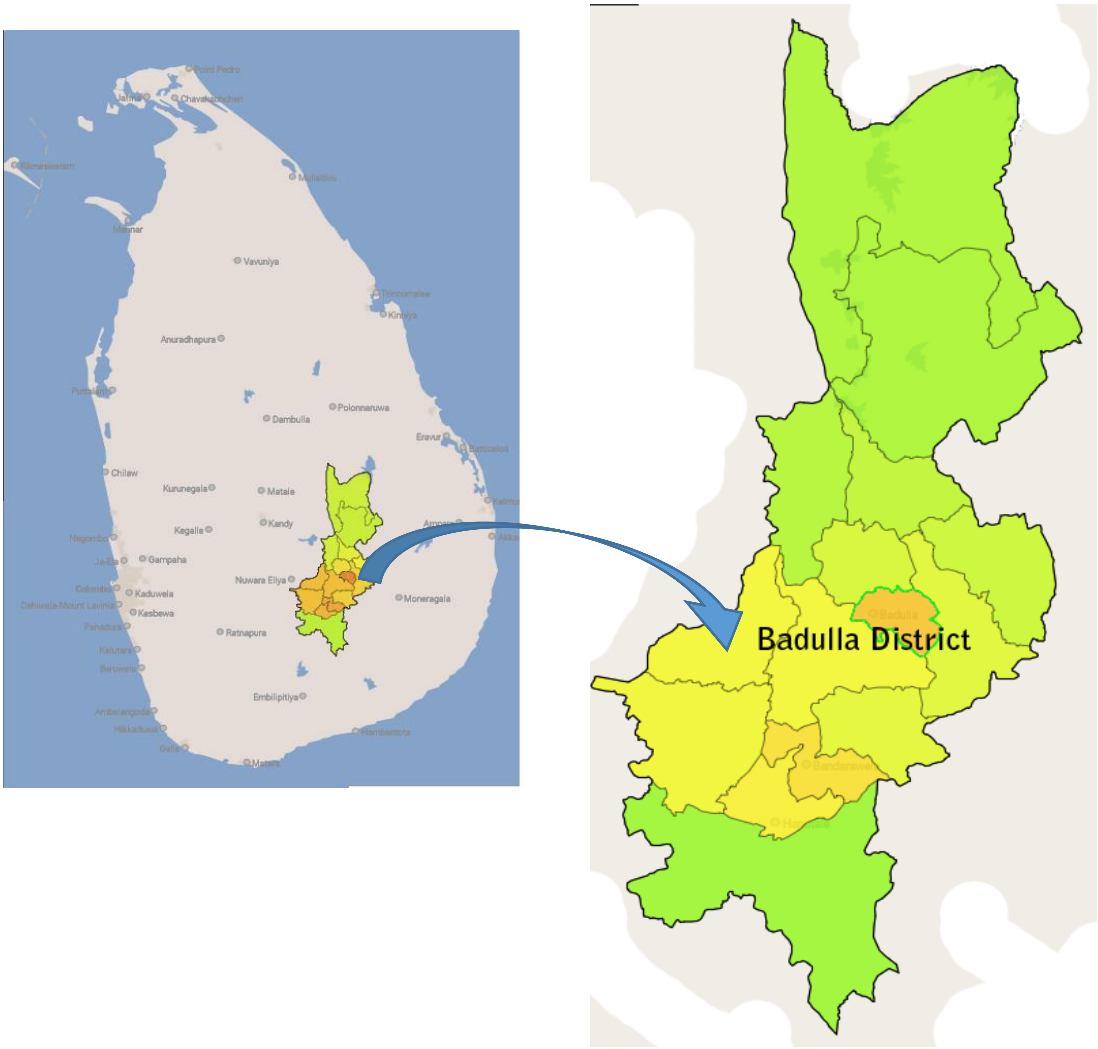


Figure 3:2: Badulla District

3.3 Data Acquisition

Table 3.1 shows all collected data and maps which used for this research. Some data had been rearranged and analyzed in order to suit the proposed model. Topographical data collected from Survey department and GIS data Collected from NBRO are merged to obtain optimal geographical output.

Data	Data Type	Source
Topographical Maps **(1:10000, 1:50000)**	GIS (Shapefiles)	Survey Department, NBRO
Land Use Map **(1:50000)**	Landuse (2015/2017/2019/2020) Detail Type of land use in previous years, occurred time and current	Derived from Satellite image (Landsat 8) analysis USGS (30m x30m resolution) Google earth
Ground Elevation Maps **(1:10000, 1:50000)**	Satellite image files (.TIFF format)	Landsat 8 USGS, True color full ALPSMLC30_N007E081_DSM satellite images
Soil type map	Satellite image files (.BILL format)	USGS 30as IGBPSOILCD_565 satellite images
Hydrology Maps	Shapefiles, JPEG image files	Survey department Internet

Rainfall data	Data Sheet	Department of Meteorology PERSIANN Cloud Classification System (PERSIANN CCS) is a real-time global high resolution (0.04 x 0.04 or 4km x 4km daily
Landslide data	GIS & Data Sheet	National Building Research Organization
Administrative Map	GIS (Shapefiles)	Survey Department

Table 3:1: Factor data sources

The above collected data are preprocessed and rearranged in order to extract the landslide related factors which are used to further computations. There were some restrictions due to technical issues (data format, resolution, lack of metadata etc..) and restrictions according to agreements and the conditions rested by the data issued organizations.

The following satellite image services are used to gather images and acquire data.

- USGS Earth Explorer
- Sentinel Open Access Hub
- NASA Earthdata Search
- DigitalGlobe Open Data Program
- Globcover

Landslide factors identified in the literature study there are several factors which may affect landslides. Some factors are constant and some are varying continuously. Therefore, these two categories are should be further analyzes before map them in to a prediction model. Accordingly, suitable weight values for varying factors are also to be computed before building the model. Based on the states factors can be further categorized as internal and external factors. [20] . Table 3:2 shows internal and external landslide factors identified so far,

Internal Factors	**External Factors**
• Geological structure	• Hydrology
• Soil overburden	• Building and other constructions
• Land slopes	• Earthworks
• Land use	
• Landform	
• Drainage system	

Table 3:2: Landslide Internal & External factors

A knowledge of the general setting is essential in the recognition of either potential or actual landslides. Similar geologic and soil conditions tend to give rise to similar landslides and recognition features. But only under constant climatic conditions.

Even though modified, the slump blocks are still discernible; the depressions between them are disconnected and there is no well-defined drainage system. In regions of heavy rainfall, however, and in the same amount of elapsed time, the topographic expression of identical geologic conditions is entirely different. [21]

3.4 Factor data layers

The following landslide related factors were selected by using the above data analysis and other landslide factor identification techniques. These seven factors would be used to create landslide inventory and further processed data would be used to ANN prediction model which is described in Chapter 04 in detail.

Figure 3:3 shows factor data layers selected for the research study. Rainfall is the most varying and triggering factor for most of landslides. The other factors are related to shape and the content of the terrain and they are considered as static when compared to rainfall factor. These factors take very long time period to change when compared to rainfall.

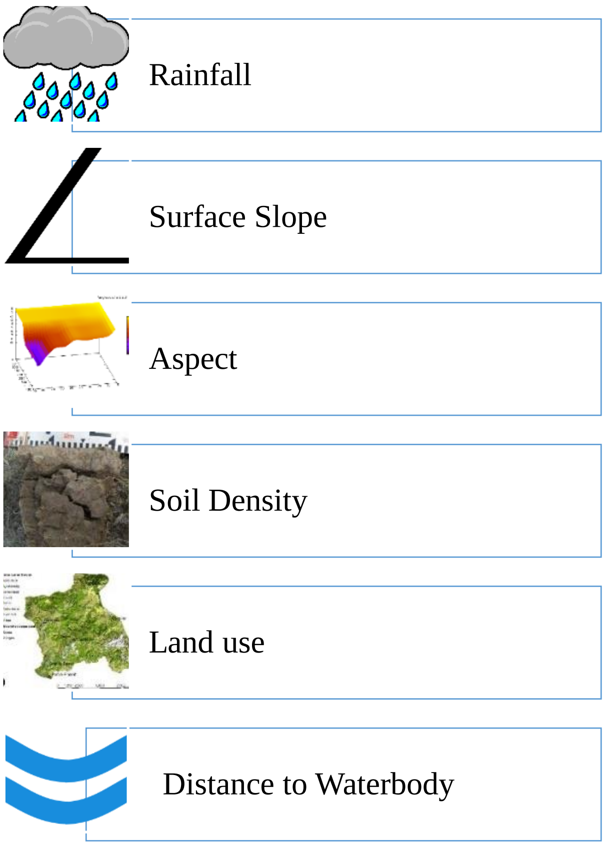


Figure 3:3: Factor data layers

3.5 Artificial Neural Network for Landslide prediction

The network is trained using the data taken from the geospatial database. MATLAB R2015a is used as the supporting software to develop ANN and train the network. And also to generate testing reports.

Matlab supports developing different types of neural networks like classification, distribution, fitting and pattern recognition. Pattern recognition tool is used to develop the neural network to predict landslide events of this research. The neural network contains 07 neurons in input layer, 01 hidden layer with 10 neurons, 02 outputs in the output layer. See Figure 3:4.

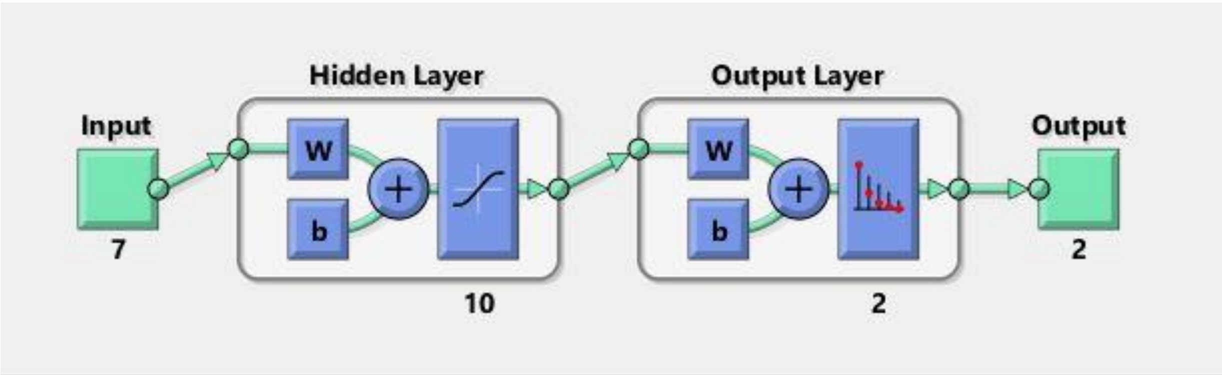


Figure 3:4 : Structure of the Neural Network

The input layer input the data to the network for training/ prediction purposes. Data contains 07 landslide factor data (LSHZ Range Val, Rainfall, Surface Slope, Aspect, Soil Density, Land use, Distance to Waterbody) This will be further discussed in Chapter 04 (Implementation Chapter)

Then initial weights for each node is randomly assigned. These weights and bias are adjusted according to the training process. Each node is a simple processing element that responds to the weighted inputs it receives from other nodes. The arrangement of the nodes is referred to as the network architecture. The receiving node sums the weighted signals from all the nodes that it is connected to in the preceding layer. Formally, the input that a single node receives is weighted according to Equation 3:1.

$$net_J = \sum_i w_{ij} \; . O_i$$

Equation 3:1: Calculation of input for each node

“MATLAB is a programming language developed by MathWorks. It started out as a matrix programming language where linear algebra programming was simple. It can be run both under interactive sessions and as a batch job.” [22] Matlab scripting language is used to create the network modify and adjust the variables/parameters. Sample set of the script is shown in Figure 3:5. Complete script for the ANN network can be found in Appendix B.

```
x = input;
t = target;

% Training Function
% trainlm , trainbr , trainscg , traingdx - with learnin rate
trainFcn = 'trainscg';  % Scaled conjugate gradient backpropagation

% Creating Nural Network
hiddenLayerSize = 12;
net = patternnet(hiddenLayerSize);

% Input and Output Pre/Post-Processing Functions
net.input.processFcns = {'removeconstantrows','mapminmax'};
net.output.processFcns = {'removeconstantrows','mapminmax'};


net.divideFcn = 'dividerand';  % Divide data randomly
net.divideMode = 'sample';  % Divide up every sample
net.divideParam.trainRatio = 70/100;
net.divideParam.valRatio = 15/100;
net.divideParam.testRatio = 15/100;


% Performance Function
net.performFcn = 'crossentropy';  % Cross-Entropy
% net.performFcn = 'msereg';  % Mean Aquare Error
```

Figure 3:5: Sample script of the ANN

3.5.1 Architecture of the Artificial Neural Network

Architecture of the developed Neural network for landslide prediction is shown in Figure 3:6 (FCNN style). This consists of Input layer, one hidden layer and output layer with two outputs.

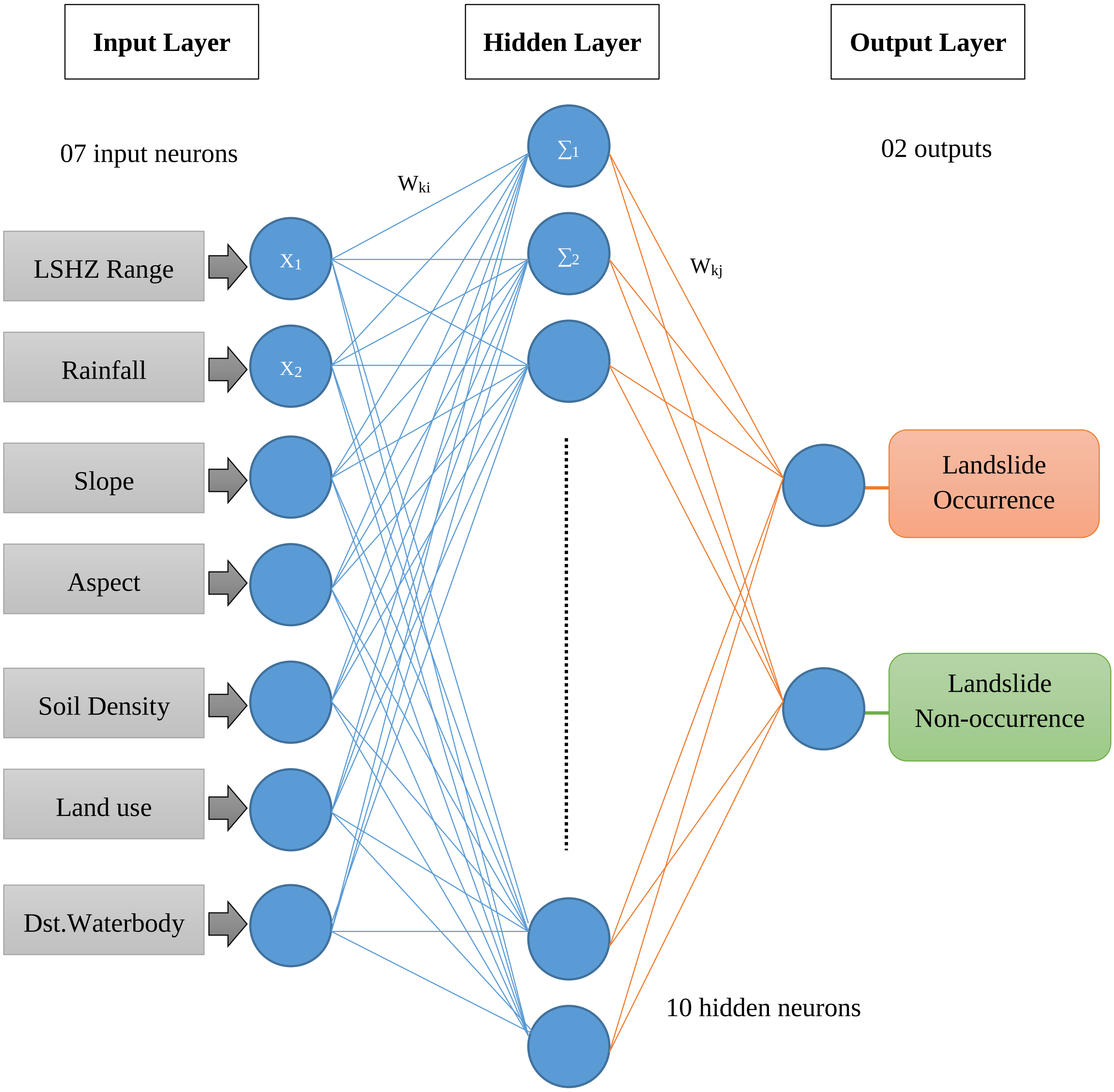


Figure 3:6: Architecture of the Artificial Neural Network

3.6 Landslide Hazard Zonation

Landslide hazard is commonly shown on maps, which display the spatial distribution of hazard classes (Landslide Hazard Zonation). Landslide hazard zonation refers to "the division of the land in homogeneous areas or domains and their ranking according to degrees of actual / potential hazard caused by mass movement" [23].

Landslide failures have caused untold number of causalities and huge economic losses. In many countries, economic losses due to landslides are great and apparently are growing as development expands into unstable hillside areas under the pressure of expanding populations. In spite of improvements in recognition, prediction, and mitigation measures, worldwide landslide activity is increasing.

3.6.1 Methods for Landslide Hazard Zonation

Direct method: Consists of Geomorphological mapping where the earth scientist evaluates the direct relationship between the hazard and the environmental setting during the survey.

Indirect method: This method includes two different approaches, namely the heuristic (knowledge driven) and Statistical techniques.

Figure 3.4 shows direct and indirect approaches for Landslide Hazard Zonation

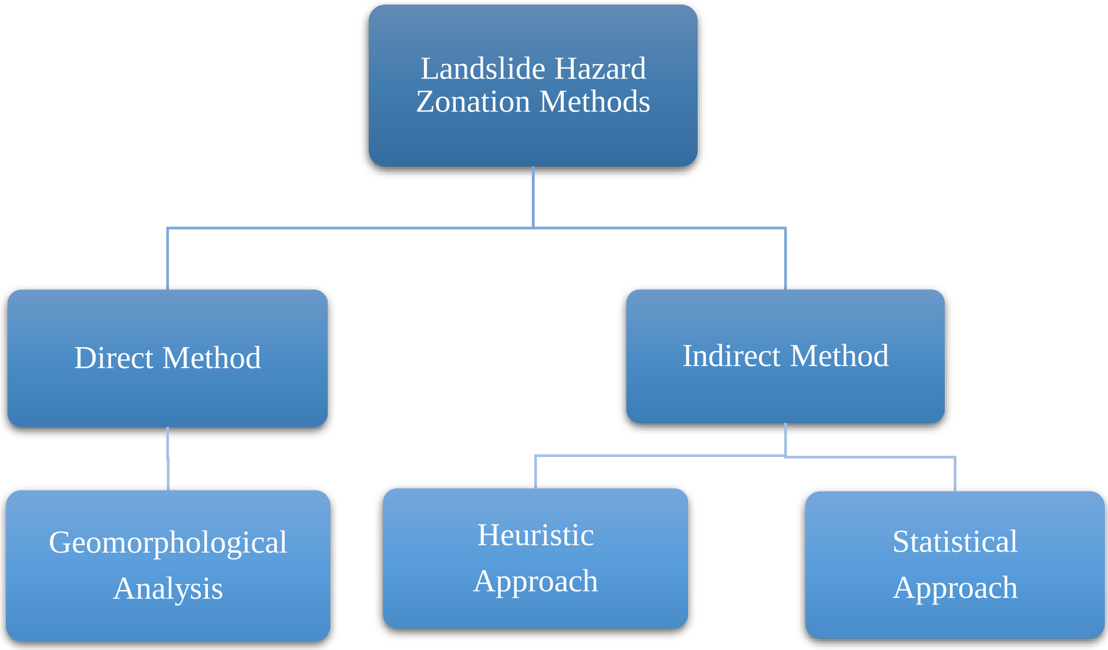


Figure 3:7: Approaches for LHZ

3.6.2 Geospatial Database

To apply the artificial neural network for landslide susceptibility mapping, a spatial database is created which consists of topography, soil, forest, landuse, rainfall analysis data etc., into consideration. Landslide occurrence areas are detected from satellite images and source data. Both the calculated and extracted factors are converted to form a grid (ARC/INFO grid type), and then it is converted to ASCII data for use with the artificial neural networks program. Then the back-propagation algorithm is used for training the network.

Evaluation factors for landslide hazard study are assessed through a detailed analysis of terrain characteristics such as geology, geomorphology, soil, slope, aspect, rainfall, data and landuse/land cover which are inter-related and interdependent and play a major role in determining the areas susceptible to landslide hazard.

CHAPTER 04

IMPLEMENTATION

CHAPTER 04

IMPLEMENTATION

4.1 Implementation of the model

The implementation process consists of eight main steps which are used to generate a landslide susceptibility prediction.

1. Data preprocessing
2. Generating Factor maps
    I. Generating Rainfall map
    II. Generating Slope map
    III. Generating Surface Aspect map
    IV. Generating Landuse map
    V. Generating Soil Density Map
    VI. Calculating distance to waterbodies.
3. Data conversion and Rectification
4. Extracting factor data
5. Creating geospatial database

6. ANN training process

7. Comparing Results

Figure 4:1 shows the simplified model of obtaining landslide susceptibility results by processing factors and applying machine learning approach. The landslide database is then used to network training, verifying, and for evaluation purposes. Finally, this database (geospatial database) is converted in to a landslide susceptibility map for a given time frame.

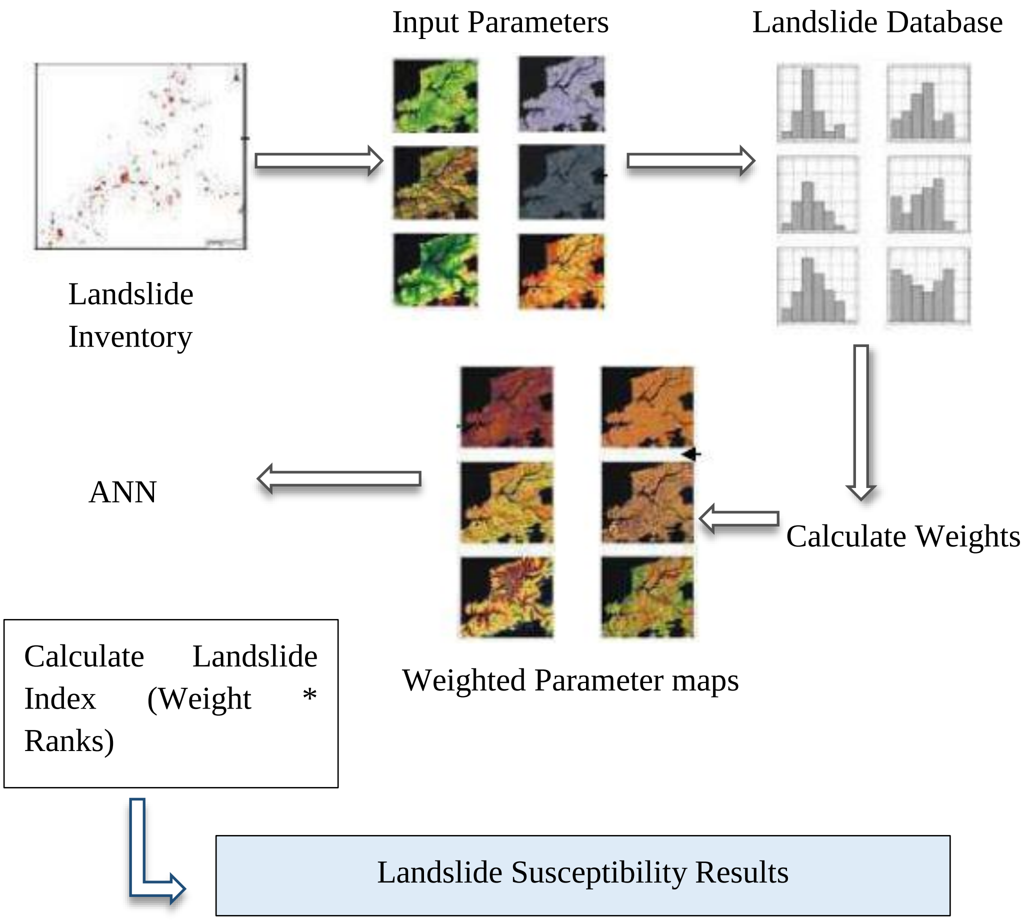


Figure 4:1: Landslide data process model

4.1.1 Data pre-processing

This is the process of pre-processing data (GIS, Raster) before performing any GIS analysis. This includes following processing tasks and further data types and completeness of data is checked before any major GIS operations.

1. **Data conversion** – Collected data should be in proper format (shapefiles, size etc..). All data is stored in an ESRI geodatabase

2. **Projection** – Since lot of GIS operations are performed for data, everything should be in same correct spatial projection system. Therefore, all data are transformed in to WGS84 coordinate system.

3. **Topology** – Topology of data is checked in order to ensure the completeness of spatial data (maps)

4. **Clipping** – Any unwanted data are clipped out in order to keep everything in a clean spatial environment.

5. **Merging data** –Multiple source satellite images (.tiff) are merged together to get more detail picture.

4.1.2 Generating Factor maps

This step includes generation all factor data maps which would be used to extract point data. Most factor data are extracted by processing satellite images. Satellite images are downloaded from internet and obtained from Survey Department.

The main problem arises with factor maps generation is, unavailability of data for required locations. (E.g.: Department of Metrology collect rainfall levels for fixed locations as point data) Therefor several geoprocessing techniques are used to generate and compute data for required locations. These techniques include data interpolation, merging, union and reclass.

All IS operations are performed as geodatabase feature classes and then converted back in to raster format for raster operations.

I. Generating Rainfall map

Precipitation is the most significant and the triggering factor when considering landslides. Therefore, it is an essential step to generate more detail rainfall map to extract rainfall levels when analysing landslides.

Rainfall information are collected from Department of Meteorology and satellite images are analysed and merged to create date wise rainfall maps for each past landslide event, since each landslide is taken place in different dates and different environmental conditions. Department of Meteorology collect rainfall data at their fixed locations in all over the country. Rainfall for each location (Station ID) is shown with their coordinate information. Table 4:1 shows sample dataset of rainfall datasheet for several locations in Badulla District.

Date	Datatype	Station ID	value	fl_miss	fl_cmiss	dd	obs_val
2010-01-01T00:00:00	TPCP	GHCND:CE000434730	746	0	0	1	34
2010-01-01T00:00:00	TPCP	GHCND:CEM00043418	474	0	0	2	0
2010-02-01T00:00:00	TPCP	GHCND:CEM00043418	38	0	0	7	0
2010-02-01T00:00:00	TPCP	GHCND:CEM00043424	10	0	0	8	0
2010-02-01T00:00:00	TPCP	GHCND:CEM00043441	15	0	0	9	0
2010-02-01T00:00:00	TPCP	GHCND:CEM00043466	54	0	0	10	1.6
2010-03-01T00:00:00	TPCP	GHCND:CEM00043441	1657	0	0	14	8.8
2010-03-01T00:00:00	TPCP	GHCND:CEM00043466	304	0	0	15	0
2010-04-01T00:00:00	TPCP	GHCND:CE000434730	1481	0	0	16	15
2010-04-01T00:00:00	TPCP	GHCND:CEM00043418	20	0	0	17	4.1
2010-04-01T00:00:00	TPCP	GHCND:CEM00043424	1317	0	0	18	0
2010-04-01T00:00:00	TPCP	GHCND:CEM00043441	2597	0	0	19	1.6
2010-04-01T00:00:00	TPCP	GHCND:CEM00043466	4413	0	0	20	8
2010-05-01T00:00:00	TPCP	GHCND:CEM00043466	7874	0	0	25	0
2010-06-01T00:00:00	TPCP	GHCND:CE000434730	3131	0	0	26	24.5
2010-06-01T00:00:00	TPCP	GHCND:CEM00043418	193	0	0	27	0

Table 4:1: Rainfall Datasheet (Source - DMET)

Each landslide event is occurred in different dates and times, Therefor separate rainfall map should be generated for each event. The following process shows generation of rainfall map for the landslide event which occurred in December 2006 at Batawatta Group, Division 4 in Badulla district. The above event is taken as an example to describe the map creation process.

ArcGIS for desktop (10.5) is used as the GIS tool to analyse data and generate all factor maps including rainfall detail maps. All rainfall values from all collection points should be loaded in to the map, since there might be no collection point(s) at the exact event location. (Figure 4:2) Then values from the relevant datasheet are merged to the location attribute table as a table join as shown in Figure 4:3.

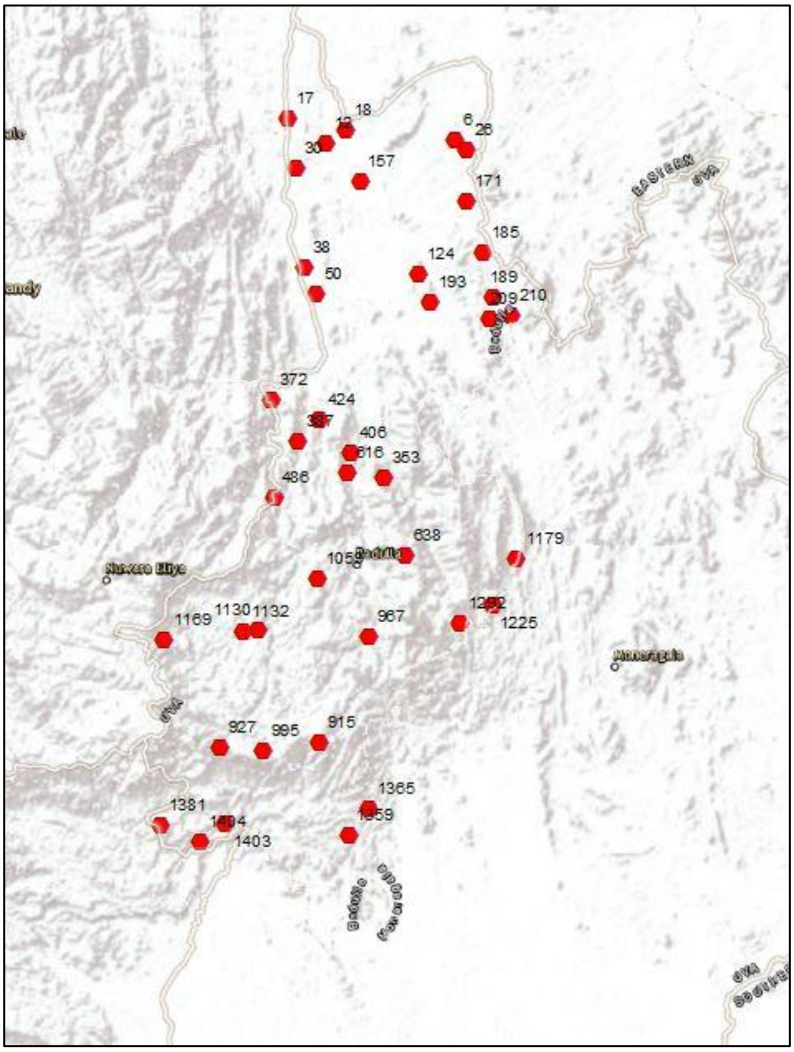


Figure 4:2 : Rainfall collecting points map (Badulla District)

OBJECTID_1 *	Shape *	RFC_Point	RF_Level
6	Point	Muthugalawela	75
12	Point	Girandurukotte	86
17	Point	Ginnoruwa	103
18	Point	Pahala Ratkinda	46
26	Point	Weerapokuna	317
30	Point	Aluttarama	177
38	Point	Puhulpola	297
50	Point	Model Village	352
124	Point	Eldeniya	69
157	Point	Cashew Farm	269
171	Point	Padukumbura	28
185	Point	Puskivula	34
189	Point	Gonabokka	113

Figure 4:3: Location and value merged data

This is a rainfall collection point map. The data values can be obtained only from the specified locations. But landslide occurred points are rested in different locations. Therefor point data interpolation method is applied in order to obtain data values for the required locations. “Interpolation predicts values for cells in a raster from a limited number of sample data points. It can be used to predict unknown values for any geographic point data such as elevation, rainfall and so on” [24] Following point interpolation methods are supported by ArcGIS software,

- IDW
- Kriging
- Natural neighbour

In order to select the best interpolation method for rainfall, analysis for the output is performed by calculating Root Mean Square Error (RMSE) by using the Equation 4:1 [25]

$$RMSE_{fo} = \left[\sum_{i=1}^{N} \left(z_{fi} - Z_{oi} \right)^{2} \; / \; N \right]^{\frac{1}{2}}$$

Equation 4:1 : RMSE

Interpolation Method	IDW	Kriging	Natural neighbour
SD	120.9	123.7	117.4
RMSE (+-)	2.31	2.43	2.48

Table 4:2: Interpolation Method analysis

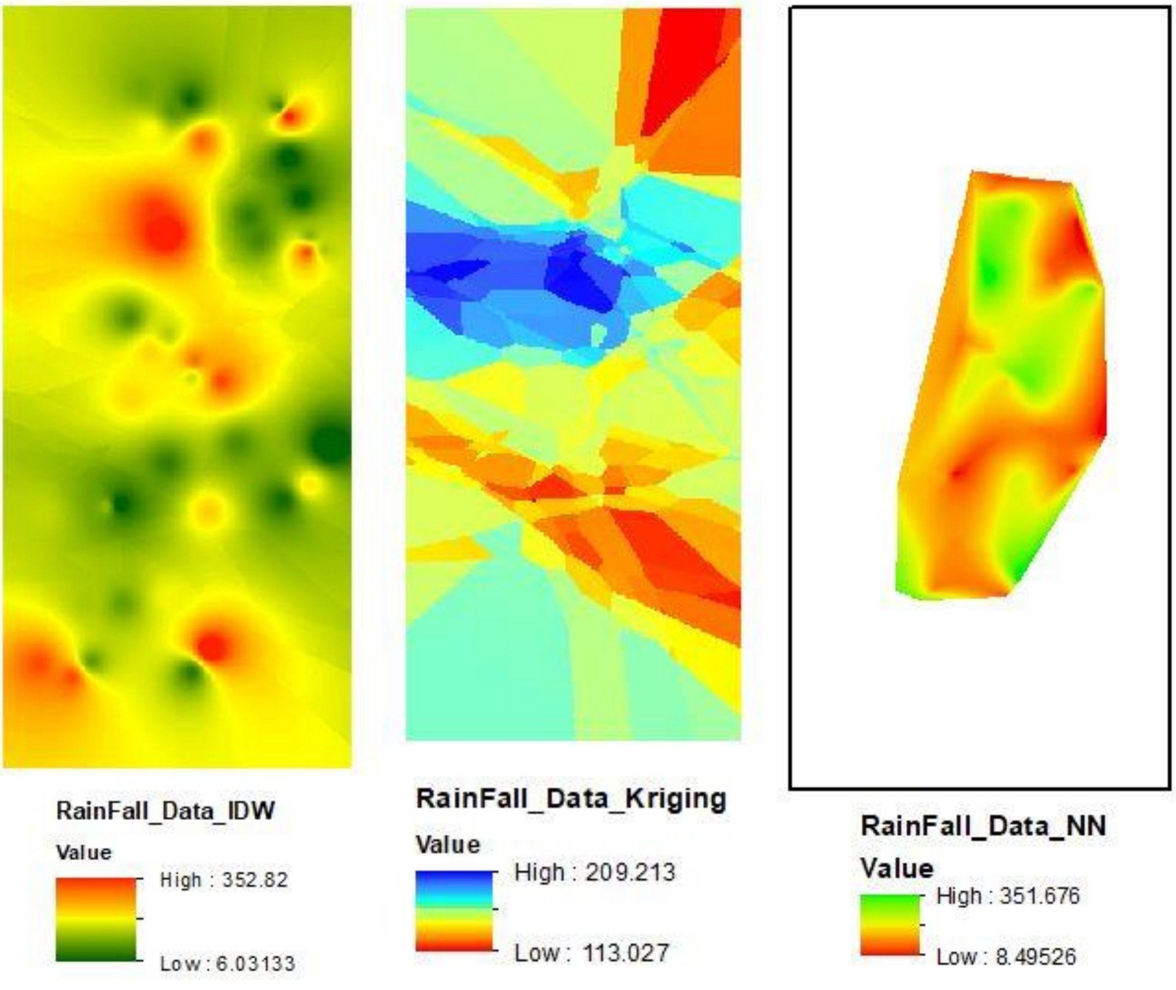


Figure 4:4 : Interpolation methods comparison

Statistical evaluation for each interpolation method in depicted in Table 4:2 and the output results are shown in Figure 4:4. According to the RMSE value calculated using different interpolation methods it could be decided that the IDW as the most suitable methods to distribute values for the coverage area.

Then the data values for the specified location can be obtained using the above rainfall distributed map. Values are extracted by using special geoprocessing tools which are available in ArcGIS software (Data extraction is explained in details in section 4.1.4)

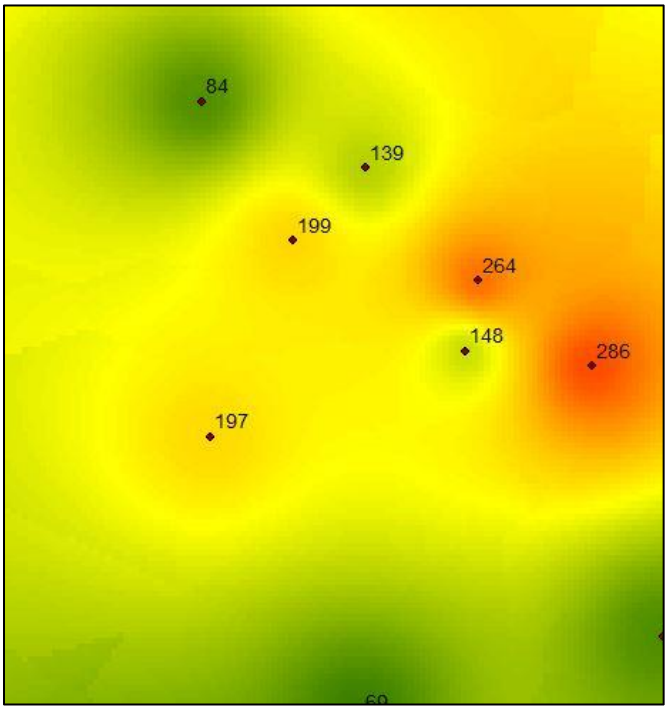


Figure 4:5 : Extracted rainfall values

The above map generation and point calculation is performed for the landslide event which occurred in December 2006 at Batawatta Group, Division 4 . Therefor this is calculated only for that landslide event. This process should be repeated for all the selected landslide events by generating rainfall interpolation maps for appropriate dates. These values will be used as the rain fall factor data for further processes.

II. Generating Slope map

Slope map is generated by using a DTM which is further improved by using a satellite image. DTM is "a statistical representation of the continuous surface of the ground by a large number of selected points with known X, Y, Z coordinates in an arbitrary coordinate field" [26]. Sentinel-2A L1C, False colour (urban)-2020-04-02 (Figure 4:6) is used as the satellite image for the slope generation. Sentinal images are taken from the sential satellite which are available at *https://apps.sentinel-hub.com.*

Figure 4:6 : Sentinel-2A L1C, False colour (urban)-2020-04-02

Then the values of the pixels are analysed for all three bands including the meta data of the image. Then the required bands are filter in order to extract data only from the relevant band (InfraRed, Visible Light, Radio waves etc.) And colour improvements are applied in order to extract data more precisely. The projection system should be adjusted according to the are which is accessing.

ArcGIS geotools are used to generate slope map for the selected area. The above improved satellite image is applied to the tool with relevant parameters as shown in Figure 4:7.

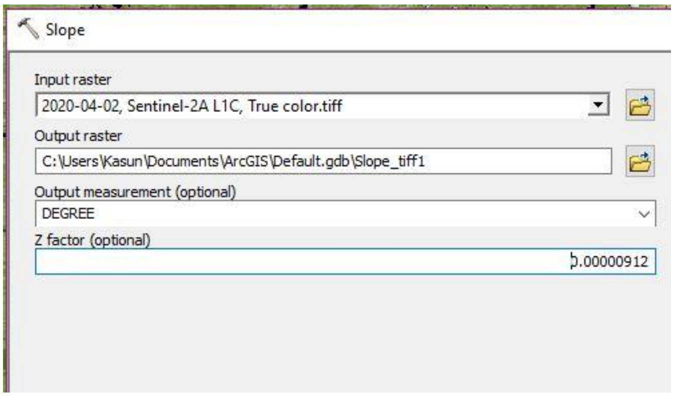


Figure 4:7: Attributes for slope map generation

The z-factor adjusts the units of measure for the z units when they are different from the x,y units of the input surface. The z-values of the input surface are multiplied by the z-factor when calculating the final output surface. Here the value 0.00000912 is applied according to the DTM properties.

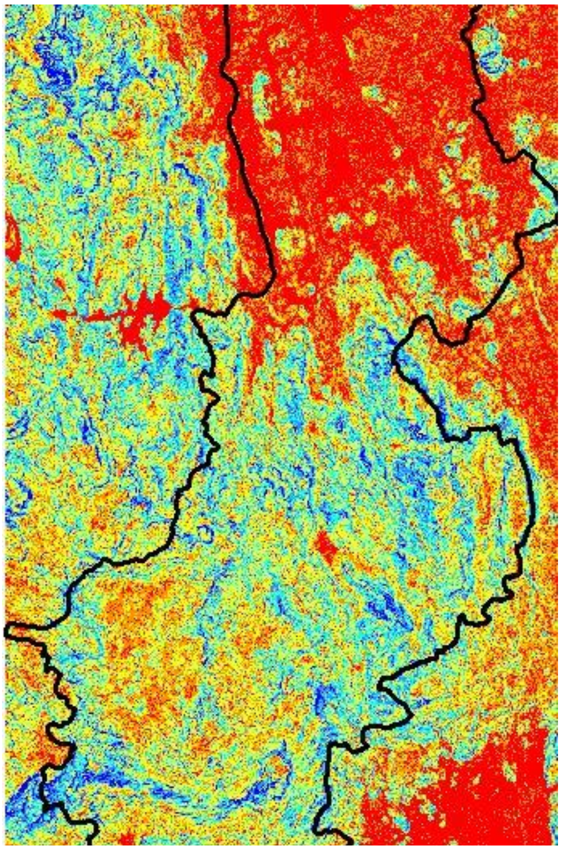


Figure 4:8 : Slope map

The Figure 4:8 shows the final reclassified slope map with attribute data for the specified area. As explained in previous point (rainfall map) this slope map is used to extract slope details of the landslide occurred locations.

III. Generating Surface Aspect map

Aspect represents the horizontal orientation of a surface and is determined in units of degrees. Each facet of the surface is assigned a code value which represents the cardinal or ordinal direction of its slope, and contiguous areas with the same slope directions are merged into one feature. This map defines the direction of the slope of the terrain.

ArcGIS Aspect generation tool is used to generate aspect map for the area. The same Sentinel-2A L1C, False colour (urban)-2020-04-02 satellite image is used to generate aspect detail map. The final surface aspect map is shown in Figure 4:9.

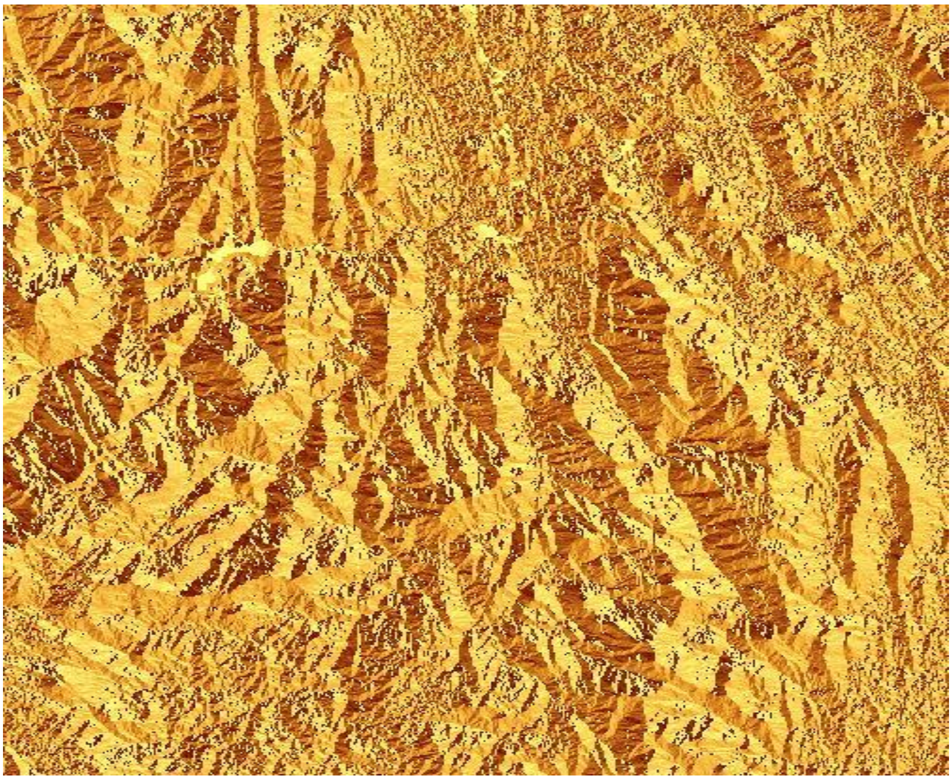


Figure 4:9: Surface Aspect Map

IV. Generating Landuse map

Land use illustrates how a particular part of the earth surface area is created and the features of the land. Landuse is important when landslide prediction since landuse directly affect the moisture level of soil. “Globcover_L4_200901_200912_V2.3_CLA_QL” satellite image is used to generate landuse map. Data source is GlobCover 2009 land cover maps (*https://datacatalog.worldbank.org/dataset/global-land-cover-2009*). The landuse is categorized in to 23 categories of landuse types as shown in the Figure 4:10 below.

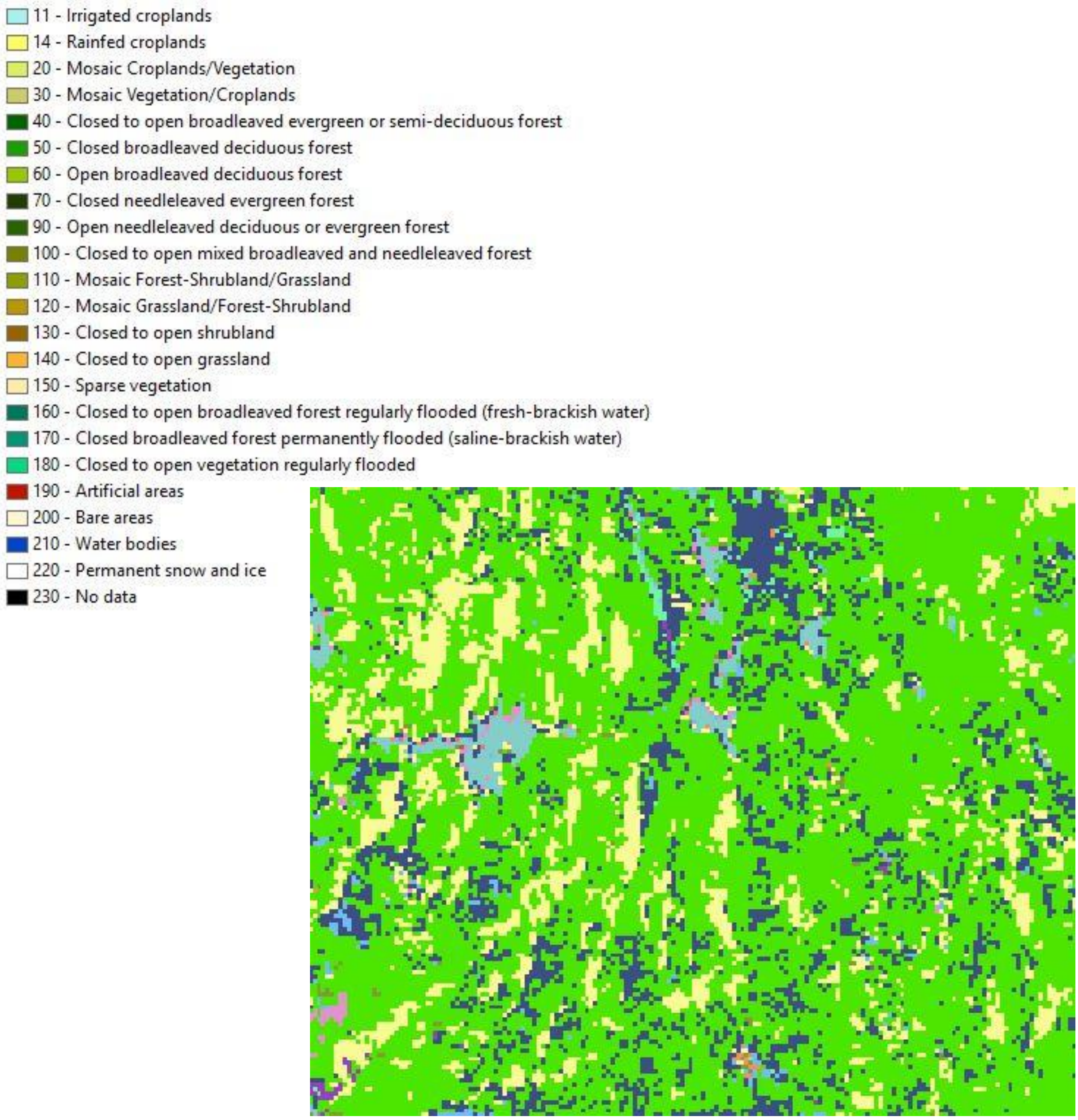


Figure 4:10: Landuse map

V. Generating Soil Density Map

” The soil bulk density (BD), also known as dry bulk density, is the weight of dry soil (Msolids) divided by the total soil volume (Vsoil). The total soil volume is the combined volume of solids and pores which may contain air (Vair) or water (Vwater), or both . The average values of air, water and solid in soil are easily measured and are a useful indication of a soils physical condition.” [27]

Soil type is widely used as a landslide factor for landslide prediction models and also for GIS analysis tasks. But in this research the factor Soil density is use for the first time as a landslide causing factor for predictions by using Artificial Neural Networks. The accuracy level of the results clealy shows that the factor soil density is a suitable measurement factor for landslide predictions.

Soil density map is generated using the data sheets taken from NBRO and the data obtained by processing satellite images. (Data source – SoilGrids-*https://www.isric.org/explore/soilgrids*)

Soil density reclassified map (Figure 4:11) is used to extract data for prediction model.

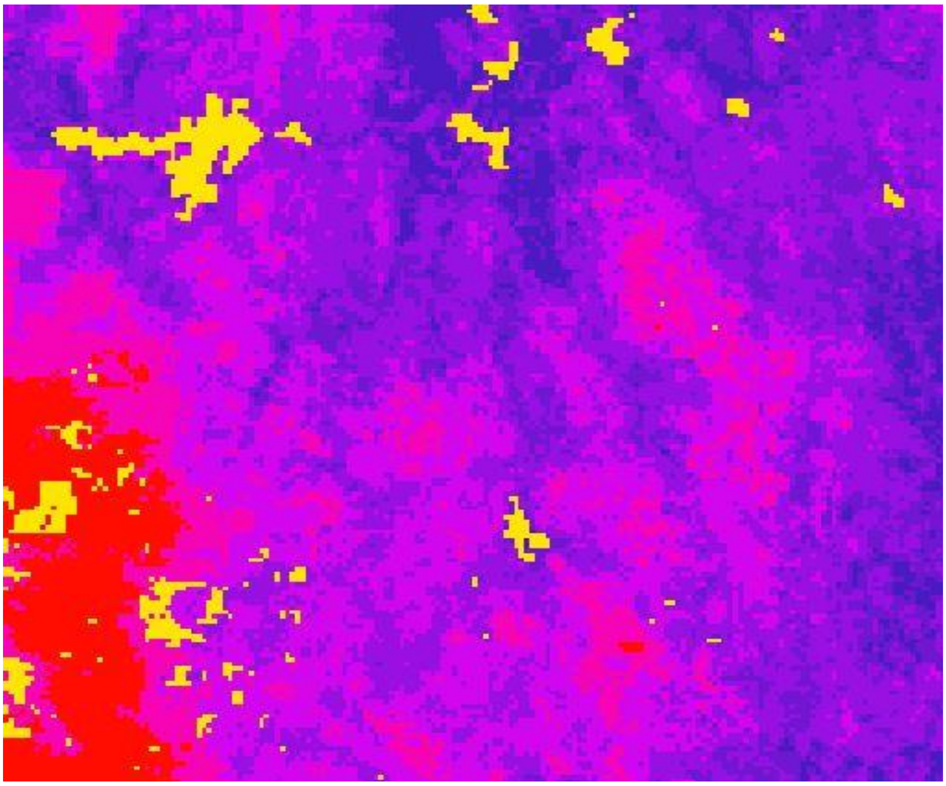


Figure 4:11: Soil density reclassified map

VI. Calculating distance to waterbodies

There is a direct connection between the hydrological effectivity and the landslide incidents. Therefore, it is important to calculate the distance to the nearest waterbody as a factor for the prediction model. This is performed stream network identification by hydrology analysis in ArcGIS.

The distance is calculated using the Euclidean distance calculation method for the stream lines and the specified locations. 2D Euclidean distance d between two cells between the linier index i and j is calculated with X and Y being matrices with the same size as Z where each element refers to the x,y coordinates respectively. Euclidean distance defined as squared distance between two vectors in multidimensional space is the sum of squared differences in their coordinates as shown in Equation 4:2.

$$d_{ij} = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2}$$

Equation 4:2 : 2D Euclidean distance

The shows the streams network of Badulla district area and the distance calculated map by applying the above equation.

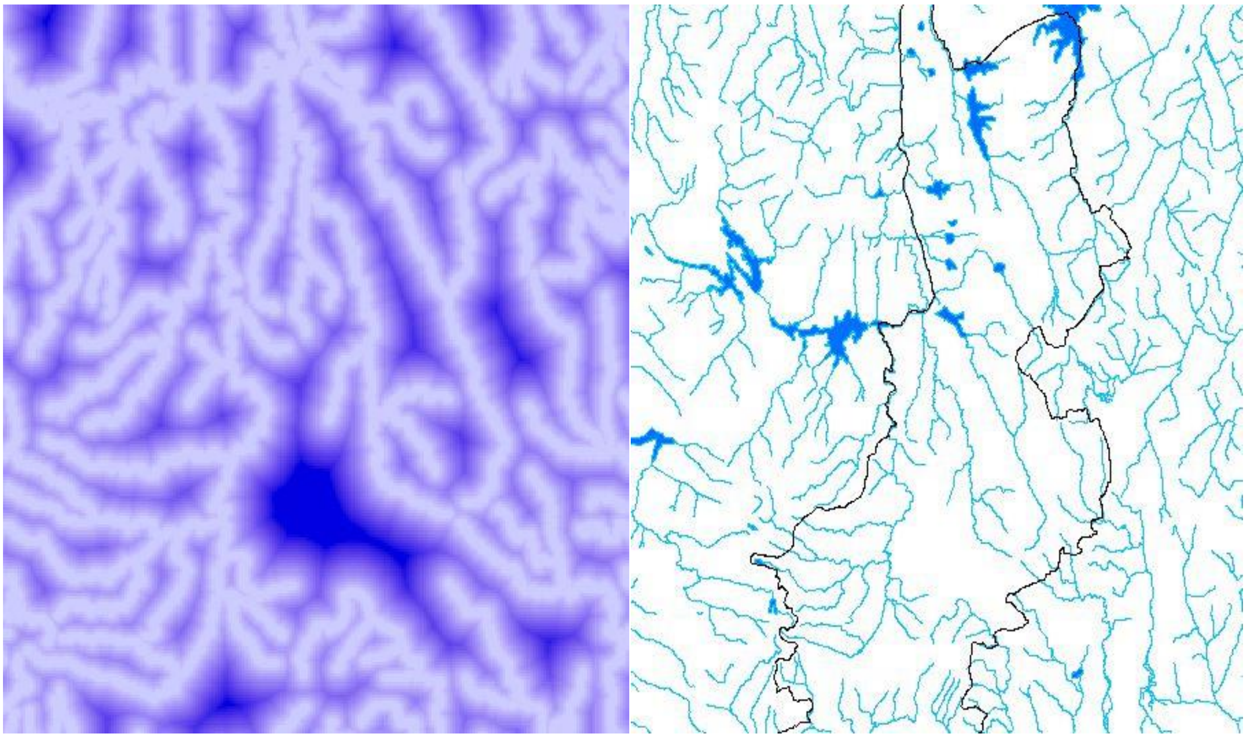


Figure 4:12: Distance to waterbodies

4.1.3 Data conversion and Rectification

All the above generated factor maps should be converted in to raster images for the data extraction process. Therefore, maps which are in GIS format (feature classes) are exported to File GeoDatabase Raster Dataset format. ArcGIS facilitates several data conversion tools (ASCII to raster, DEM to raster etc) for raster vector data conversions.

All factor maps including raster converted maps are then checked for the coordinate system differences and projection errors. If there are any differences, they are corrected using geoprocessing tools available in ArcGIS software. Stream network and lakes data layer is georeferenced (All layers are in WGS84 coordinate system) in order to overlay to the other layers. Some factor layers are reclassified in to smaller number of classes to optimize for neural network input dataset.

4.1.4 Extracting factor data

Data values for each landslide occurred point should be extracted from all of above factor layers in order to create geospatial database. The Spatial Analyst extension in ArcGIS software offers several tools that can do this process such as,

- Extract Values to Points
- Extract Multi Values to Points
- Sample

According to the size, band and the other properties of factor layer raster datasets, “Extract Values to Points” Spatial Analyst tool is used to extract data values. This tool only takes a single raster input. Even though the input is a multiband raster, it will only process on the first band by default or the single band defined. The raster values are stored in a predefined field called “RASTERVALU”.

However, the main problem arises when creating geospatial database as input dataset for a Neural network, is lack of known events (landslide events) for the training. Because ANN requires large amount of data to successful training of the network. As a solution for this problem, following two methods are used to increase the number of **Landslide Initiation Points**.

1. Selecting (extracting data) more than one point through the landslide initiation boundary line which is identified by NBRO for each landslide. (Also known as the crown of a landslide) (See Figure 4:13)

2. Selecting initiation points from other types of landslide events (Earthflow, Mud/rock flow)

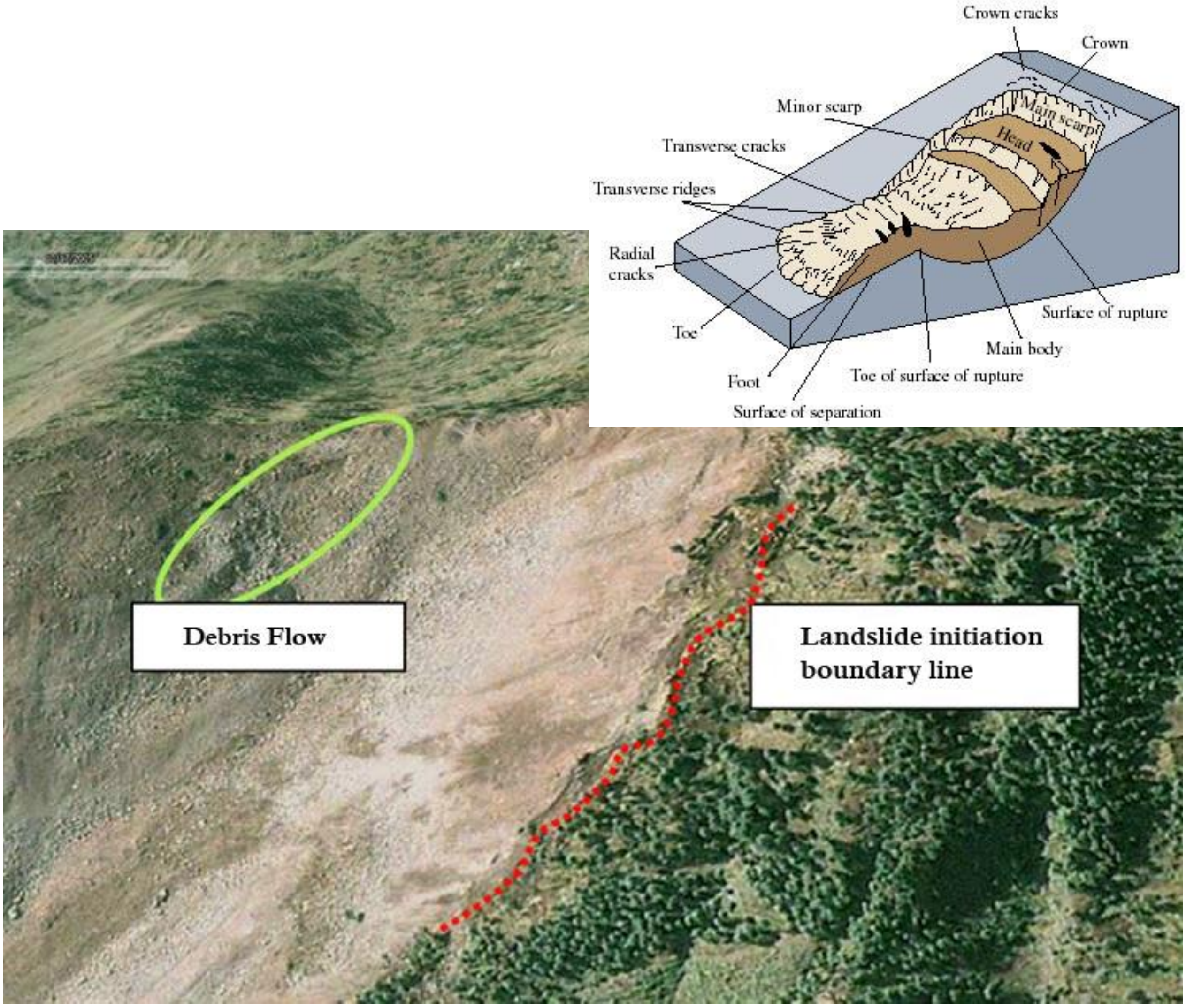


Figure 4:13 : Landslide Initiation boundary line

Number of landslide events - 93

Number of total **landslide initiation points** - 826

Therefor the above solution methods have efficiently increased the number of input datasets for the ANN network training while maintaining the value differences of the factor data.

Then data is extracted from each layer by using "Extract Values to Points" tool (Figure 4:14). This tool appends the extracted new data to the existing layer attribute data table.

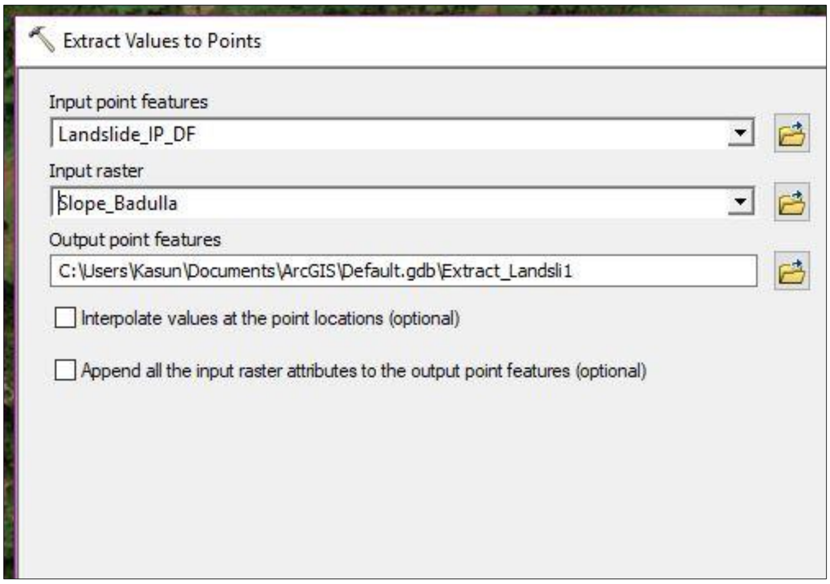


Figure 4:14 : Extract Values to Points

Generating Non-Landslide points

Similar number of points (826 points) are selected from the other areas (Ares identified as no previous landslides occurred). Then data extraction is again applied (as explained in previous point) to this point layer and result dataset is also used to create the geospatial database. These **non-occurred data** are required to perform a better training process of the network.

4.1.5 Creating geospatial database

The target of all previous points and workflows is to create a geospatial database (Figure 4:15) in order to store all the factor information and their corresponding values. This database table contains following data,

Training Dataset	• LHZM_Range • RainFall_Data_IDW • Slope data • Aspect data • Soil density • Landuse • Distance to waterbodies
Target Dataset	Contains target data for ANN. Data is binary represented as Two states (landslide occurrence and landslide non-occurrence data is represented)

At the previous feature extraction phase, altogether 1652 points taken as final dataset. But these points are further categorized as Training dataset and Evaluation dataset. Several points (04 points) are removed since they contain identical data rows. Finally, the table contains 1648 events. From this 1348 tuples are selected for training (Approximately 80% for training and 20% for validation and testing) and 300 tuples for evaluation and testing.

Input - Training Set 01								Target	
OBJECTID	LHZM_Range	RainFall_Data_IDW	Slope_Badulla	Aspect_Badulla	Soil_Density_Reclass	Landuse_Badulla	EucDist_Waterbodies	Occurrence_OP1	Occurrence_OP2
100	2	147.5480804	19.06338692	105.8519287	4	100	0.011533928	1	0
101	1	147.5480804	18.64581871	108.4349518	4	100	0.011533928	1	0
102	1	147.5480804	12.96311092	101.3099289	4	100	0.011533928	1	0
106	2	147.5480804	18.64581871	108.4349518	4	100	0.011533928	1	0
107	2	106.7199783	24.94881821	355.446228	3	40	0.018236741	1	0
108	2	106.7199783	26.53945923	353.8677673	3	40	0.018236741	1	0
109	1	147.8196564	11.14412022	90	4	100	0.011868314	1	0
110	1	147.8196564	12.3746748	97.52381897	4	100	0.011868314	1	0
111	2	147.8196564	12.3746748	97.52381897	4	100	0.011868314	1	0
112	2	147.8196564	17.04953003	105.5241089	4	100	0.011868314	1	0
113	2	147.8196564	17.35670471	119.9315109	4	100	0.011868314	1	0
114	1	147.8196564	10.24628258	92.60256195	4	100	0.011868314	1	0
115	2	106.7199783	16.52171516	345.5792236	3	40	0.018236741	1	0
116	2	106.7199783	15.88791561	24.71744156	3	40	0.018236741	1	0
117	2	106.7199783	16.46778297	1.59114027	3	40	0.018236741	1	0
118	2	106.7199783	16.52171516	345.5792236	3	40	0.018236741	1	0
119	1	147.8196564	10.97769928	83.92754364	4	100	0.011868314	1	0
120	1	147.8196564	10.24628258	92.60256195	4	100	0.011868314	1	0
121	1	147.8196564	10.04774761	103.3924942	4	100	0.011868314	1	0
123	2	147.8196564	13.81851578	115.7099533	4	100	0.011868314	1	0
124	3	106.7199783	14.06748199	31.60750198	3	40	0.018236741	1	0
125	2	106.7199783	14.06748199	31.60750198	3	40	0.018236741	1	0
126	2	106.7199783	11.92925262	7.815293312	3	40	0.018236741	1	0
127	2	106.7199783	12.58664417	323.9726257	3	40	0.018236741	1	0
128	3	106.7199783	15.10915089	19.53665543	3	40	0.018236741	1	0
129	2	106.7199783	15.10915089	19.53665543	3	40	0.018236741	1	0
130	2	106.7199783	9.823251724	354.5596619	3	40	0.018236741	1	0
131	2	106.7199783	12.72574329	289.0935059	3	40	0.018236741	1	0
132	3	104.3350906	23.14795303	356.6981201	3	40	0.017912025	1	0
133	3	106.7199783	17.57161713	356.2846985	3	40	0.018236741	1	0
135	2	106.7199783	12.12423706	313.4518433	3	40	0.018236741	1	0
140	3	104.3350906	15.95049381	338.9624939	3	40	0.017912025	1	0
141	3	106.7199783	12.89923	323.7461548	3	40	0.018236741	1	0
143	3	104.3350906	12.52472019	281.7250977	3	40	0.017912025	1	0
144	2	104.3350906	9.368200302	264.2893982	3	40	0.017912025	1	0
145	3	106.7199783	9.166574478	262.6942444	3	40	0.018236741	1	0
146	2	167.418396	4.897940636	196.6992493	4	40	0.001978052	1	0
147	3	167.418396	13.29661942	110.3231354	4	130	0.001978052	1	0
148	3	167.418396	18.01594734	222.9545898	4	40	0.001978052	1	0
149	3	167.418396	12.62363338	166.218399	4	130	0.001978052	1	0
150	3	167.418396	21.53964996	224.1574707	4	130	0.001978052	1	0
151	3	167.0500336	18.07218552	83.50065613	4	130	0.001978052	1	0
152	3	167.418396	14.5679903	105.5725403	4	130	0.001978052	1	0
154	2	167.0500336	25.03046417	102.6926651	4	130	0.001978052	1	0

Figure 4:15: Geospatial Database (Sample Dataset)

4.1.6 ANN training process

This is the second phase of the research. In this phase, Artificial Neural Network is developed using Matlab inbuilt nprtool (Neural network Pattern Recognition) and Matlab scripting language. Training data of each sample are simplified/standardized in order to improve the accuracy level.

As described in the previous section, geospatial database is the data source for the neural network train process. The following steps required to import data and execute the training process of the network

input
7x1462 double

	1	2	3	4	5	6	7	8	9	10
1	3	3	3	3	2	2	2	2	2	3
2	162.1272	162.1272	162.1272	162.1272	162.1272	162.1272	162.1272	163.1382	161.3824	163.1382
3	26.6335	22.0453	22.5628	15.6186	15.6186	14.9144	15.0323	19.9820	7.0189	13.5378
4	158.8952	162.2996	147.0948	130.2364	130.2364	146.3099	141.2034	290.4803	306.8699	240.3763
5	4	4	4	4	4	4	4	4	4	4

target
2x1462 double

	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	1	1	1	1	1
2	0	0	0	0	0	0	0	0	0	0

Figure 4:16: Input & target data

- Creating input and target variables. – Two variables (input, target) are created in Matlab workspace and load input/target data to the variables respectively. See Figure 4:16. Data should be transposed in to horizontal format in order to feed the data to the network.

- Loading the script to the workspace and set the folder path to the current folder. And then run the following command the command windows to run the script.

  >> *Landslide_NN_Adv_Script*

- The above command will run the script and train the network according to the input and target dataset. Once the training process is completed, the trained network object, output data, performance data are save to the workspace by the script. The progress of the training process is shown in Figure 4:17.

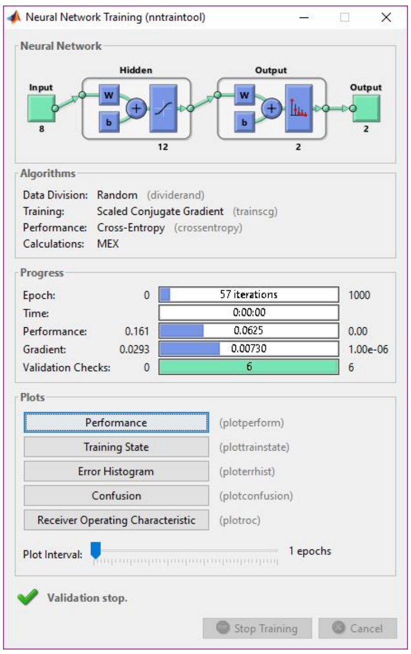


Figure 4:17: Progress of the training

- Creating new variable (testdata) to load the test dataset. This dataset is used to test the network before the predictions. This test dataset contains 20% of the total dataset. See Figure 4:18

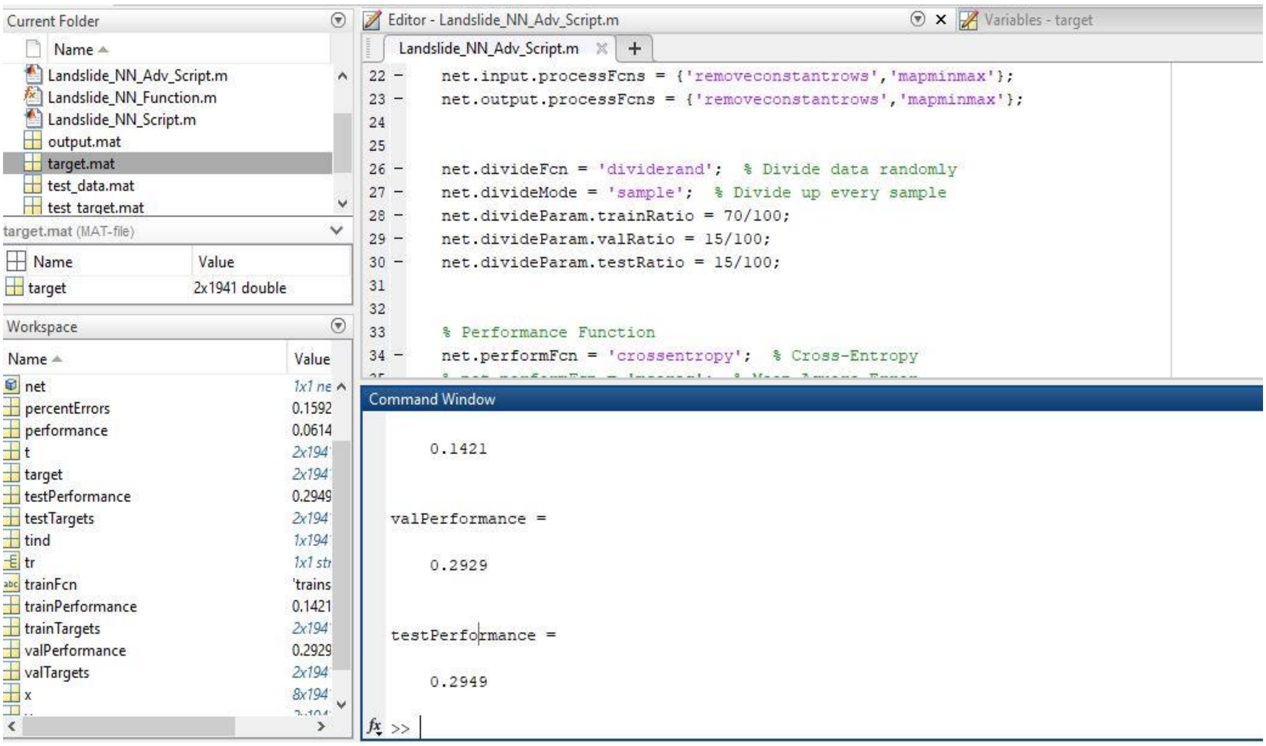
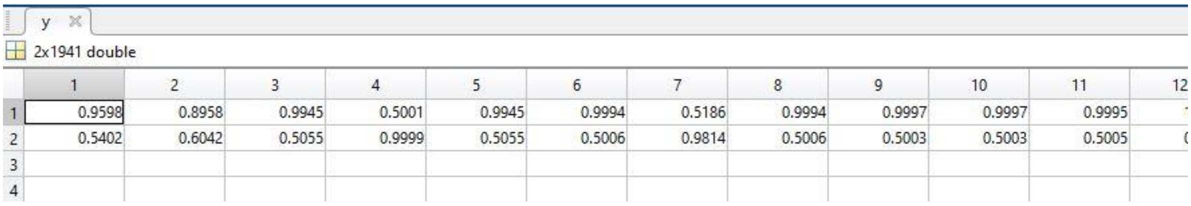


Figure 4:18: Training process

- Executing the network for prediction process. Following command is used to get prediction results. "Result" is the name of the output variable, "landnet" is the trained network object. "testdata" is the variable containing test data set.

  >> *Results=landnet(testdata)*

- Exporting result data. (Data transpose is done prior to exporting data). This returns decimal format data as prediction results. See Figure 4:19.



y

2x1941 double

	1	2	3	4	5	6	7	8	9	10	11	12
1	0.9598	0.8958	0.9945	0.5001	0.9945	0.9994	0.5186	0.9994	0.9997	0.9997	0.9995	
2	0.5402	0.6042	0.5055	0.9999	0.5055	0.5006	0.9814	0.5006	0.5003	0.5003	0.5005	
3												
4												

Figure 4:19 : Prediction results (decimal format)

4.1.7 Comparing Results

The above exported resultant data is compared with the expected output. The accuracy and performance of the prediction results and calculated in this step. Although the expected data is in binary format (1,0 and 0,1), prediction results are given in decimal format.

Therefore, this decimal format numbers are categorized in to binary format (True – Accurate prediction or False – Wrongful prediction) . Values which are closer to ‘1’ are categorized as “1” and valued which are closer to “0” are categorized as “0”. According to the categorization, the combined result (True or False) is computed for each data record as shown in Figure 4:2.

Landuse_Badulla	8.EucDist_Waterbodies	Expected_OP1	Expected_OP2	ANN Result1	ANN Result2	True/False
40	0.006255151	1	0	0.999575165	0.000424835	1
40	0.011533928	1	0	0.999528717	0.000471283	1
40	0.011533928	1	0	0.999761068	0.000238932	1
40	0.011533928	1	0	0.999767903	0.000232097	1
130	0.001978052	1	0	0.998532253	0.001467747	1
40	0.016900487	1	0	0.993359788	0.006640212	1
40	0.008846119	1	0	0.071702095	0.928297905	0
40	0.010652138	1	0	0.992160836	0.007839164	1
40	0.010652138	1	0	0.999355757	0.000644243	1
100	0.011533928	1	0	0.99949621	0.00050379	1
40	0.018236741	1	0	0.920789635	0.079210365	1
100	0.011868314	1	0	0.995760373	0.004239627	1
40	0.018236741	1	0	0.035910459	0.964089541	0
40	0.018236741	1	0	0.988045691	0.011954309	1
130	0.001978052	1	0	0.591770894	0.408229106	0
40	0.006255151	1	0	0.999829638	0.000170362	1
130	0.009890262	1	0	0.617367064	0.382632936	0
130	0.003956105	1	0	0.513890074	0.486109926	0
130	0.003956105	1	0	0.500899284	0.499100716	0
130	0.003956105	1	0	0.139217549	0.860782451	0
130	0.003956105	1	0	0.519520875	0.480479125	0
40	0.012510302	1	0	0.999880102	0.000119898	1
100	0.001978052	1	0	0.948686162	0.051313838	1
40	0.012510302	1	0	0.999880102	0.000119898	1
40	0.048128091	1	0	0.999654752	0.000345248	1
40	0.048128091	1	0	0.999613805	0.000386195	1
130	0.009890262	1	0	0.997626247	0.002373753	1

Figure 4:20: Compared Results

Finally, Accuracy of the prediction is calculated based on the successful prediction count (True count) of the compared data.

If the accuracy is not in satisfactory level the network should be retrained until the results are good.

The above dataset returned 71% accuracy for landslide prediction when considering overall results.

CHAPTER 05

EVALUATION

CHAPTER 05

EVALUATION

5.1 Evaluation Criteria

The developed GIS model and Neural Network model is evaluated using the past landslide data and the Landslide susceptibility GIS data and results returned by the neural network model. The ANN model was evaluated and statistical data analysis was done using

- Input data (data from Landslide geospatial database which are used to train the network)
- Test data (Set of data which are randomly selected from the landslide database. These are new data which are not used to train the ANN model)

Mainly the evaluation of the study is performed by calculating Accuracy, plotting precision and recall (ROC) and analysing performance of the neural network. In addition to that weight determination analysis is performed in order to identify effectiveness of selected landslide causing factors. Accuracy is calculated comparing results returned from the Neural network and the expected outputs in the geospatial database. A separate new data set is used for this process. All factors are used to perform the above calculations.

5.2 Evaluation Approach

The Evaluation methodology for the research study is shown in Figure 5:1.

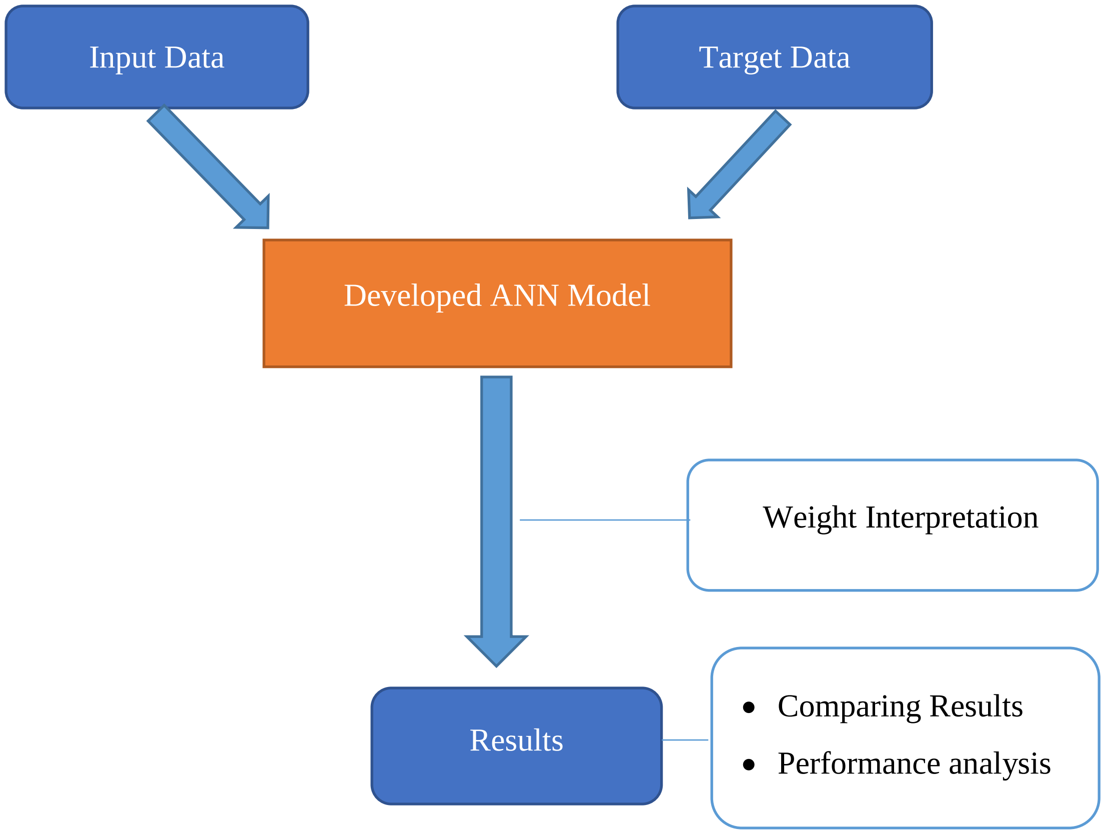


Figure 5:1: Evaluation methodology

5.3. Effectiveness of the factors

The effectiveness of the factor parameters is calculated from the weights interpretation which are determined from back-propagation neural network. The total dataset is divided in to 05 separate datasets and each dataset is loaded to the network in each case. These datasets were trained several times until the performance of the results prove to be higher.

After each successful training process weights are calculated from the all input nodes. This is considered as the landslide susceptibility index for each case of train process. The weight calculation process is shown in Figure 5:2.

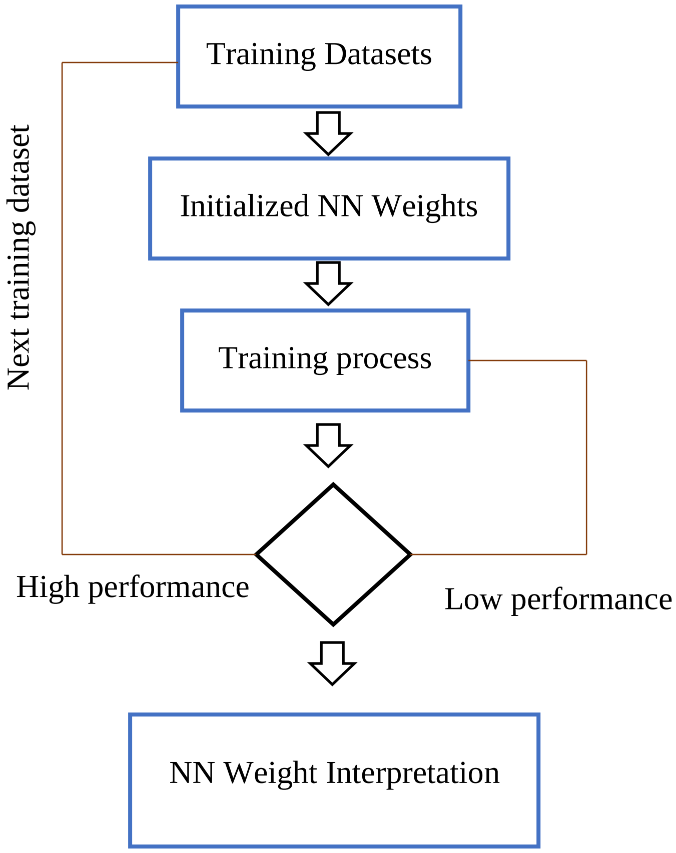


Figure 5:2: Weight Calculation process

The training datasets are selected from the geospatial database covering different areas of the surface. The back-propagation algorithm is applied to calculate weights between the input layer and the hidden layer and between the hidden layer and the output layer by modifying the number of hidden layers neurons and the learning rate.

The weights were applied to the entire study area. The calculated weight values are then listed vs each case of the training iteration. Calculated values are shown in Table 5:1.

Case No:	Case 01	Case 02	Case 03	Case 04	Case 05	Case 06	Case 07	Case 08	Case 09	Case 10	Mean	SD
LHZM_Range	0.9788	1.2259	-0.8719	-0.4544	-0.6197	0.1383	0.6226	0.7311	-0.2358	0.4885	0.2003	0.718474
RainFall Data	1.4842	1.1792	0.7626	1.0668	1.1567	0.5991	0.2352	1.0040	0.3596	-0.5259	0.7321	0.588649
Slope	0.0664	0.6742	0.5971	1.0660	-0.2621	1.7033	-0.9812	1.0411	-0.2943	-0.5393	0.3071	0.84346
Aspect	1.2299	-0.8048	0.7887	-0.2000	-0.7778	-0.5927	0.4109	-0.0910	0.2288	0.9001	0.1092	0.722568
Soil Density	0.0232	0.3290	0.2620	1.3417	-1.0705	0.8714	-0.4703	0.0536	0.3602	-0.9535	0.0747	0.752057
Landuse	0.2583	0.7626	0.8838	-1.0384	0.5168	0.0098	-1.2653	-0.6023	-0.1901	0.8452	0.0180	0.780738
Distance to Waterbodies	-0.6529	0.0927	-0.7219	0.2916	-0.7907	0.5179	0.6127	0.8570	-0.1243	-0.8854	-0.0803	0.648008

Table 5:1: Weight Calculations

Plot diagrams for each landslide factor is depicted in Figure 5:8 to Figure 5.9.

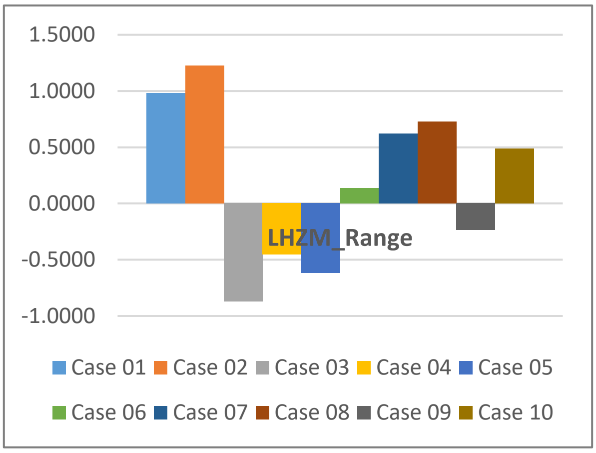


Figure 5:3: LHZM Range

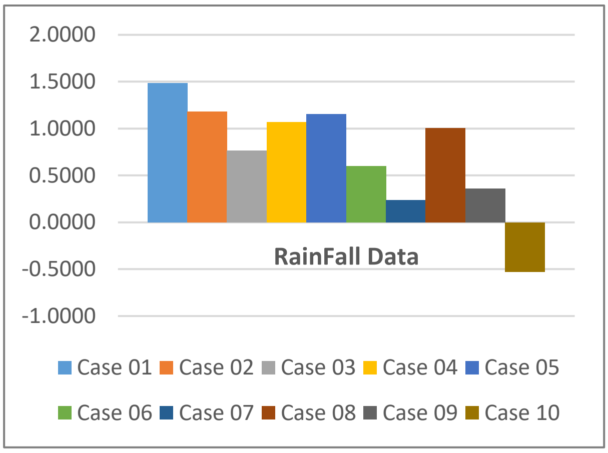


Figure 5:4: Rainfall Data

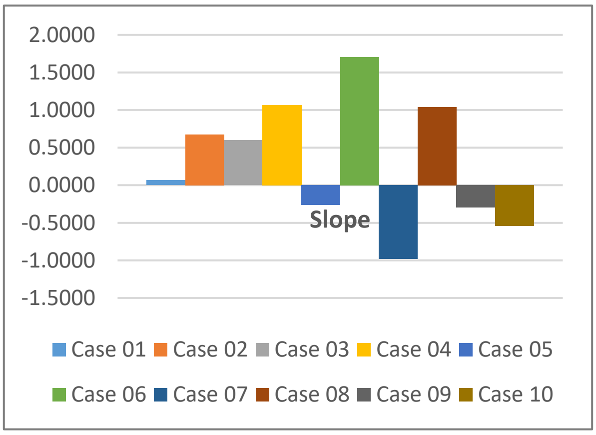


Figure 5:5: Slope

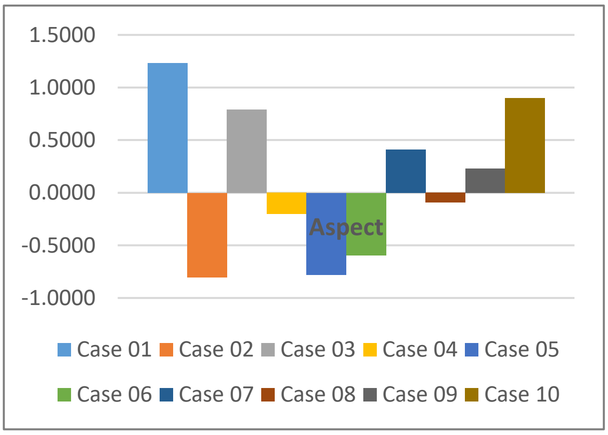


Figure 5:6: Aspect

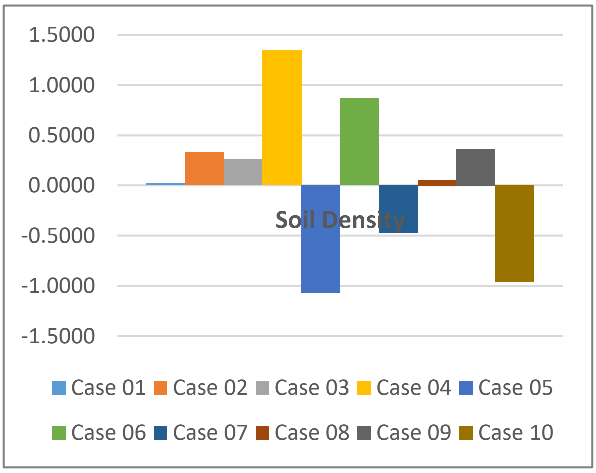


Figure 5:7: Soil Density

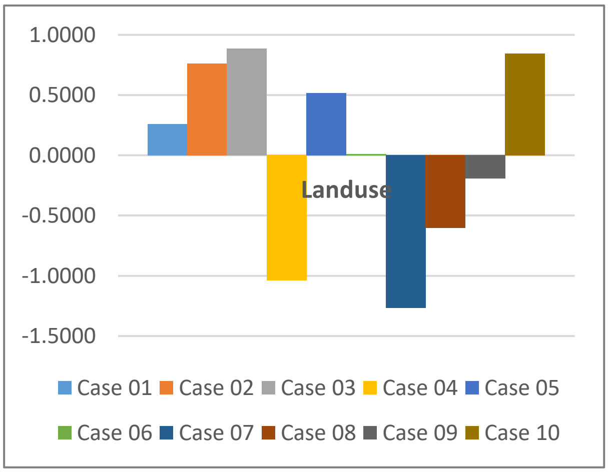


Figure 5:8: Landuse

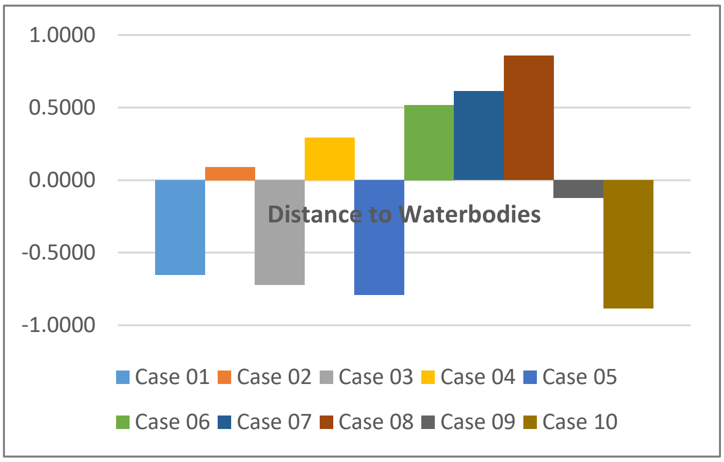


Figure 5:9: Distance to Waterbodies

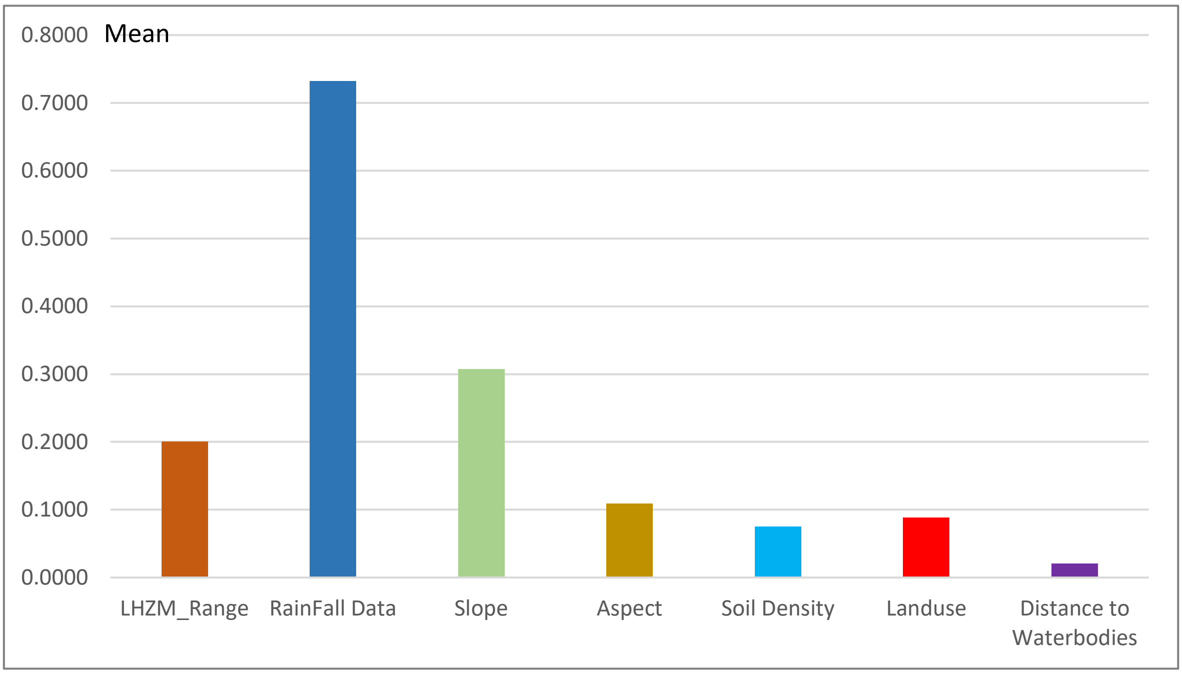


Figure 5:10: Factors with mean weights

Figure 5.10 shows the mean weight values calculated for each factor considering all iterations.

The figure depicts comparative analysis of landslide causing factors with the normalized weights. This graph clearly indicates that the **rainfall** and **slope** as dominant factors in causing the landslides in the study area.

5.4. Performance of Artificial Neural Network Model

5.4.1. Classification Accuracy

The training is done by using 1348 Landslide/Non-landslide points and the accuracy for training process are 79.40% and 63.34% for Landslides occurrence and Landslide Non occurrence respectively. (See Figure 5:11). The overall accuracy was 71.33%.

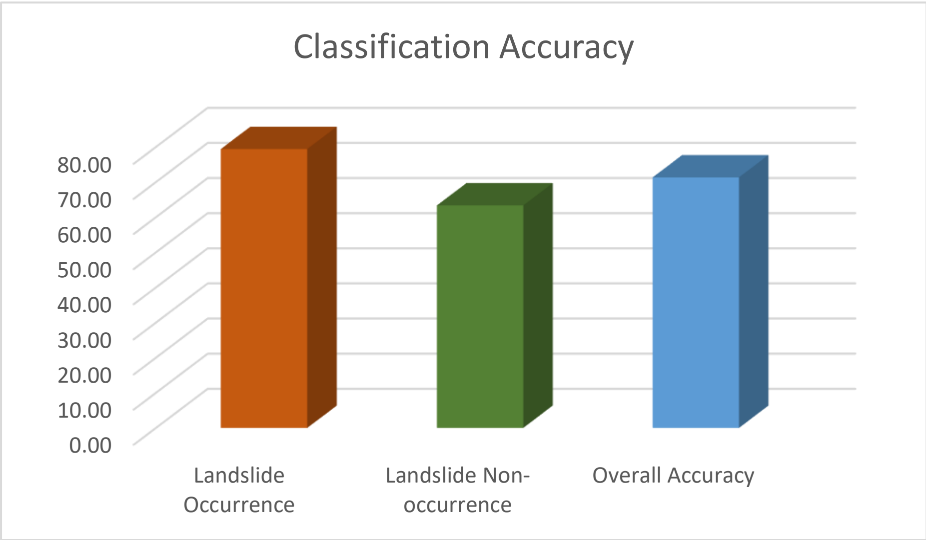


Figure 5:3: Classification Accuracy

5.4.2. Regression Analysis

The regression analysis is done between the network response and the corresponding targets for training data set using Matlab regression plot tool (Figure 5:12). The regression analysis for prediction results is shown in Figure 5:13.

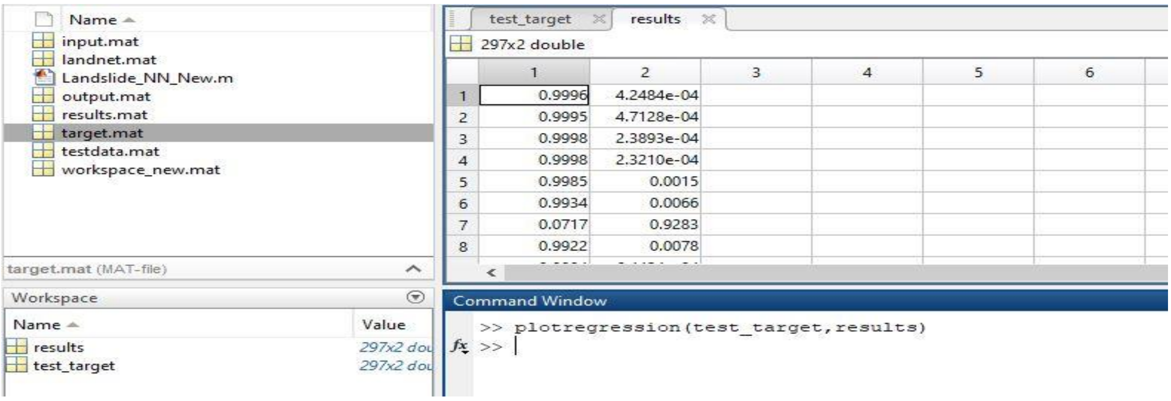


Figure 5:4: Matlab regression plot tool

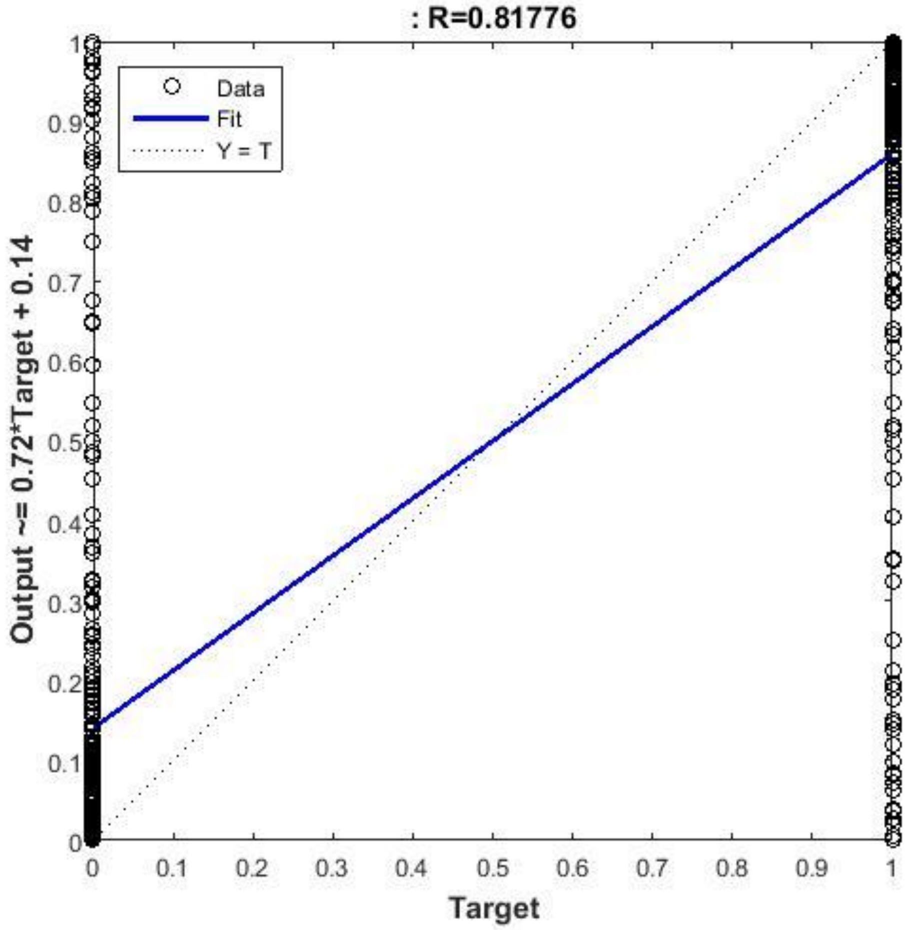


Figure 5:5: Regression analysis

5.4.3. Performance of the Neural network

Performance of an Artificial Neural Network is computed by analyzing various techniques and by statistically analyzing output data. Here the performance is calculated analyzing No of epochs vs Cross-entropy. Best performance is found as 0.052 at epoch 37. Figure 5:14 shows the overall performance of the trained neural network.

"The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The true-positive rate is also known as sensitivity, recall or probability of detection" [28] It is a graphical plot that illustrates the diagnostic ability of the model. Figure 5:15 shows the ROC curve for the developed model.

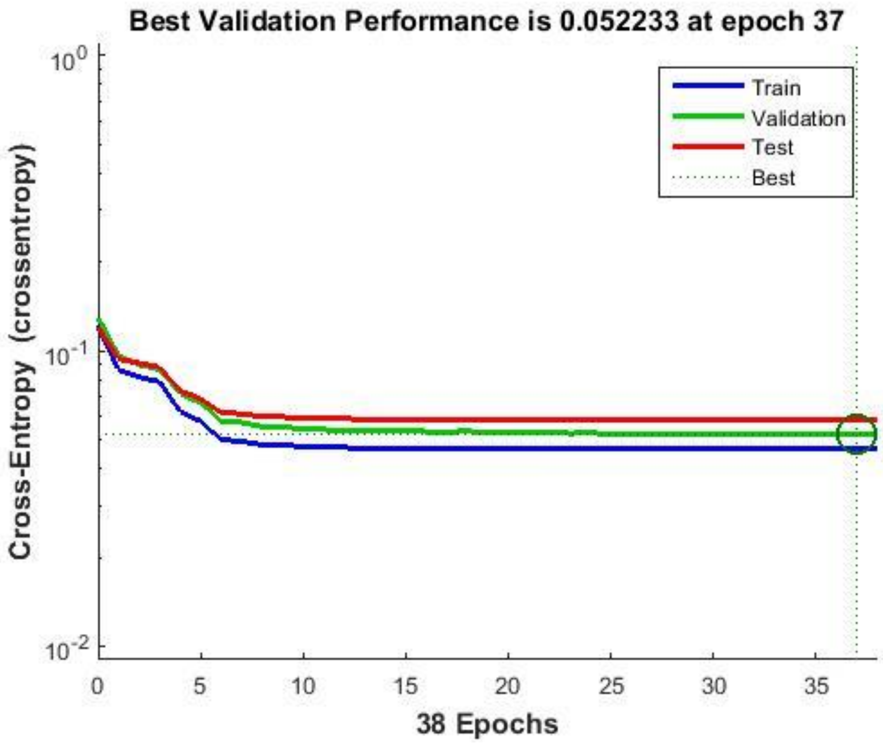


Figure 5:6: Performance of ANN

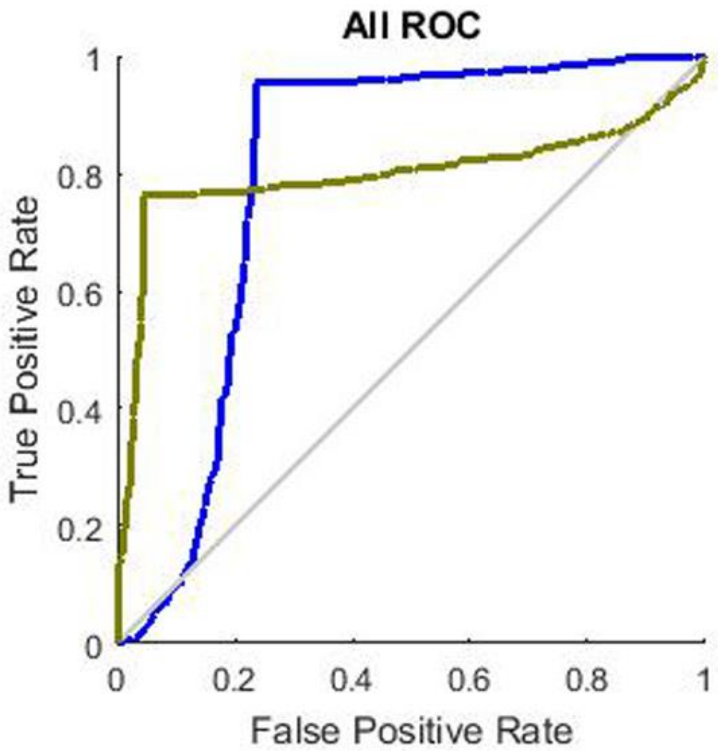


Figure 5:7: ROC curve for the ANN model

5.5. Accuracy assessment for use of Artificial Neural Network tools for Landslide Prediction

The results and the accuracy level of the predictions demonstrate that the neural network can be used to produce a landslide susceptibility and, consequently, to manage landslide hazards effectively. However, in this study GIS is used as a tool to geospatial analysis and factor data extraction other than the ANN tools and techniques. Sample set of test data prediction results are shown in Table 5:2.

OBJECTID	Factor Data		Expected OP1	Expected OP2		ANN Output1	ANN Output2		Result
6	FD1		1	0		0.999575165	0.000424835		TP
13	FD2		1	0		0.999528717	0.000471283		TP
14	FD3		1	0		0.999761068	0.000238932		TP
19	FD4		1	0		0.999767903	0.000232097		TP
39	FD5		1	0		0.998532253	0.001467747		TP
47	FD6		1	0		0.993359788	0.006640212		TP
61	FD7		1	0		0.071702095	0.928297905		FN
66	FD8		1	0		0.992160836	0.007839164		TP
81	FD9		1	0		0.999355757	0.000644243		TP
88	FD10		1	0		0.99949621	0.00050379		TP
98	FD11		1	0		0.920789635	0.079210365		TP
101	FD12		1	0		0.995760373	0.004239627		TP
115	FD13		1	0		0.005910459	0.964089541		FN
120	FD14		1	0		0.988045691	0.011954309		TP
125	FD15		1	0		0.007708939	0.999229106		FN
162	FD16		1	0		0.999829638	0.000170362		TP
189	FD17		1	0		0.000367064	0.996329363		FN
224	FD18		1	0		0.003890074	0.986109926		FN
225	FD19		1	0		0.000899284	0.999100716		FN
226	FD20		1	0		0.039217549	0.960782451		FN
227	FD21		1	0		0.019520875	0.990479125		FN
248	FD22		1	0		0.999880102	0.000119898		TP
249	FD23		1	0		0.948686162	0.051313838		TP
250	FD24		1	0		0.999880102	0.000119898		TP
251	FD25		1	0		0.999654752	0.000345248		TP
252	FD26		1	0		0.999613805	0.000386195		TP
283	FD27		1	0		0.997626247	0.002373753		TP
284	FD28		1	0		0.998362542	0.001637458		TP
285	FD29		1	0		0.9980071	0.0019929		TP
286	FD30		1	0		0.039067403	0.960932597		FN
:	:		:	:		:	:		:
:	:		:	:		:	:		:

Table 5:2:Test data prediction results

Figure 5:8 shows the confusion matrix for above results

n=300	**Predicted (No)**	**Predicted (Yes)**	
Actual (No)	TN - 95	FP - 55	150
Actual (Yes)	FN - 31	TP - 119	150
	126	174	214

Figure 5:8: Confusion matrix for test results

Accuracy = (TP+TN)/total

= 71.33%

Recall =TP/(TP+FN)

=0.79

Precision =TP/(TP+FP)

= 0.46

Moreover, the weighting given to the various factors that are significant in this landslide susceptibility analysis provides a ranking of their relative significance. In this neural network method, it is difficult to follow the internal processes of the procedure and the method requires a long execution time. This is because high distribution of the factor data and the amount of data content.

In addition, artificial neural network models are adaptive and capable of generalization. They can handle imperfect or incomplete data and can capture non-linear and complex interactions among variables of a system. Therefore, in this research Artificial Neural Network tools are successfully used as a prediction model foe landslide susceptibility prediction.

CHAPTER 06

CONCLUSION

CHAPTER 06

CONCLUSION

Landsides have become a major natural disaster in Sri Lanka causing several damage to the human life and property. Early prediction of landslides will be highly beneficial to the society. This research was carried out to develop a mechanism for early warning of landslides.

Due to the nature of uncertainty in predicting landslides, the overall prediction accuracy of 71.3% by the ANN can be considered as an encouraging achievement. After extracting the factors considered by the model, those factors are input to the model..

The Landslide occurrences predicted by the implemented tool shows good agreement with the actual landslide data collected for the selected research study.

6.1. Future Work

This research study is mainly based on Badulla district as the selected study area. But the identified factors affecting the landsides are also applicable to other districts of the country. Accordingly, this prediction model can be used to predict the landslide susceptibility of the other districts.

However sufficient test and analysis must be performed prior to the application of the model to the other areas.

In this research landslide related 07 factors (LSHZ Range Rainfall, Surface Slope, Aspect, Soil Density, Land use, Distance to Waterbody) are used for the prediction model. But there are some other factors which can be identified as landslide causing factors like Curvature, Soil type, geological structure etc. Data from these factors can be applied by increasing the number of events. Then prediction model will be expanded to accepts wide range of factor data.

As identified in this research, rainfall is a prominent factor when considering landslide prediction. Thus this model can be connected to an real time rainfall forecasting service and can be further developed to perform landslide predictions automatically and use as a early warning system with a satisfactory accuracy level of 71.3%.

References

[1] Sangchini, E.K., Nowjavan, M.R., Arami, A., “Landslide susceptibility mapping using logistic statistical regression in Babaheydar Watershed, Chaharmahal Va Bakhtiari Province, Iran,” *Journal of the Faculty of Forestry Istanbul University 65(1),* pp. 30-40, 2015.

[2] T.D.C. Pushpakumara, K.G.H.N.R.P Kodagoda, D.M.T Ranasinghe, A.S.J Abeysinghe, U.G.A. Puswewala, Indishe Senanayake, “Analyzing and Predicting of Landslide Prone,” in *Neural Processing Letters, April 2015, Volume 41, Issue 2*, 2015, p. 169–178.

[3] Cees van Westen, “Introduction to landslides Part 1:Types and causes,” International Institute for Aerospace Survey and Earth Sciences (ITC), Enschede, The Netherlands.

[4] “Aranayake Landslide - Hiru News,” Lotus Technologies (PVT) Ltd, May 2016. [Online]. Available: http://www.hirunews.lk/133503/120-missing-in-aranayake-landslide-17-bodies-recovere. [Accessed 15 Jan 2020].

[5] Mihail E. Popescu, “Landslide causal factors and landslide remediatial options,” in *Illinois Institute of Technology, Chicago, USA.*

[6] Varnes, D.J, “Slope Movements And Types And Processes,” *Landslides Analysis and Control. Transportation Research Board Special Report,* vol. 176, pp. 11-33, 1978.

[7] Crozier, M.J., Landslides - Causes, Consequences And Environment, Croom Helm, London, 1986.

[8] Biswajeet Pradhan, Saro Lee, “Landslide risk analysis using artificial neural network model focusing on different training sites,” *International journal of physical sciences,* pp. 03-07, February 2009.

[9] Lippmann, Richard, “An introduction to computing with neural nets,” in *ASSP Magazine*, pp. 4-22.

[10] Paola, Justin & Schowengerdt, Robert, “The Effect of Neural-Network Structure on a Multispectral Land-Use/Land-Cover Classification,” *Photogrammetric Engineering & Remote Sensing.*, pp. 15-20, 1997.

[11] Kalyana, T., Mehta, P., Bansal, R., Parekh, C., Merchant, S.N. et al, “Routing Protocols for Landslide Prediction using Wireless Sensor Networks,” in *Intelligent Sensing and Information Processing*, pp. 43-47.

[12] C. Huangqiong, Z. Zhigang and Huiming Tang, “Landslide Deformation Prediction Based on Recurrent Neural Network,” *Neural Processing Letters*, vol. 41, no. 2, p. 169–178, April 2015.

[13] Emmanuel Léonard, Shou-Hao Chiang, “Landslide Susceptibility Analysis using an Artificial Neural Network (ANN) Model: A Case Study In Yushan National Park, Taiwan,” in *38th Asian Conference on Remote Sensing*, 27 Oct. 2017.

[14] Jie Dou, Hiromitsu Yamagishi, H R Pourghasemi, Ali P. Yunu, Xuan Song, Yueren Xu, Zhongfan Zhu, “An integrated artificial neural network model for the landslide susceptibility assessment of Osado Island, Japan,” *Neural Processing Letters*, vol. 41, no. 2, p. 169–178, April 2015.

[15] Joaquín Andrés Valencia Ortiz & Antonio Miguel Martínez-Graña, “A neural network model applied to landslide susceptibility analysis (Capitanejo, Colombia),” in *Geomatics, Natural Hazards and Risk*, pp. 1106-1128.

[16] T. Jiang and D. Wand, “A landslide stability calculation method based on Bayesian network,” in *Instrumentation and Measurement, Sensor Network and Automation*, Toronto, ON, Canada, December 2013.

[17] Chung C.F. Fabbri A.G, “Probabilistic Prediction Models for Landslide Hazard Mapping,” in *hotogrammetric Engineering & Remote Sensing*, December 1999, pp. 1389-1399.

[18] “Landslide Maps,” National Building Research Organisation, Sri Lanka, [Online]. Available: http://nbro.gov.lk/index.php?option=com_content&view=article&id=25&Itemid=179&lang=en. [Accessed 18 Jan 2020].

[19] “Badulla District,” Wikipedia, 2018. [Online]. Available: https://en.wikipedia.org/wiki/Badulla_District. [Accessed 15 Jan 2020].

[20] N.B.N Prasad, “Landslides-Causes and Mitigation,” Centre for Water Resources Development, Kerala , India, 1995.

[21] A. M. Ritchie, “Chapter 04 - Recognition and Identification of Landslides,” in *Recognition and Identification landslide*, pp. 48-54.

[22] Tutorialspoint, “MATLAB Tutorial,” TutorialsPoint, [Online]. Available: https://www.tutorialspoint.com/matlab/index.htm. [Accessed 10 09 2020].

[23] Halbouty M.T., “D J Varnes, Slope-stability problems of the Circum- Pacific Region as related to mineral and energy resources,” in *Energy Resources of the Pacific Region Vol. 12*, American Association of Petroleum Geologists, Jan 1981 , pp. 489-505.

[24] ESRI, “Comparing interpolation methods,” Institute, Environmental Systems Research, [Online]. Available: https://pro.arcgis.com/en/pro-app/tool-reference/3d-analyst/comparing-interpolation-methods.htm. [Accessed 10 June 2020].

[25] Stephanie Glen, “ StatisticsHowTo.com: Elementary Statistics for the rest of us!,” [Online]. Available: https://www.statisticshowto.com/probability-and-statistics/regression-analysis/rmse-root-mean-square-error/. [Accessed 15 05 2020].

[26] Miller C L ,R A Laflamme, “The Digital Terrain Model - Theory and Application,” in *Photogrammetric Engineering, Vol.XXIV, No.3*, 1958 June, p. 433.

[27] Katharine Brown, Andrew Wherrett , “Bulk Density – Measurement,” [Online]. Available: http://soilquality.org.au/factsheets/bulk-density-measurement. [Accessed 20 May 2020].

[28] Wikipedia, “Receiver operating characteristic,” Wikipedia, the free encyclopedia, [Online]. Available: https://en.wikipedia.org/wiki/Receiver_operating_characteristic. [Accessed 25 09 2020].

[29] Radbruch-Hall D.H., Varnes D.T., Landslide: cause and Effect, vol. 14, Int. Association of Engineering Geology Bulletin, pp. 205-216.

[30] H. A, “Landslide Hazard Analysis,” in *Slope Instability*, Wiley and Sons, New York, pp. 523-602.

[31] Van Western C.J., “GIS in Landslide Hazard Zonation: a review with examples from the Andes of Colombia,” in *Mountain Environments and Geographic Information*, London, pp. 135-165.

[32] “Analytic hierarchy process,” Wikipedia, 2000. [Online]. Available: https://en.wikipedia.org/wiki/Analytic_hierarchy_process. [Accessed 15 01 2020].

[33] ESRI, “Map Map Use: Reading, Analysis, Interpretation, Seventh EditionUse: Reading, Analysis, Interpretation, Seventh Edition,” p. 360.

APPENDIX A

List of past landslide data in Badulla district - Sri Lanka

Year	Month	Location	Type	No or Failures	Rainfall/day
2012	February	Between culverts 51/10 and 51111 on Kandy	Landslide	1	160mm
2012	December	Welangahawatta	Landslide	1	120mm
2011	January	Udunuwara, Yatinuwara, Gangawata Korale,Patha	Cut slope failures		
2011	January	Rattota, Ukuwela	Landslide	1	98mm
2010	December	Udunuwara, Yatinuwara, Gangawata Korale,	Cut slope failures and	25	122mm
2009	January	Undunuwangoda,	Rock fall and potential	1	
2009	December	Lindula maussa ella estate	Landslide	1	
2008	June	Kotagala	Landslide	1	119mm
2008	August	Mathurata, Hanguranketha	Creep	1	101mm
2007	January	Walapane & Hanguranketha	Landslide	131	95mm
2006	December	Galahitiyawa, Janathapura	Reactivated old landslide	1	120mm
2006	December	Batawatta Group, Division 4, Batawatta GN, Lunugala	Debris flow		
2004	December	Kapuruwatta , Boragas,	Cutting failure	1	89mm
2004	December	Rendapola/Siliniyapura/W	Cutting failure	1	103mm
2002	April	Demodara	Landslide	1	100mm
2002	April	Sarnia estate	Landslide	1	94mm
2002	November	Bathgoda, Haldummmulla	Landslide	1	134mm
1999	October	Between 17 km post and 18 km	Landslide	1	
1997	November	Naketiya, Koslanda	Landslide	1	161mm
1995	April	Maussagolla, Passara	Landslide	1	126mm
1994	August	Mabak:umbura, Nawalapitiya	Landslide	1	99mm

1994	August	North Pundalu Oya village	Debris flow	1	
1994	August	2 km away from Nawalapitya on	Rock fall	1	102mm
1993	October	Beragala junction	Cutting failure	1	
1993	October	Koslanda	Rock fall & debris flow	1	108 mm
1993	December	Ambolike	Cutting failure	1	
1993	December	1.5 km away from Namunukula on	Rock fall	1	
1993	December	Soranatota junction	Cutting failure	1	103mm
1993	December	Hotel Hilltop, Pilimathalawa, Kandy	Earthslide	1	89mm
1993	June	Watawala	Earthslide	1	125mm
1992	November	Viharagala	Debris flow	1	162mm
1992	June	Watawala			146mm
1992	November	Pattipola	Landslide	1	
1990	June	Uthuwankanda , Kadugannawa	Landslide		148mm
1989	May	Diyabibila, Thalawakele		1	
1989	June	Paraiyagala, Florence area		1	220 mm
1987	May	Beragala		1	
1986	January	Uduwara		1	90mm
1986	January	Badulusirigama and Kirimetiya		2	
1986	January	Katupallegama, Mahakele, Agaratenna		3	
1986	January	Ambaliyadda and Rupaha		2	182 mm
1986	January	Ketyapathana, Maturata		1	266.6 mm
1986	January	Mandaramnuwara		1	130 mm
1986	January	Kurupanawela		1	160 mm
1986	January	Madulla		1	182 mm
1986	January	Watumulla hospital, Mulhalkele		1	185 mm
1986	January	Marabedda, Walapane		1	

APPENDIX B

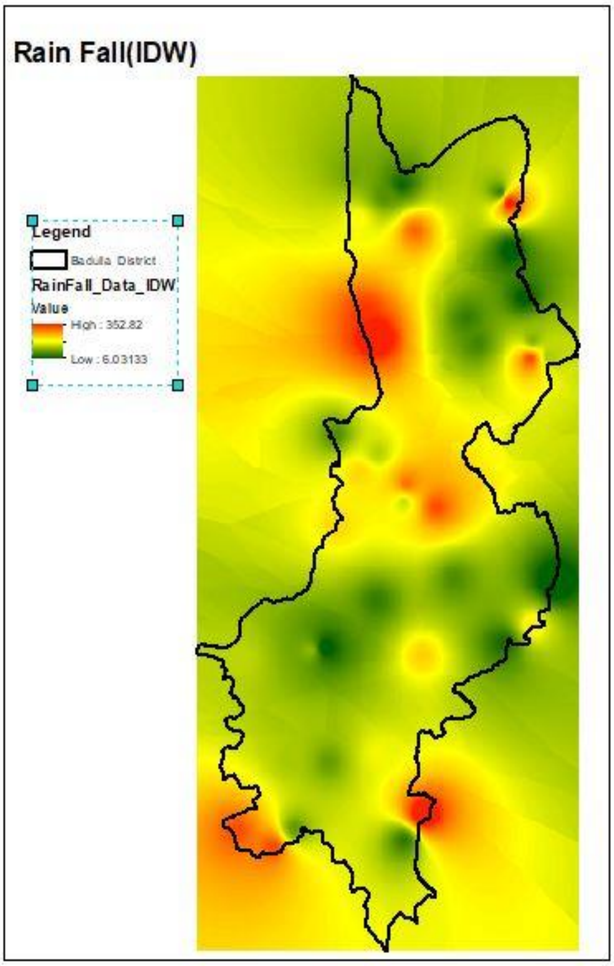
ANN Script

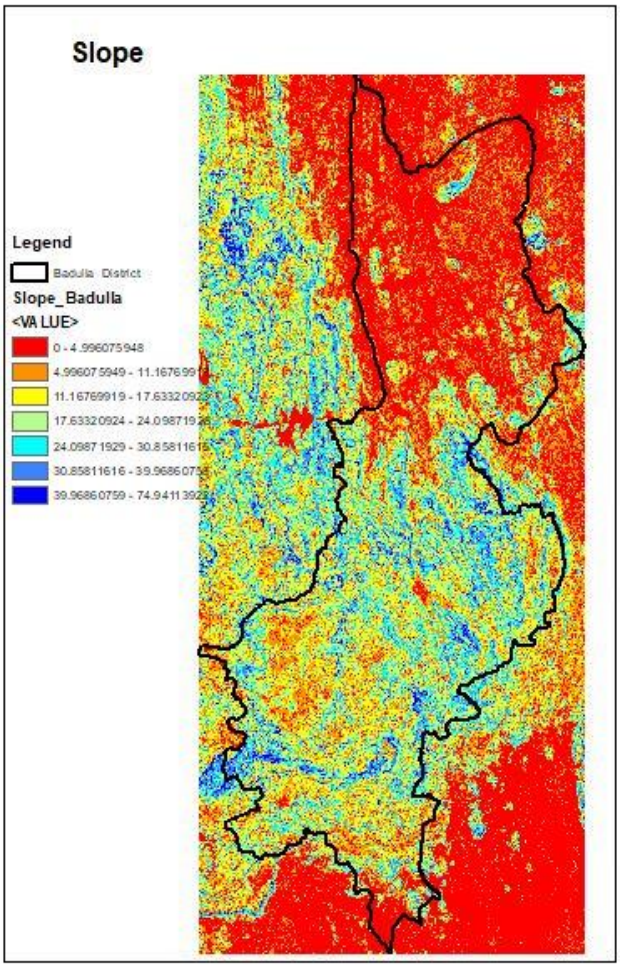
```
% Landslide Pattern Recognition using MATLAB with Artificial Neural Network
% Optimized for identify landslide occurrence pattern with use of eight
selected factors.
% Training data of each sample are simplified/standardized in order to
improve the accuracy level


% Variables for input-x(landslide factor data) and target-t data
% t - 1 0 - landslide occurrence event
% t - 0 1 - landslide nonoccurrence event

x = input;
t = target;

% Training Function
% trainlm , trainbr , trainscg , traingdx - with learnin rate
trainFcn = 'trainscg';  % Scaled conjugate gradient backpropagation.


% Creating Nural Network
hiddenLayerSize = 10;
net = patternnet(hiddenLayerSize);


% Input and Output Pre/Post-Processing Functions
net.input.processFcns = {'removeconstantrows','mapminmax'};
net.output.processFcns = {'removeconstantrows','mapminmax'};


net.divideFcn = 'dividerand';  % Divide data randomly
net.divideMode = 'sample';  % Divide up every sample
net.divideParam.trainRatio = 70/100;
net.divideParam.valRatio = 15/100;
net.divideParam.testRatio = 15/100;

% Performance Function
net.performFcn = 'crossentropy';  % Cross-Entropy

% Defining learning rate of the network
% net.trainParam.lr = 0.1;


% Plot Functions
net.plotFcns = {'plotperform','plottrainstate','ploterrhist', ...
    'plotconfusion', 'plotroc'};


% Train the Network
[net,tr] = train(net,x,t);
```

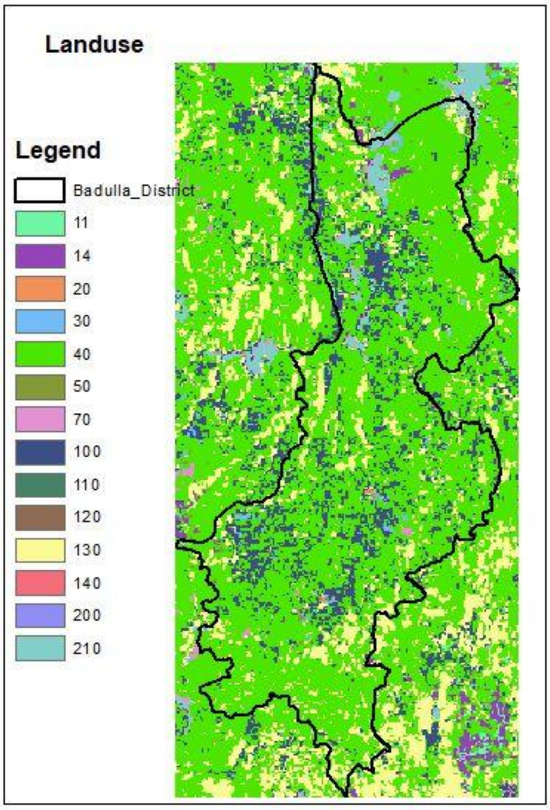
```
% Test the Network
y = net(x);
e = gsubtract(t,y);
performance = perform(net,t,y)
tind = vec2ind(t);
yind = vec2ind(y);
percentErrors = sum(tind ~= yind)/numel(tind);

% Recalculate Training, Validation and Test Performance
trainTargets = t .* tr.trainMask{1};
valTargets = t .* tr.valMask{1};
testTargets = t .* tr.testMask{1};
trainPerformance = perform(net,trainTargets,y)
valPerformance = perform(net,valTargets,y)
testPerformance = perform(net,testTargets,y)

% View the Network
view(net)

% Plots
%plotperform(tr)
%plottrainstate(tr)
%ploterrhist(e)
%plotconfusion(t,y)
%plotroc(t,y)



% Deployment
% Change the (false) values to (true) to enable the following code blocks.
% See the help for each generation function for more information.
if (false)
    % Generate MATLAB function for neural network for application
    % deployment in MATLAB scripts or with MATLAB Compiler and Builder
    % tools, or simply to examine the calculations your trained neural
    % network performs.
    genFunction(net,'myNeuralNetworkFunction');
    y = myNeuralNetworkFunction(x);
end
if (false)
    % Generate a matrix-only MATLAB function for neural network code
    % generation with MATLAB Coder tools.
    genFunction(net,'myNeuralNetworkFunction','MatrixOnly','yes');
    y = myNeuralNetworkFunction(x);
end
if (false)
    % Generate a Simulink diagram for simulation or deployment with.
    % Simulink Coder tools.
    gensim(net);
end
```

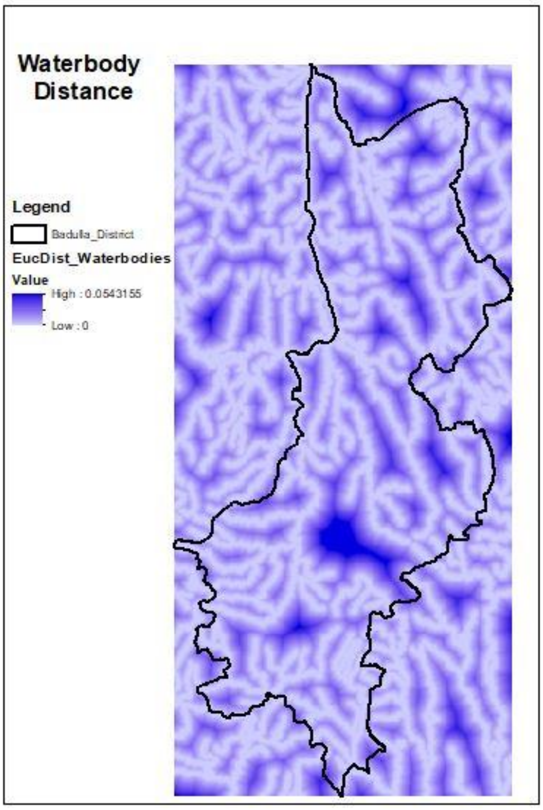
APPENDIX C

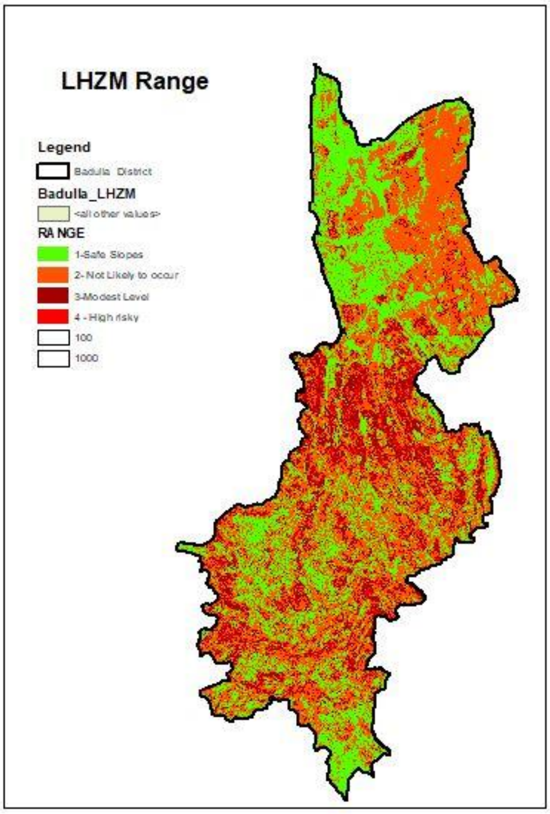
Factor Data GIS Images

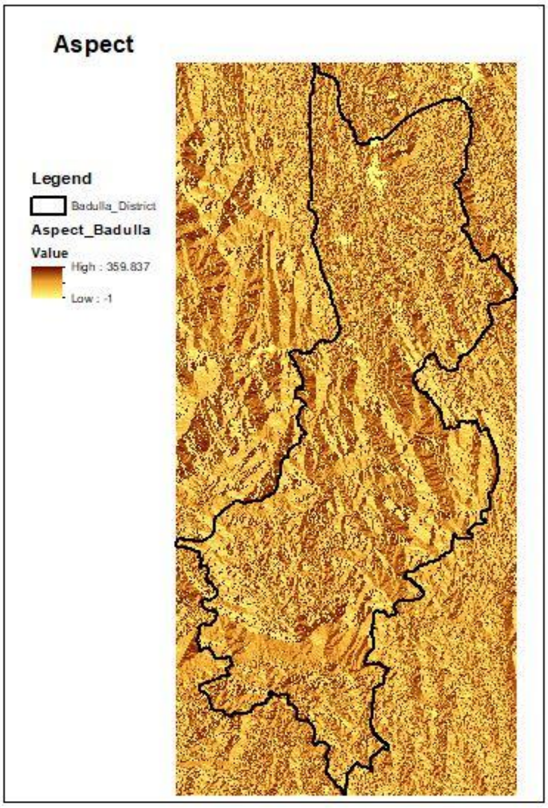





Landuse
Legend
Badulla_District
11
14
20
30
40
50
70
100
110
120
130
140
200
210


Waterbody Distance
Legend
Badulla_District
EucDist_Waterbodies
Value
High : 0.0543155
Low : 0


LHZM Range
Legend
Badulla District
Badulla_LHZM
<all other values>
RA NGE
1-Safe Slopes
2- Not Likely to occur
3-Modest Level
4 - High risky
100
1000


Aspect
Legend
Badulla_District
Aspect_Badulla
Value
High : 359.837
Low : -1