AI & Machine Learning based early stage Landslide detection using GIS

By

Mehak Eman

2016-NUST-MSIT-103711

Supervisor

Dr. Sharifullah Khan

A thesis submitted in conformity with the requirements for
the degree of *Master of Science* in
Information Technology

Department of Computer Science

School of Electrical Engineering and Computer Sciences (SEECS)

National University of Sciences and Technology (NUST)

Islamabad, Pakistan

Certificate of Originality

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any degree or diploma at NUST SEECS or at any other educational institute, except where due acknowledgement has been made in the thesis. Any contribution made to the research by others, with whom I have worked at NUST SEECS or elsewhere, is explicitly acknowledged in the thesis.

I also declare that the intellectual content of this thesis is the product of my own work, except for the assistance from others in the project's design and conception or in style, presentation and linguistics which has been acknowledged.

Mehak Eman

2016-NUST-MSIT-103711

I am dedicating this thesis to *my beloved Family*

Abstract

Number of population and areas affected by natural disasters are increasing day by day due to lack of proper planning and response/regularity authorities. Better practices and models must be adopted in order to reduce human and economic loss due to such disastrous events. Landslide is one of natural disaster happens due to downward and outward movement of rock debris and earth materials resulting in vibrations which blocks drainage and roads. In Pakistan, northern areas comprises of regions susceptible to landslides due to extensive mountains and rugged terrain. Landslides disaster can lead to enormous casualties and loss of economy. Landslide hazard mitigation can be done effectively with the help of new methodologies, that can develop better landslide hazard understanding and help to make rational decisions for management of landslide risk. The primary objective of this study is i)To identify factors influencing occurrence of landslides, through a quantitative methodology ii)To identify Artificial Intelligence and Machine Learning based models that are effective in detecting landslides, iii) Creation of landslide inventory map data for landslide modeling using open source resources. In our research work, we are proposing to identify landslide inventories using Satellite imagery and field data, calculating susceptibility analysis using Geographical Information Systems (GIS) tools for Muzaffarabad area and then based on that data we will detect landslide prone areas in other regions using Machine Learning tools. For that purpose we will be using Analytical Hierarchy Process (AHP) for landslide parameters determination and Support Vector Machine (SVM), Linear regression, Decision Tree, K nearest neighbor (KNN) classifiers and Neuro Evolutionary Algorithm named as Cartesian Genetic Programming Artificial Neural Network for landslide susceptibility. The results shows that almost 90% accuracy when correlate with the landslide inventories. In this project we focused on susceptibility of landslides using Geographical Information Systems (GIS) tools, machine learning techniques and Artificial Intelligence algorithm.

Using Analytical Hierarchy Process we identified landslide prone parameters to be used in susceptibility analysis which were later divided into four categories, i.e. low , Moderate , High and very High landslide prone areas. The analysis show that almost 30% of the area comes under high and very high prone areas of landslides. Barren land and grassland in land cover, fault lines and Muzaffarabad formation, Hazara formation and Holocene in geology are found to be most susceptible factors in Muzaffarabad areas which contributes to landslides.The classification results show the model performance. In the given analysis, Cartesian Genetic Programming Evolved Artificial Neural Network (CGPANN) shows better performance as compared to others, Support Vector Machine, K Nearest Neighbours (KNN) and Logistic regression performance is also good.The performance score shows 0.81 for knn, 0.83 for Decision Trees, 0.85 for Support Vector Machine and 0.87 for Logistic Regression. Cartesian Genetic Programming Evolved Artificial Neural Network outperformed other techniques like SVM & Logistic Regression with 0.96 accuracy. Our proposed methodology will help the government to improve the landslide prediction system and utilize available professional resources efficiently in order to deal with the situation of increasing occurrence of landslides in Pakistan. In future we are going to extend this work by installing sensors and cameras, developing heterogeneous sensor network along with Artificial Intelligence algorithms that are effective in developing new landslide reduction services to predict landslides beforehand to save the community and infrastructure from big losses.

Keywords: *Analytical hierarchy Process, Landslide, Susceptibility, Machine Learning, Support Vector Machine, Logistic Regression, Decision Tree*

Acknowledgments

I am grateful to Allah who granted me success and it was impossible without His blessings. Dr. Sharifullah Khan and my GEC members have always been very kind and helpful to me. I want to thank them for their support and cooperation. I want to extend my gratitude to all my lovely friends and family for supporting throughout the degree. I would like to pay my thanks to everyone who directly or indirectly contributed in this journey.

Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

Abbreviations

GIS	Geographical Information Systems
AHP	Analytical Hierarchy Process
SVM	Support Vector Machine
AI	Artificial Intelligence
CGPANN	Cartesian Genetic Programming evolved Artificial Neural Network
WHO	World Health Organization
UN	United Nations

Chapter 1

Introduction

Landslides disaster can lead to extensive economical losses and immense human causalities in mountainous regions [1]. Because of the damage, millions of dollars and hundreds of lives are lost each year. To cater for this problem, we introduce an approach to identify landslide susceptible areas in Muzaffarabad, Pakistan where a lot of landslides had taken place in the past. In this chapter, a brief introduction of landslide, landslide occurrence all over the world and in Pakistan, motivation, problem statement and research objectives has been explained.

1.1 Background and Motivation

Landslide is a natural phenomenon which means downward motion of slope in the form of a rock mass, debris, or soil [2]. There are destabilizing forces which weakens the mountains slope. In return slopes strive to achieve its stability by grasping a natural state of equilibrium under a particular arrangement of conditions but gets unstable and weakens once the equilibrium that is interrupted thus it leads to landslide occurrence [3]. They have caused destruction and loss to life of human beings and infrastructure worldwide.

Out of all natural disasters, it is ranked as seventh killer in number when compared with other natural disasters [4]. They are named seventh killer after volcano, floods, drought, windstorms, earthquakes, and extreme temperature, claiming lives from 800-1000 on average during the 20 years [5]. Among all the continents, 220 landslides occurred in Asia only during the last three decades which is by far the most when compared to other world regions. However, deaths and injuries are being suffered by America the most, as the number is greater than 25000 while Europe suffered a huge damage that cost them the damage of almost $23 million [6]. Tens of billions of dollars is estimated to be the infrastructure loss damage worldwide for landslides with annual losses alone of more than $1.5 billion in USA alone [7]

1.2 LANDSLIDES IN PAKISTAN

There are many ways to identify and analyze landslides, some of them are landslide monitoring system that are sensor based [8] [9] [10], and others are through artificial intelligence landslide prediction[11] [12][13]. In Pakistan, northern areas comprises of regions susceptible to landslides due to extensive mountains and rugged terrain [14]. The Karakoram, the Himalayas, and the Hindu Kush mountain ranges are located in Pakistan and have faced many landslides which have been resulted from the shifting of the Indian plate northward and Asian Plate clashing with them [15]. These ranges extents are seismically dynamic locales amidst two impacting continents [16]. Landslides are the most common risk in the Himalaya and Karakoram ranges due to its high dynamic seismicity, huge topographic geomorphology, eroded land and massive precipitation factor etc. Pakistan has experienced several catastrophic landslide disasters in recent years.In April 2016, a massive landslide in Khyber Pakhtunkhwa (KP) killed at least 262 people, with 56% of the fatalities [17].It is very important to understand the factors related to landslides in order to solve complex problems related to landslides. There are many factors that influence landslides, most of them are discussed below.

1.3 FACTORS CAUSING LANDSLIDE

Parameters for landslides are very important as they contribute towards landslide occurrence. There are many factors that contribute to landslides. They are divided into two categories. Natural Factors and Anthropogenic factors.

1.3.1 Natural Factors

Slope

One of the most important factors in landslide susceptibility is Slope [18]. Slope is a gradient. The more steep it will be, the more it is likely that landslide will occur. Due to availability of Digital Elevation models, it is now more convenient to get Slope from them even of remote areas.

Rainfall

Precipitation is also an important and common factor for triggering landslides. There is a direct relationship between precipitation and landslide occurrence. The more the rainfall, the more soil will get moisture and the more cohesion the soil will lose as the soil will get saturated hence the landslide. In Pakistan during the monsoon season due to heavy precipitation and antecedent rainfall, the soil gets a lot of moisture and pore pressure increases and water infiltration is high thus a lot of landslides had been recorded over the years[19].

Erosion

Due to weathering, intermittent running water like river streams etc, high winds and snowfall the top part of soil is removed that kept the soil binded thus slope becomes steep and landslide occurs [20].

Earthquakes

Due to shaking of earth vibration occurs and sometimes its very devastating. Earthquake put strain on the soil and thus weak slope fails and landslide occurs[21].

1.3.2 Anthropogenic Factors

Deforestation

Due to urbanization, a lot of construction has been occurring especially in mountain regions, bridges and roads constructions are taking place rapidly. Trees are being cut down for constructions in urban areas which has a drastic impact on stability of land[22].

Vibrations

Vibrations due to movements of vehicles, construction etc. lead to instability in land and thus creates a landslide[23].

1.4 Problem Statement

The landslides processes are the extensively damaging, among all the other disasters, as due to land sliding a lot of homes and properties are devastated as everything on that area comes down. Because of the damage millions of dollars and hundreds of lives are lost each year. In order to identify landslide prone factors relation to landslide occurrence, it is important to find techniques that can help in landslide feature engineering. Foremost step in landslide detection is to find out accurately which factor weighs more in landslide occurrence. Once we know the parameters which causes the landslides, we can easily predict landslide in the future. Secondly, prediction of landslide with high accuracy is very important as less accuracy can lead to wrong results which will further cause great damage to people and infrastructure, if they are wrongly been informed. For that reason, prediction of landslide prone areas with high accuracy is needed to reduce the damage caused by landslides and also to mitigate planning and development strategies. Previously, different researches have been done (see section 3.1,3.2 of Chapter 3) to identify susceptibility of landslides using different techniques but the accuracy of susceptibility results could not achieve the high mark which could help in saving the community.

1.5 Research Objectives

This research has two main objectives. The first objective is identify AI and Machine Learning based models that are effective in detecting landslides with high accuracy . The second objective is to create landslide inventory map data for landslide modeling. They are briefly stated as follows:

1. **To identify factors influencing occurrence of landslides**

2. **To identify AI and Machine Learning based models that are effective in detecting landslides with high accuracy**

3. **Creation of landslide inventory map data for landslide modeling.**

1.6 Proposed Framework

In our research work, we are proposing to identify landslide parameters used for landslide prone area detection.For that reason we are proposing novel method for feature engineer using GIS tools and decision making techniques. Firstly, we are using frequency ratio to find parameters of landslides and then after we get to know about the weights of parameters we will use Analytical Hierarchy process to find susceptible areas of landslide classified into Low landslide, Moderate Landslide,High landslide and very high landslide prone areas called landslide susceptible areas. Using that information we will detect other areas landslide susceptibility using Artificial intelligence and Machine learning Algorithms. Initially the primary data for landslide susceptibility was extracted using Satellite imagery, Digital Elevation model, and GPS points of landslide inventory. Parameters were then extracted using GIS tools and softwares. Later they were converted into numeric data to be used in Machine Learning and AI models.

1.7 Organization of Thesis

Chapter 2 contains the background information of Analytical Hierarchy Process, and Machine learning Classifiers. Chapter 3 provides the literature review carried out on susceptibility analysis and landslide detection. Chapter 4 describes the proposed framework of this research and workflow of solving the problem. Chapter 5 describes the results and Chapter 6 will describe the discussion, conclusion limitations and future work of this research.

Chapter 2

Background

This chapter describes the basic concepts of the models being used in developing the Framework. Three models are used in the proposed framework. The models are Analytical Hierarchy Process, Support Vector Machine. Linear regression, Decision Tree and K nearest neighbor.

2.1 Analytical Hierarchy Process

AHP uses multi-criterial matrix based on ranking of experts. This method is used in decision making process for example site selection, disaster management etc. In the recent years several researchers have been using this technique as it proves to be a convenient procedure that deals with multi-criteria hierarchical structures [24]. Pairwise comparisons matrix are involved in this process against decision variables[25]. The factors of landslides are assigned a numeric value from 0-1 individually where 1 is the least and 9 is the most, depending on their relative rank. It shows the intensity of importance of landslide factors.The table introduced by Saaty is given below.

Intensity of Importance	Definition	Description
1	Equal importance	Two activities contribute equally to the objective
2	Equal to moderate importance	Used to represent compromise between 1 & 3
3	Moderate importance	Experience and judgment slightly to moderately favor one activity over the other
4	Moderate to strong importance	Used to represent compromise between 3 & 5
5	Strong importance	Experience and judgment strongly or essentially favor one activity
6	Strong to very strong importance	Used to represent compromise between 5 & 7
7	Very strong importance	Experience and judgment strongly favored one activity over the other and its dominance is shown in practice
8	Very strong to extremely strong importance	Used to represent compromise between 7 & 9
9	Extremely strong importance	The evidence of favoring one activity over the other is of highest degree possible of an affirmation

Figure 2.1: Intensity Importance

Pairwise comparison was done using matrix. For its preparation, the utilization of comparison matrix was done using matrix of landslide causative parameters, knowledge by the experts and existing research to determine the ranks among its parameters. Using the table developed by Saaty, for this research a pairwise comparison matrix was prepared to identify prone areas. Weights of criteria were computed by each column values' sum of matrix of pair-wise comparison, further it divided each matrix element by its column sum. The mean of elements was computed in row of each element. Several researchers have applied this technique for landslide susceptibility mapping [26][27][28] to find landslide prone areas by giving ranks to landslide causative parameters. The consistency of judgement in this process is improved by inconsistency measurement. The formula for consistency index (CI) for comparison of matrix is given below.

2.2 AI and Machine Learning Techniques

2.2.1 Support vector Machine

A separator is used in support vector machine to classify the data. First the data is mapped into high dimension feature space to categorize the data, then the separator is drawn for the data to be separated. The separator is in the form of a hyperplane[29]. SVM was selected in this research because it gives good accuracy with limited data. The image below shoes the data divided by hyper plane in SVM methodology [30].

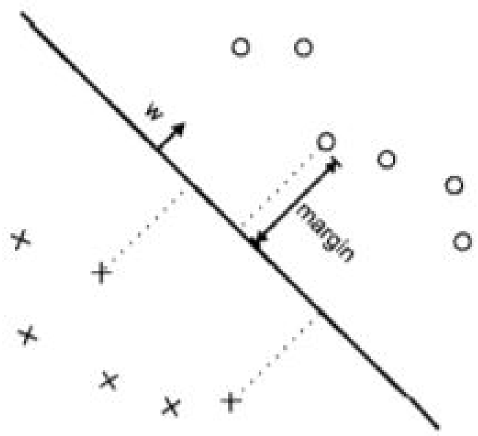

Figure 2.2: Support Vector Machine

2.2.2 Decision Trees

Decision Tree: In decision trees training set are split into nodes, where one category of data is contained in one node. Construction of decision tree can be done by considering the attributes explicitly[31]. First data is chosen from data one by one then each parameter is split in tress by its significance. Decision trees are used in this research because of their simplicity and less computational power.

2.2.3 Logistic Regression

McFadden (1973) introduced Logistic Regression model. In logistic regression, there is one or more independent variables. Logistic regression is an analogous to linear regression, but tries to predict a categorical or discrete target field instead of a numeric one[32].Linear regression works by considering the feature set and a bias term that can be fitted into training set. Regularization in linear regression assist in model fitting prevention

2.2.4 K Nearest Neighbour

The K-Nearest Neighbors algorithm takes the data that is labeled and use the signature to label the ones that are not labeled. It classifies the data by seeing its similarity to other datasets. In KNN the data points closest to one another are considered neighbors and label the other points according to this paradigm[33]. We used KNN in our project because it works better with static data.

2.2.5 Cartesian Genetic Programming

Cartesian Genetic Programming (CGP) initially developed by Miller and Thomson is an Evolutionary Programming method. It works as a feedforward connected network. It is inspired by genes in human body where genotype consists of neurons called nodes which have inputs and functions. In CGP genotypes are evolving from first generation to next. In this we have parents genotype and later offspring genotypes that are produced after mutation.To add more flexibility to NeuroEvolution based cartesian programming, artificial neural networks are evolved in it. It takes direct encoding scheme which means architecture,weights, functions are encoded by CGPANN in one genotype which then optimize it to the best possible option through mutation by adding and removing features based on its best suitability [34].

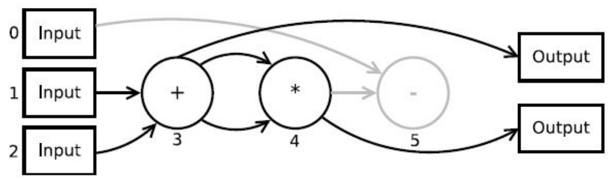

Figure 2.3: Cartesian Genetic Programming Architecture Diagram

In this figure above shows three inputs i.e. 0,1,2, three nodes in which two of them are active (black) and one is inactive (grey) and an output.

Chapter 3

Literature Review

In this chapter, we present our conclusion on previously accomplished analysis and research regarding landslide susceptibility and its prediction using different techniques. With the assistance of related work, we can get deep knowledge of our work with which we can further refine our research problem, assists us in carrying out methodology, guide us in relevance of our work and deduce the research gap.

3.1 Landslide Susceptibility Parameters

Spatial landslide prediction with the assistance of landslide parameters that caused the landslides previously are refer to as landslide susceptibility, which aims to identify landslide occurrence over a region [35]. However, there are some basic assumptions that are being applied whenever landslide susceptibility mapping is done. Firstly, it implies that landslides that going to happen afterwards will happen under similar geographical, geomorphological, hydrological and climatic states which were responsible in the past. Secondly, through remote sensing and field survery, distinct features can be identified for landslides [36]. Thirdly, landslides are results of identifiable internal factors [37].

Over the last few years, developments of GIS data for spatial analysis have played a significant role. Tools of GIS and Remote sensing techniques have been widely used for landslide detection and prediction. Comparative analysis of different landslide parameters used in different research papers and evaluation of different landslide susceptibility mapping is done in this literature review.

Factors of landslides depend on few things like study purpose, scalability and availability data. Landslide causative factors can be many, it all depends on the research objective [38]. Different researchers have chosen different parameters. KT Change et.al [39] worked using statistical and machine learning techniques to produce landslide maps for susceptibility in which facts influencing landslides were twelve parameters. Jaewon Choi et al. 2012 [40] proposed to obtain landslide factors for landslide susceptibility using satellite images. Factors like bedrock and surface lithology, bedding attitude, structure, conditions of in ground water, vegetation cover, climate, land use and human activity were included in the research. They removed lithology from the paper as the information of lithology cannot be extracted directly from the imagery. Biswajeet Pradhan et.al (2010) [41] presented neuro-fuzzy inference system to analyze landslide causative factors to produce susceptibility mapping. Factors like altitude, curvature, slope gradient, distance from road, distance from drainage, lithology, Normalized Difference Vegetation Index, distance from faults were identified using previous papers. The study depicted that distance to road have close relation to landslides. Also slope is directly proportional to landslides in the given study. Thus it makes slope an important parameter to assess landslides. Aafaf El Jazouli et al. (2019) [42] proposed Analytical Hierarchy Process for assessment of landslide causing factors and their effects on its parameters. Among the parameters of landslides slope, drainage, land use, lithology, slope aspect, roads, elevation and faults were included. The reason to include these factors as they intervene with stabilization of rocks and are expose to landslide susceptibility. After analysis, it showed that distance to faults, slope, distance to drainage network and lithology were considered most important. Kuan-Tsung Chang et al. (2019) [43] selected parameters like topography, hydrology, tectonics, geology, and geomorphology based on aforesaid summarization of spatial relationships between landslide causing factors and its occurrences. Tectonics were later removed as the research was rainfall triggered landslide susceptibility. Sajid Ali et al. (2019) [44] selected factors like seismicity, lithology,rain fall intensity, elevation, slope, faults, aspect, curvature, hydrology and landcover. Based on statistical and spatial analysis results it was discovered that seismicity, faults and slope angle mainly control the landslide's spatial distribution. Himam Shahabi et al. (2015) [45] selected parameters on landslides by keeping in mind the study area and data availability. They presented ten factors i-e slope, soil, aspect,lithology,Normalised DVI,distance to drainage,landcover,precipitation, distance to faults and distance to roads.

In landslide susceptibility mapping causative factors play a crucial part. Almost all of the research papers have included Slope, in landslide susceptibility mapping. Lithology and Aspect are being used by many researchers as well. The parameters are used by every researcher according to the research goad and study area. Landuse and buffer of roads (human induced), then factors induced by water content e.g. streams, rainfall etc. and finally geomorphic factors that are induced by nature i-e weathering or earthquake e.g. Faults, lithology, soil etc. Factors are chosen according to the goal of the research and availability of data.

3.2 Landslide Susceptibility Mapping Approaches

Landslide Susceptibility Mapping is a multiplex task which helps in mitigation, and planning [46] [47]. There are multiple ways that are being proposed to map landslide prone areas. They are as follows:

3.2.1 Qualitative Approach

Expert knowledge and past experiences in landslide susceptibility assessment lead to Qualitative approach. Geomorphic analysis and Distribution analysis were widely used during 1970-1980 time. different types of data are introduced in qualitative methods based on subjectivity to generate landslide susceptibility mapping.

3.2.2 Distribution Analysis

Distribution analysis refers to as spatial location of landslides of a particular area using aerial imagery, historic landslide data and field survey as a polygon on point event [48]. Landslide susceptibility mapping is produced on the basis of landslide inventory. Distribution analysis also refers to as density map of landslides that shows frequency of landslides and their loss alongside to determine susceptibility of landslides by interpolating density data [49]. However, this method of analysis could not define the relationship between causative factors of landslides and landslide occurrences, moreover, it was time consuming and expensive. In conclusion, distribution analysis could not find landslide susceptibility for the future landslide occurrence.

3.2.3 Geomorphic Analysis

Geomorphological mapping of landslides is done through expert ability to approximate existing and prospective slope failures [35] . This is a direct method where field work is involved by the experts based on their subjective knowledge of the area [50] produced map based on geomorphic analysis for landslide mapping using causative factors of landslides to extrapolate for area having similar physical characteristics to map different level of susceptibility [51]. However, the drawback of these kind of analysis is that subjectivity is involved as well as extensive field data is required [52].

In Quantitative approach expert opinion plays a vital role. It is a knowledge driven method where experts directly predicts the susceptibility based on the observations and history of that particular area. Techniques like Analytical Hierarchy Process, Boolean overlay, Multi-class overlay, Spatial evaluation based on a criteria are developed to make qualitative susceptibility map. AHP uses multi-criterial matrix based on ranking of experts. This method is used in decision making process for example site selection, disaster management etc. In the recent years several researchers have been using this technique as it proves to be a convenient procedure that deals with multi-criteria hierarchical structures [24]. The factors of landslides are assigned a numeric value from 0-1 individually where 1 is the least and 9 is the most, depending on their relative rank.

Aafaf El Jazouli et al. [42] proposed decision making technique called AHP for assessment of landslide causing factors and their effects on its parameters. Spatially each parameter was analyzed based on expert judgement. The result showed the percentage of different susceptibility prone areas classified into differnt risk factors. The validation was done through Receiving Operating Characteristics method. They were successful in identifying present and future landslides. The results of susceptibility maps will help in planning and management for decision makers.

Sajid Ali et.al [44] selected 11 landslide factors to be assigned its weights according to its rank based on expert opinion given in the previous literature. Based on statistical and spatial analysis results, it was discovered that seismicity, faults and slope angle play significant role in landslide's location when it came to selecting landslide parameters. They were assigned weights in Analytical Hierarchy Process technique which resulted in susceptibility map that was grouped into different categories. The result showed the prone area of landslide in percentage which is likely to getting landslide in the future.

For validation ROC curve was used which resulted in accuracy of 72% which is suitable for future planning, management and mitigation.

Himam Shahabi et al. [53] utilized a decision making technique called AHP. The credibility of the map was validated by using ROC method and R-Index method.

Shamsa Kanwal et al [13] [54] proposed in her paper the use of GIS and AHP to determine the weights of parameters for landslide identification and generate susceptibility maps. Basic landslide causing factors were used in this research. Parameters were further grouped into susceptibility categories and ranked them based on significance slope in stability hence the landslide. The ranks were given to each class based on literature studies. Validation was done using landslide inventories.

3.2.4 Quantitative Approach

Quantitative approaches were used to minimize the use of subjectivity in weight allocation techniques and maximize objectivity in quantifying the considerate signification of causative factors of earthquakes to be used in landslide susceptibility mapping. There are a number of techniques that are quantitative approaches in prone areas mapping.

3.2.5 Statistical Analysis

Statistical analysis juxtapose the landslides causative parameters with existing spatial landslide distribution [55]. It minimizes subjectivity when assignments of weights on parameters is done. GIS tools and technologies are useful when dealing with statistical analysis.They are classified into two classes.

3.2.6 Bivariate Statistical analysis

In bivariate analysis each topographic layer is assessed in comparison to landslide spatial distribution. The value of weights to be assigned to each category of landslide parameters is in accordance of landslide frequency. Thus landslide density layer which is its spatial frequency is overlaid on GIS thematic layers where respective landslide density values are calculated.

Many researchers have used statistical methods to find susceptibility ranks (Brabb 1984) [46], information value method [56] [57] and weights of evidence modeling method [58].

Mukhiddin Juliev et al. (2019) [59] utilized different statistical analysis to derive susceptibility maps of landslide prone areas. In their research they compared different statistical methods to be employed for the purpose of landslide susceptibility. Factors of landslide were used. The results were classified into five groups od susceptibility. Among all SI resulted in better accuracy i-e 80%. For future study they would research on observing the influence of statistical landslide susceptibility mapping on type of landslide inventory. Moreover, they also wanted to extend the area of their research with the help of local authorities.

3.2.7 Multivariate statistical Analysis

In Multivariate statistical analysis the weighted numbers of landslide causal factors confined to land unit. The association between factors are also part of the analysis.

3.2.8 Probabilistic Approach

Probabilistic techniques in landslide susceptibility mapping is used to diminish subjectivity during weight allocation procedure. In this approach comparison is done for landslide spatial distribution with respect to causative factors within the framework of probability. Methods like conditional probability, Bayesian probability model are included. Kuan-Tsung Chang et al. (2019) [60] selected parameters like topography, hydrology, tectonics, geology, and geomorphology based on aforesaid summarization of spatial relationships between landslide causing factors and its occurrences. Tectonics were later removed as the research was rainfall triggered landslide susceptibility. They utilized different Machine learning algorithms to identify landslides prone areas.

Ataollah Shirzad et.al (2017) [61] presented ensambled hybrid approach in order to identify landslide susceptibility map. Subsets were generated from the data gathered in training. NBT was used to produce a classifier which is utilized by a subset.

KT Change et al. (2009) [60] produced landslide susceptibility maps using statistical and machine learning techniques in which facts influencing landslides were twelve. The factors were based on susceptibility. Later they rule out faults from their data as it was not corresponding to landslide occurring while they realized rainfall is the main factor to trigger landlide in their area.

Jaewon Choi et al. (2012) [62] proposed to obtain landslide factors for landslide susceptibility using Satellite images. He referred in his paper that identification and mapping of parameters of landslide require a priori information and knowledge of the causative landslides . Factors like bedrock and surface lithology, bedding attitude, structure, conditions of in ground water, vegetation cover, climate, land use and human activity were included in the research. They removed lithology from the paper as the information of lithology cannot be extracted directly from the imagery. The factors were quantifies and identified usingdifferent approaches. The factors weights were then used to map susceptibility of landslides. The disadvantage of using the factors by images is that the resolution of image was not good enough to distinguish large scale landslide area.

3.2.9 Distribution-Free Approaches

Some new approaches such as fuzzy logic, artificial neural Networks (ANNs) etc. are adopted for landslide susceptibility mapping to eliminate subjectivity and increase accuracy and reliability. Neural Networks, fuzzy approaches are being used to identify landslide prone areas. Artificial Neural Networks deal with non-linearity in the processes to find solution to problems same like human brain does. They focus on objectivity in which data and weights are free of any spatial distributional assumptions.

Dieu Tien Bui et al (2016) [63] explained for training and validation purpose a framework was constructed using machine learning techniques of shallow landslide prone models. Past landslide data was used using two locations for landslide susceptible areas. Conditioning factors like slope, aspect, weather conditions etc for landslides using different sources were produced.

Huangqiong Chen et.al (2013) [64] proposed a novel technique based on RNN to predict landslide and its susceptibility. Optimization of initial weights and biases of network architecture were done by Genetic Algorithm. RNN model prediction accuracy showed better and accurate results as compared to neural network model that used feedforward techniques for the area of Baishuihehe landslide. For landslide displacement prediction RNN models are good enough Biswajeet Pradhan et.al (2010) [65] assessed landslide susceptible area using landslide prone areas identification analysis by using different methods like artificial neural network , GIS tools and RS techniques. Back propagation model of artificial "intelligence neural network" was selected for the using five random sites of training and hazard maps were constructed after getting the factors weight. To validate the results inventory landslides were used which showed 83.45% accuracy. However, lack of data was the limitation of the paper

3.3 Critical Analysis

The critical analysis includes the analysis and limitations of the techniques used in the studies discussed and limitations of the studies itself.

The landslide parameters impact the susceptibility of an area. The selection of landslide parameters is crucial to the analysis of landslide early detection.To identify correct parameters, co relation and other analysis should be done. Kuan-Tsung Chang et.al (2019) [65], Ataollah Shirzadi (2017) [61], Aafaf El Jazouli (2019) [42], Biswajeet Pradhan (2010) [65], Sajid Ali (2019) [44], KT Change et al. (2009) [60] have done no feature engineering before applying the technique to find susceptibile area. This leads to increase subjectivity and less accuracy. Some of the techniques used by researchers had followed more conventional ways to deal with landslide susceptibility problem [59] [66][46]

The systematic study of literature shows that for landslide prone areas mapping different techniques and approaches are there with every method has its own pros and cons. This study presents a critical review of landslide susceptibility methods having their pros and cons. For different methods serve different purpose in landslide susceptibility. For small region and short term, Slope stability analysis i.e. morphological methods [52] is relevant but for a longer term, this method is not suitable as it does not consider the landslide causing factors in its procedure. To know an overview of the area for landslide susceptibility analysis in a longer term, Qualitative methods [42] [53] [54] are appropriate for experienced designers and experts as it consider subjectivity in an initial stage using their their knowledge. Quantitative analysis [63] [64] prevent subjectivity and weigh down every parameter independently. The investigations on methodologies suggests that Quantitative methods engender better results as compared to any other. Recently the work on landslide susceptibility has been done extensively but there are some gaps that need to be addressed properly. Data scarcity is one of them. The limited data (landslide occurrences and its time and date) and resources to find the data, restrict the researchers to get the required results. Secondly, landslide susceptibility analysis is done to provide a mitigation plan and to prevent the risk of losses caused by landslides, so we need better algorithms to get better accuracy in identifying susceptible areas of landslides and to predict landslides beforehand.

Chapter 4

Proposed Framework

In this chapter we will identify the parameters used in landslide susceptibility using feature engineering, apply several approaches for landslide susceptibility and identify which model work best for landslide susceptibility analysis. In this chapter we will discuss the steps taken to acquire the data, labeling of the data, parameter selection and model implementation.

4.1 Framework Steps

The research procedure followed to achieve the objectives is discussed in this and coming sections. The proposed framework is divided into different phases shown in Figure 5.2.

- Study Area Selection
- Data Acquisition
- Feature Engineering
- Feature Selection
- Model Implementation

4.2 Data Selection

Muzaffarabad in northern Pakistan is the area of our research. It covers 1,241 km2 of the area. It is the captital of Kashmir administered by Pakistan. It is known for its high dynamic seismicity, huge topographic geomorphology, diverse mountain ranges, rivers, eroded land and massive precipitation factor. Weather consists of humid to moderate. The mean temperature in July is from 23 Degree Calcius to 35 Degree Celcius and in January it ranges from 3 Degree Celcius to 16 Degree Celcius respectively. Mean annual precipitation was recorded as 1,511mm.

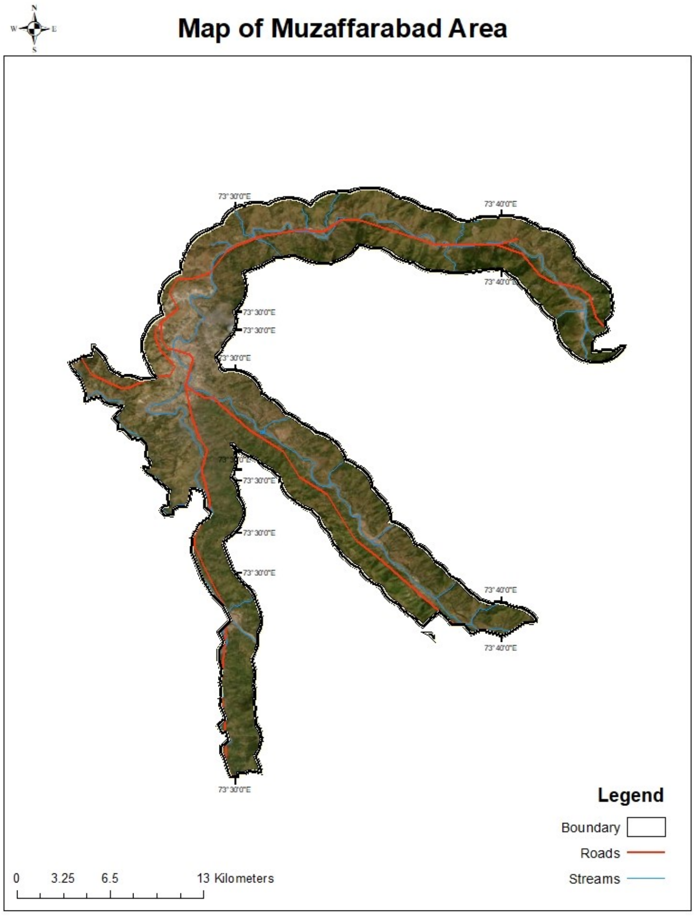

Figure 4.1: Study Area: Map of Muzaffarabad

4.3 Data Acquisition

4.3.1 Land cover

Land cover: It was acquired using Satellite imagery. Sentinel Imagery was classified into different classes of Land use using supervised technique. The classified classes Bare Land, Vegetation, Snow, Water Body, Urban Area.The classification accuracy was 90%

which was checked using Google Earth Pro.

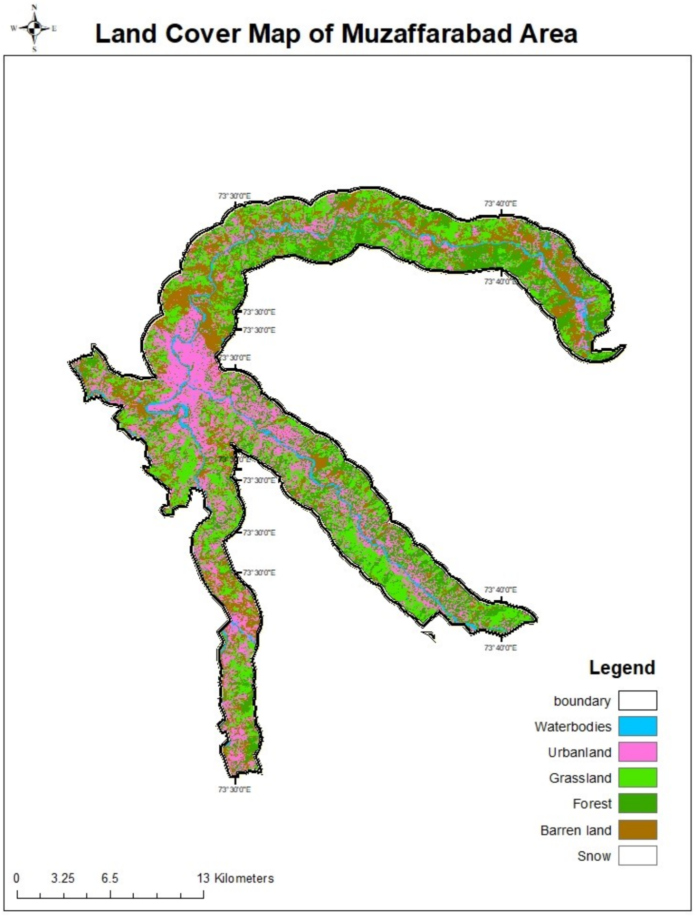

Figure 4.2: Landcover Map of Muzaffarabad

4.3.2 Slope

Slope: Digital Elevation model of Satellite ASTER, (30m) resolution was used to extract Slope from it. The Slope was then classified 0-10 degrees,10-20,20-30,30-40,40-50,50-60 and Above 60

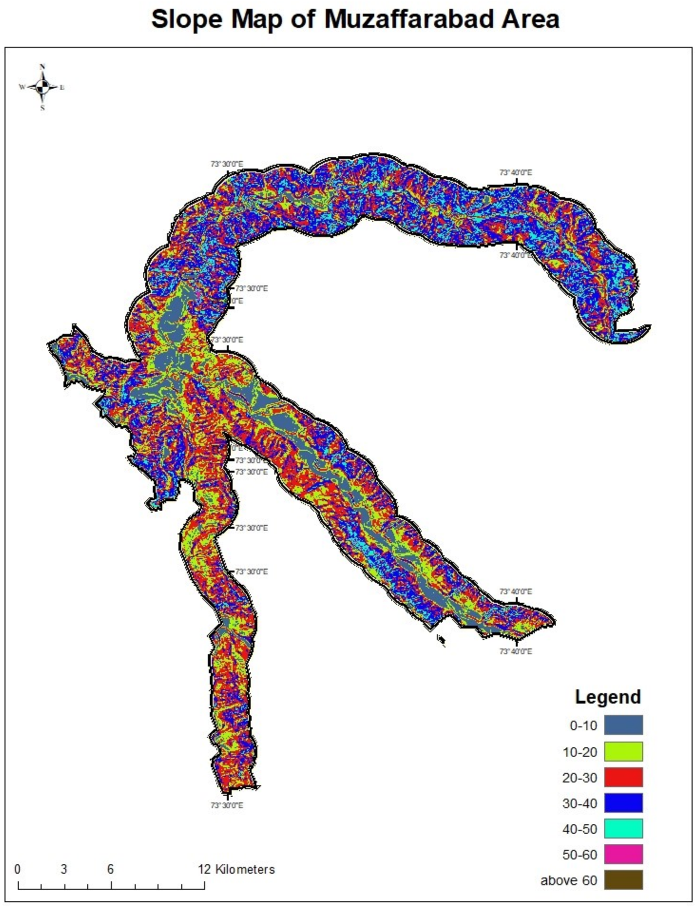

Figure 4.3: Slope Map of Muzaffarabad

4.3.3 Elevation

Elevation: To extract elevation, we used Digital Elevation model Aster (30m resolution). The elevation were further classified into four classes. The following are the classes of Elevation. 565-1000,1000-1500,1500-2000 and 2000-2500.

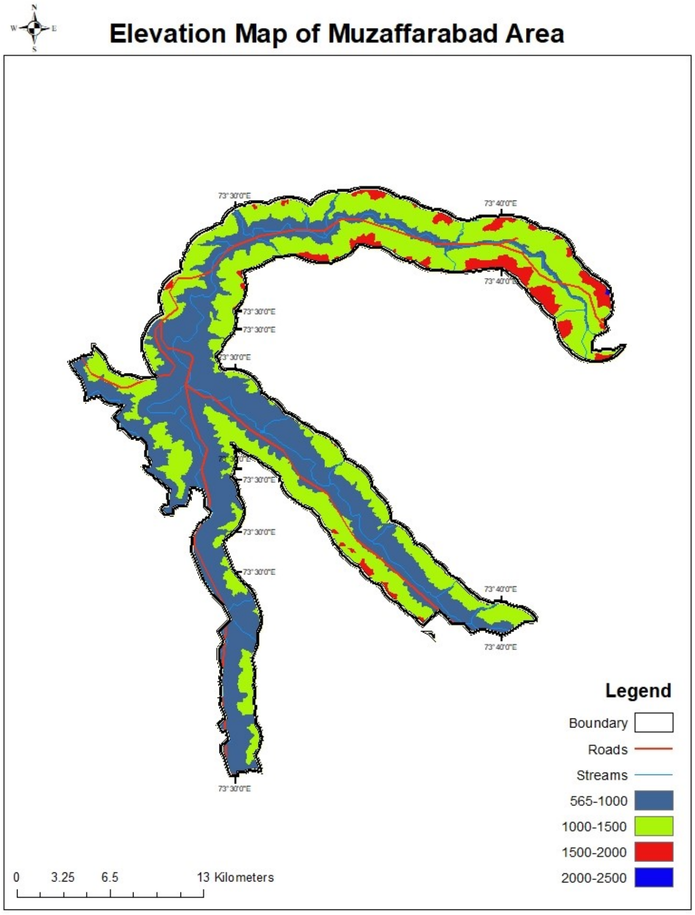

Figure 4.4: Elevation Map of Muzaffarabad

4.3.4 Aspect

Aspect: with the help of Digital Elevation Model of Aster Satellite of 30m resolution, Aspect was acquired. Aspect gives direction information of a place. It was classified into eight classes, which are North, North-East, East, South-East, South, South-West, West, and North-West

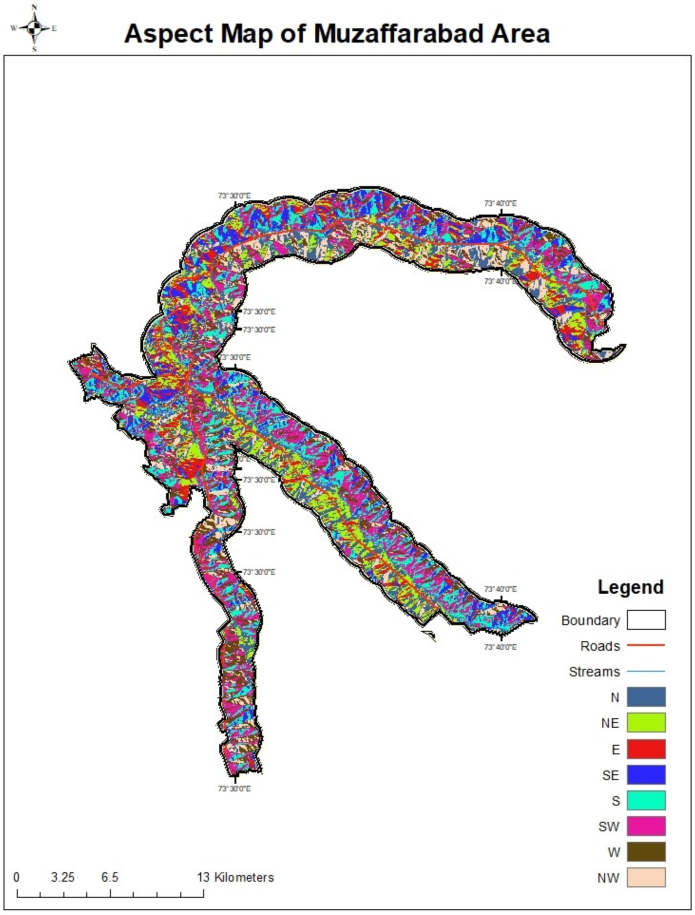

Figure 4.5: Aspect Map of Muzaffarabad

4.3.5 Geology

Geology: Geology was obtained from Tayyab et.al. It was further classified into different formations. The following are the names of those formations. Muree Formation, Tanol Formation, Panjal Metasediments, Panjal Volcanics, Hazara Formation, Holocene, Muzaffarabad Formation, Paleocene/Eocene Sequence

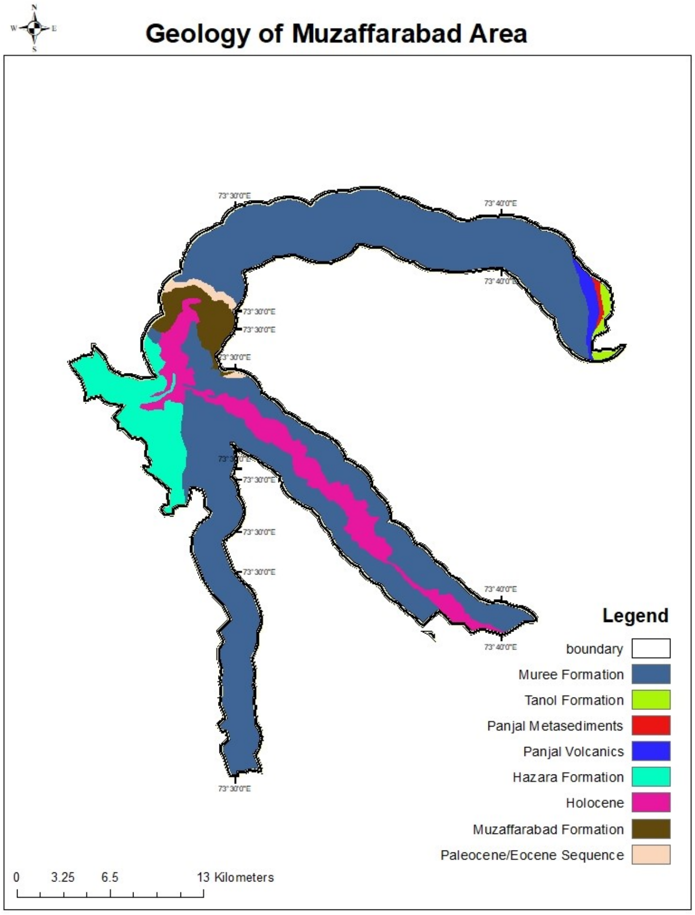

Figure 4.6: Geology Map of Muzaffarabad

4.3.6 Buffers of Faults, Streams and Roads

Buffers: In order to find streams, faults and roads near the study area, they were converted into buffer zones at a particular distance. For Faults, the buffers were 100, 200, 300, and Above 300. For Roads, the buffers were 50, 100 and above 100. For Streams 50, 100 and above 100.

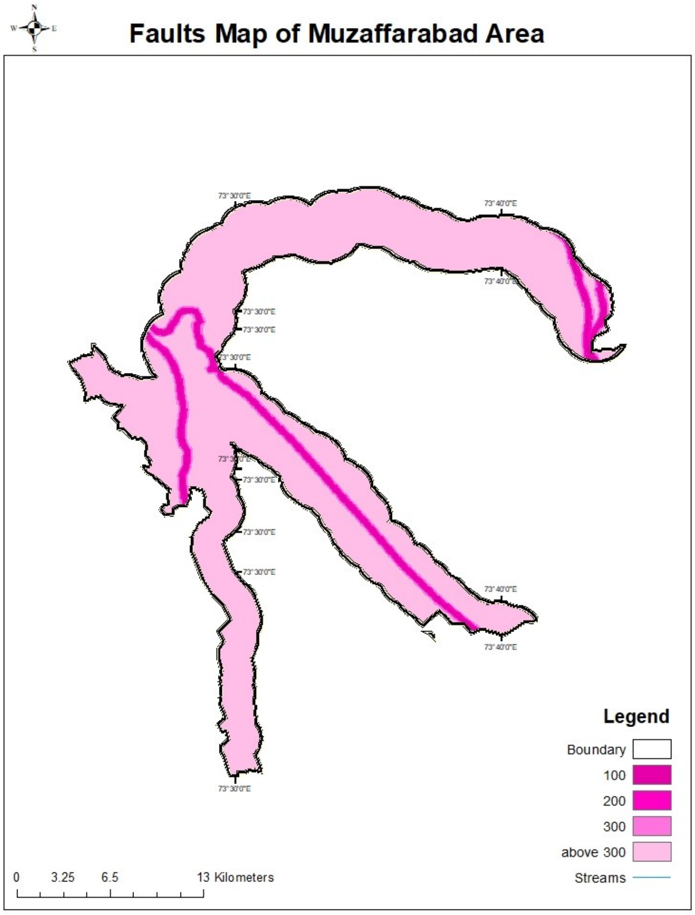

Figure 4.7: Faults Map of Muzaffarabad

4.4 Susceptibility Analysis

The analysis for susceptibility mapping was Analytical Hierarchy Process. Firstly, we used frequency ratio method to find relative ratio between each parameter using landslide inventory. Then used the frequency ratio method to be used in Analytical hierarchy method in order to reduce subjectivity. AHP uses multi-criterial matrix based on ranking of experts. This method is used in decision making process for example site selection, disaster management etc. In the recent years several researchers have been using this technique as it proves to be a convenient procedure that deals with multi-criteria hierarchical structures (SAATY 2003) . Pairwise comparisons matrix are involved in this process against decision variables. The factors of landslides are assigned a numeric value from 0-1 individually where 1 is the least and 9 is the most, depending on their relative rank. It shows the intensity of importance of landslide factors.The table introduced by Saaty is given in the background chapter

Pairwise comparison was done using metrics. For its preparation, the utilization of comparison matrix was done using metrics of landslide causative parameters, knowledge by the experts and existing research to determine the ranks among its parameters. Using the table developed by Saaty, for this research a pairwise comparison matrix was prepared to identify prone areas. Weights of criteria were computed by each column values' sum of metrics of pair-wise comparison, further it divided each metrics element by its column sum. The mean of elements was computed in row of each element. Several researchers have applied this technique for landslide susceptibility mapping to identify landslide prone areas by assigning weights to landslide causative parameters. The consistency of judgement in this process is improved by inconsistency measurement. The formula for consistency index (CI) for comparison of metrics is given below.

Where the principal Eigen value and n is the order of metrics. Similarly, when Consistency index is compared with set of numbers where each number is an average random CI, that depends on order of the metrics.

The consistency ratio above 0.1 indicates factor ratings are inconsistent which calls for revision of matrix judgements. The standard RI (Random Index) of pairwise comparison matrix is given by saaty as below:

Table 4.1: Pairwise comparison metrics

Features	**Geology**	**Slope**	**Land use**	**Roads**	**Stream**	**Faults**	**Height**	**Aspect**
Geology	1	1	5	5	9	8	8	8
Slope	1	1	2	4	8	9	7	8
Land use	1/5	1/2	1	2	1	1	1	1
Roads	1/5	1/4	1/2	1	2	3	2	2
Streams	1/9	1/8	1/4	1/2	1	5	2	2
Faults	1/8	1/9	1/5	1/3	1/5	1	1	2
Elevation	1/8	1/7	1/5	1/2	1/2	1	1	1
Aspect	1/8	1/8	1/2	1/2	1/2	1/2	1	1

Table 4.2: Calculation of parameters

Features	**Geology**	**Slope**	**Land use**	**Roads**	**Stream**	**Faults**	**Height**	**Aspect**
Geology	0.35	0.31	0.52	0.39	0.41	0.28	0.35	0.32
Slope	0.35	0.31	0.21	0.31	0.36	0.32	0.30	0.32
Land use	0.07	0.15	0.10	0.08	0.05	0.04	0.04	0.04
Roads	0.07	0.08	0.05	0.08	0.09	0.11	0.09	0.08
Streams	0.03	0.04	0.03	0.04	0.05	0.18	0.09	0.08
Faults	0.04	0.03	0.02	0.03	0.009	0.04	0.04	0.08
Elevation	0.04	0.04	0.02	0.04	0.02	0.04	0.04	0.04
Aspect	0.04	0.04	0.05	0.04	0.02	0.02	0.04	0.04

.

Table 4.3: Computation of criterion weights

Criteria	**Computation of criterion weights**	**Weights**
Geology	(0.35+0.31+0.52+0.39+0.41+0.28+0.35+0.32)/8	0.364
Slope	(0.35+0.31+0.21+0.31+0.36+0.32+0.30+0.32)/8	0.353
Land use	(0.07+0.15+0.10+0.08+0.05+0.04+0.04+0.04)/8	0.081
Roads	(0.07+0.08+0.05+0.08+0.09+0.10+0.09+0.08)/8	0.091
Streams	(0.04+0.04+0.03+0.04+0.05+0.18+0.09+0.08)/8	0.076
Faults	(0.04+0.03+0.02+0.03+0.009+0.03+0.04+0.08)/8	0.042
Elevation	(0.04+0.044+0.021+0.04+0.02+0.04+0.04+0.04)/8	0.041
Aspect	(0.04+0.04+0.05+0.04+0.02+0.02+0.04+0.04)/8	0.042

By using AHP, we will analyze the impact of landslide factors on the landslide occurrence and its susceptibility in that area. Our data set was in the form of grid data (raster format). As we wanted to see the impact of factors on landslide occurrence, hence we used AHP to calculate the influence of factors on landslide occurrence through some particular weights. For this purpose, we created a metrics based on expert opinion and frequency ratio, followed by calculations and summed up weights for each parameter. The weights of parameters helped us to identify landslide susceptible areas. Later the grid data is converted into numeric data for further Machine Learning and AI algorithms.

4.5 Data Labeling

The data was in the form of raster (grid data). For Machine learning Analysis, we needed them in the form of numeric data. For that purpose, all the grid data file were converted into shapefiles (vector data). The landslide inventories were intersected with the parameters to get all the landslide incident information along with its parameters at a particular point. The data was then converted to csv file for the experimentation purpose. The converted data was in the raw form. The data was manually labelled according to its rank. The ranks and weights of the parameters were given according to experts' opinions as well as experimentation.

4.6 Classification

AI and Machine Learning techniques like SVM, Logistic Regression , Decision tree, K Nearest neighbour and ANN evolved Cartesian Genetic Programming have been successfully applied for solving challenges in various domains. Due to reason that each technique works in a unique way, over time researchers have been able to identify the extent of their usability in different domains. Since there was no significantt time or computation cost involved with respect to either of the techniques, we can only compare based on accuracy scores .Classification attempts to learn the relationship between a set of feature variables and a target variable of interest. The target attribute in classification is a categorical variable with discrete values. The classifiers that we used in our analysis were Decision Tree, Support Vector Machine (SVM), KNN and Logistic Regression. These classifiers were applied in python anaconda using libraries like Scikit-learn, numpy etc. SVM was selected because it gives good accuracy with limited data. Decision trees were used because of their simplicity and less computational power. We used KNN in our project because it works better with static data. Later CGPANN was used to classify the data which proved to be very accurate.

Chapter 5

Results Evaluation and Validation

To find the factors influencing the landslide events Analytical Hierarchy process was applied as discussed in the previous chapter (see chapter 4). With the aim of finding the causal relationship between different variables, predictive models were used in order to observe those factors. In this chapter the results will be discussed along with the evaluation measure for the results. Accuracy was used to validate the models implemented. After model evaluation, comparison of different models were also discussed.

5.1 Experimental Protocols

Experimental protocols includes software specification and hardware, used for the implementation of models.

- **Software**: Following tools were used to implement the models; Anaconda Navigator, and Arc-GIS

    1. **Anaconda Navigator**: It is a platform which let the user to launch different applications to work with e.g. Spyder, Jupiter Notebook, RStudio etc. It also provides a platform to manage different environments and packages to be installed manually instead of writing commands using command-line [67].
    2. **Q-GIS**: Q-GIS is a data managemnet and storage toolbox which is open sourced used by a wide GIS community.It has GIS related problem solving tools[68].

- **Hardware** The system used to implement these models has the specifications as;

8GB RAM, Intel core i7, 2.20 GHz Processor, 64 bit OS, Windows 10 Pro

5.2 Dataset Description

After feature extraction and engineering, we got the following parameters based on its weights with reference to landslide susceptibility.The final parameters were as followes: Slope,Elevation, Geology, Roads, Landuse, Streams, and faults.The target classes were Low, Moderate, High and very High susceptibility zones. The dataset comprised of total 2000 inputs with 70:30 ratio of test and train data in holdout method. The final features after feature engineering shows lithology (0.36) has the highest weight then slope (0.353) followed by landuse(0.081), roads(0.091),streams(0.076), faults(0.042) aspect(0.042) and elevation (0.041)

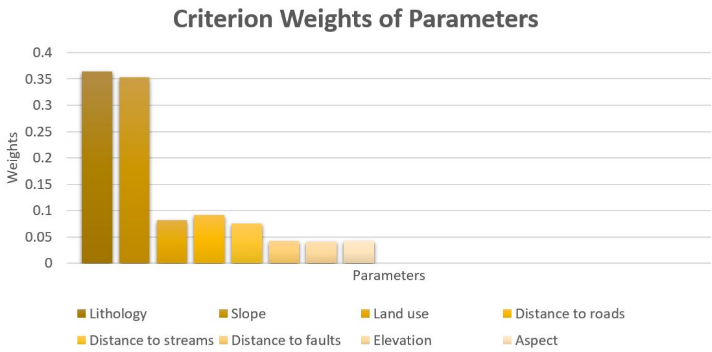

Figure 5.1: Weights of the Features after Feature Engineering

5.3 Results

5.3.1 Analytical Hierarchy Process

Several researchers have applied this technique for landslide susceptibility mapping to identify landslide prone areas by assigning weights to landslide causative parameters. The consistency of judgement in this process is improved by inconsistency measurement. The formula for consistency index (CI) for comparison of matrix is

$$\frac{\lambda_{max} - n}{n - 1}$$

Where lamda max is the principal Eigen value and n is the order of matrix. Similarly, when Consistency index is compared with set of numbers where each number is an average random CI, that depends on order of the matrix. While consistency ratio is equal to

$$\frac{CI}{RI}$$

The consistency ratio above 0.1 indicates factor ratings are inconsistent which calls for revision of matrix judgements. The standard RI (Random Index) of pairwise comparison matrix is given by saaty as below

Random Index									
n	RI	n	RI	n	RI	n	RI	n	RI
1	0.00	4	0.90	7	1.32	10	1.49	13	1.56
2	0.00	5	1.12	8	1.41	11	1.51	14	1.57
3	0.58	6	1.24	9	1.45	12	1.48	15	1.59

Figure 5.2: Random Index Standards

In our project, after all the calculations, we retrieved the Principal Eigen value as 8.518

Consistency Index:

$$\frac{8.518 - 8}{8 - 1}$$

$$0.073$$

Consistency Ratio:

$$\frac{0.073}{0.052}$$

$$0.052$$

Thus the result shows that the consistency ratio shows consistency with the judgements and preferences of the landslide parameters.

5.4 Evaluation Measures

To gauge the performance of each model, accuracy was used. Model macro level Recall, Precision, Accuracy and F1-score was reported in order to evaluate reliability of our results.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

where TP, TN, FP and FN are the number of True Positives, True Negatives, False Positives and False Negatives, respectively. Macro is chosen because it calculates mean of a particular score individually for every target value. Scores are shown for each method i.e. Holdout as well as 10 Fold cross validation. To optimize training results 10 Fold cross validation evaluation method was used. For 10 Fold cross validation method use Stratified K-Fold scheme which ensures that all folds maintain each class percentage representation. We used models that can give best results for each method. Performance of our models are given below.

5.5 Machine Learning Results

For machine learning classifiers performance, confusion matrix and ROC curve were used. The results are shown below.

	precision	recall	f1-score	support
3	1.00	0.71	0.83	7
4	0.80	0.91	0.85	22
5	0.88	0.84	0.86	25
accuracy			0.85	54
macro avg	0.89	0.82	0.85	54
weighted avg	0.86	0.85	0.85	54

Figure 5.3: SVM Result

	precision	recall	f1-score	support
3	0.50	0.71	0.59	7
4	0.84	0.73	0.78	22
5	0.92	0.92	0.92	25
accuracy			0.81	54
macro avg	0.75	0.79	0.76	54
weighted avg	0.83	0.81	0.82	54

Figure 5.4: Decision Tree Result

	precision	recall	f1-score	support
3	0.83	0.71	0.77	7
4	0.75	0.82	0.78	22
5	0.83	0.80	0.82	25
accuracy			0.80	54
macro avg	0.81	0.78	0.79	54
weighted avg	0.80	0.80	0.80	54

Figure 5.5: K Nearest Neighbours

Table 5.1: Performance score

Results	**Classifier**	**Accuracy**	**Precision**	**Recall**	**f1-score**
10 Folds C.V	SVM	0.85	0.86	0.85	0.85
	LR	0.87	0.88	0.87	0.87
	DT	0.81	0.83	0.81	0.82
	KNN	0.80	0.80	0.80	0.80
Holdout Validation	SVM	0.89	0.88	0.89	0.89
	LR	0.86	0.84	0.88	0.86
	DT	0.84	0.84	0.83	0.85
	KNN	0.88	0.86	0.87	0.88
	CGPANN	0.96	0.93	0.95	0.95

5.5.1 Performance Score

5.6 Resulting Output Map

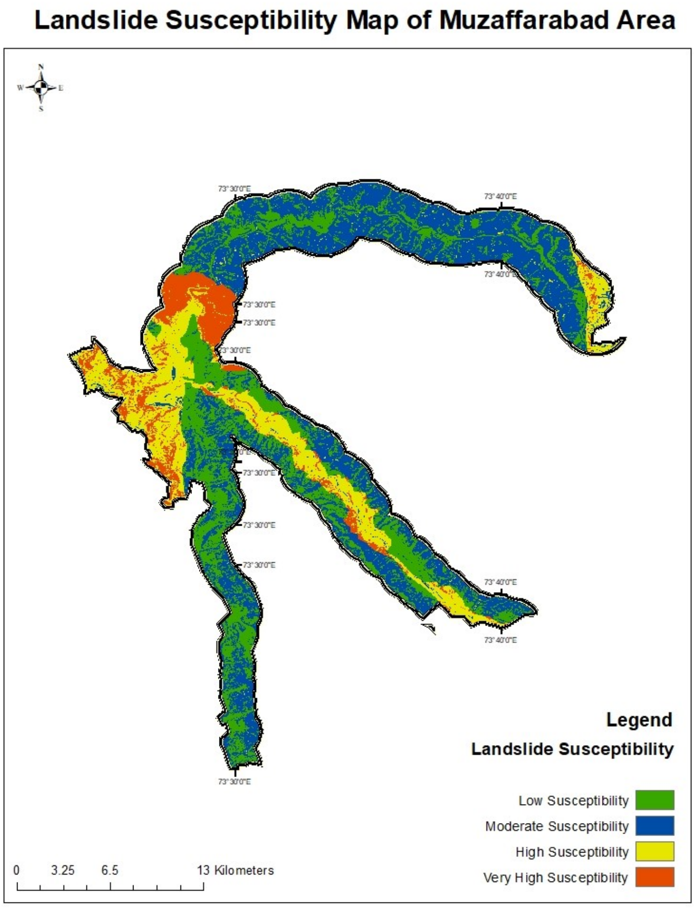

Figure 5.6: Landslide Susceptibility Map for Muzaffarabad

Figure 5.7: Landslide Susceptibility Area

5.7 Discussion

The classification was done using 10 Folds method as well as Holdout method.By looking at the results, it shows that the results performance of Logistic Regression (0.87) in 10 Folds cross validation method is better as compared to Support vector Machine (0.85), Decision Tree (0.81) and K Nearest Neighbour(0.80) while in Holdout method CGPANN (0.96) has out performed any other Machine learning classifier.Although Decision Tree (0.84), K Nearest Neighbour (0.88) and SVM (0.89) are giving results better than 10 Fold cross validation method. Moreover, the resulting Map shows categories of Landslide susceptible areas using different color schemes. Green color shows low susceptibility in the area, blue color shows moderate susceptibility, yellow color shows high susceptibility while red color shows very high susceptibility area. The graph shows percentage of Area that comes under Landslide Susceptibility. It shows that 19% of area comes under very high susceptibility, 22% comes under Low susceptibility, 50% of Area comes under moderate susceptibility and 9% of area comes under very high susceptibility.

Chapter 6

Conclusion and Future Work

In this chapter, we will discuss about the contribution of the research work, discussions over the results and summery of the whole findings in this research. In this research susceptibility analysis were done to identify landslide prone areas. It will help in hazard management and mitigation plans.

6.1 Discussion

In this project we focused on detection of landslides using GIS tools and machine learning techniques. Using AHP we identified landslide prone parameters and found areas susceptibility by dividing them into four categories i-e low landslide susceptibility, Moderate landslide susceptibility, High landslide susceptibility and very high landslide susceptibility. To identify parameters weights for landslide susceptibility AHP and frequency methods were used which helped us in feature engineering. After that AI and Machine learning classifiers were used to detect susceptible areas of landslides. Among all of them CGPANN outperformed and have given best accuracy of 0.96. The analysis show that almost 30% of the area comes under high and very high susceptibility. Barren land and grassland in land cover, fault lines and Muzaffarabad formation, Hazara formation and Holocene in geology are found to be most susceptible factors in Muzaffarabad areas which contributes to landslides.

6.2 Contribution

Landslides can result in enormous casualties and huge economic losses in mountainous regions. Because of the damage millions of dollars and hundreds of lives are lost each year.

In this research we have worked on novel method of feature engineering to rank the landslide parameters in order to predict landslide in the future using susceptibility analysis. We achieved very good accuracy that will help in reliable landslide prediction results using inexpensive and open sources tools and softwares. Susceptibility Analysis of landslides will help following authorities in different ways:

1. **1. Disaster Management Authorities** The susceptibility results will help the authority to have a better plan for hazard management and mitigation. Susceptibility analysis will help them identifying location for early warning systems and sensors placement in the area that are hazard prone.

2. **Urban Planning Authorities** Due to urbanization, a lot of constructions have been taking place recently especially in the areas with slope (mountains) like bridges, roads etc. It is best to have a landslide susceptibility map as it will help in construction planning for that area.

3. **Tourism** : Many tourists get stuck in the land sliding during heavy monsoon rainfall or during the snow. The susceptibility analysis will help them in knowing the placing that are prone to landslides.

6.3 Conclusion

Using AHP we identified landslide prone parameters and found areas susceptibility by dividing them into four categories i-e low landslide susceptibility, Moderate landslide susceptibility, High landslide susceptibility and very high landslide susceptibility. To identify parameters weights for landslide susceptibility AHP and frequency methods were used which helped us in feature engineering. After that AI and Machine learning classifiers were used to detect susceptible areas of landslides. Among all of them CGPANN outperformed and have given best accuracy of 0.96 using all open sources tools and data. In order to mitigate landslide hazard effectively, better methodologies are required to develop a better understanding of landslide hazard and to make rational decisions for management of landslide risk. For landslide mitigation and hazard management, landslide susceptibility analysis is a useful way to identify landslide prone areas. However, landslide is being controlled by geomorphic factors which means it's a complex natural phenomena, thus it is needed to be handled with better accuracy. The landslide susceptibility technique should have a better accuracy as its accuracy will help in identifying prone areas to landslides and later the most susceptible area will have proper hazard management and sensors placement for its prediction in order to keep the people, their properties and public infrastructure protected. Machine learning techniques are adopted and applied with great success for landslide susceptibility. With this information, we can place sensors on these susceptible areas in our future work and also inform the authorities about the prone areas for better landslide hazard management.

6.4 Limitations and Future Work

Dataset can be improved. More data can be added to the work. In future we are going to extend this work by installing sensors and cameras, developing heterogeneous sensor network and using better Artificial Intelligence (AI) models that are effective in developing new landslide reduction services to predict landslides beforehand to save the community and infrastructure from big losses.

References

[1] Stephen G Evans, Nicholas J Roberts, Anatoli Ischuk, Keith B Delaney, Galina S Morozova, and Olga Tutubalina. Landslides triggered by the 1949 khait earthquake, tajikistan, and associated loss of life. *Engineering Geology*, 109(3-4):195–212, 2009.

[2] MA Abdel-Latif. *Landslide hazard assessment.* PhD thesis, The Ohio State University, 1994.

[3] Lionel E Jackson. Landslides and landscape evolution in the rocky mountains and adjacent foothills area, southwestern alberta, canada. *Catastrophic landslides: effects, occurrence, and mechanisms, Geol. Soc. Am. Rev. Eng. Geol*, 15:325–344, 2002.

[4] Carolyn Kousky. Informing climate adaptation: A review of the economic costs of natural disasters. *Energy Economics*, 46:576–592, 2014.

[5] Disaster Prevention Team. Technical report on the trends of global water-related disasters. 2008.

[6] Femke Vos, José Rodríguez, Regina Below, and Debby Guha-Sapir. Annual disaster statistical review 2009: The numbers and trends. Technical report, Centre for Research on the Epidemiology of Disasters (CRED), 2010.

[7] Christian Crowley and Frederick L Joutz. Weather effects on electricity loads: Modeling and forecasting 12 december 2005. *Final report for US EPA on weather effects on electricity loads*, 2005.

[8] Dario Tarchi, Nicola Casagli, Riccardo Fanti, David D Leva, Guido Luzi, Alessandro Pasuto, Massimiliano Pieraccini, and Sandro Silvano. Landslide monitoring by using ground-based sar interferometry: an example of application to the tessina landslide in italy. *Engineering geology*, 68(1-2):15–30, 2003.

[9] Leonardo Zan, Gilberto Latini, Evasio Piscina, Giovanni Polloni, and Pieramelio Baldelli. Landslides early warning monitoring system. In *IEEE International Geoscience and Remote Sensing Symposium*, volume 1, pages 188–190. IEEE, 2002.

[10] Julien Travelletti, Christophe Delacourt, Pascal Allemand, J-P Malet, Jean Schmittbuhl, Renaud Toussaint, and Mickael Bastard. Correlation of multi-temporal ground-based optical images for landslide monitoring: Application, potential and limitations. *ISPRS Journal of Photogrammetry and Remote Sensing*, 70:39–55, 2012.

[11] Abolfazl Jaafari, Mahdi Panahi, Binh Thai Pham, Himan Shahabi, Dieu Tien Bui, Fatemeh Rezaie, and Saro Lee. Meta optimization of an adaptive neuro-fuzzy inference system with grey wolf optimizer and biogeography-based optimization algorithms for spatial prediction of landslide susceptibility. *Catena*, 175:430–445, 2019.

[12] Cheng Lian, Zhigang Zeng, Wei Yao, and Huiming Tang. Multiple neural networks switched prediction for landslide displacement. *Engineering geology*, 186:91–99, 2015.

[13] Maher Ibrahim Sameen, Biswajeet Pradhan, and Saro Lee. Application of convolutional neural networks featuring bayesian optimization for landslide susceptibility assessment. *Catena*, 186:104249, 2020.

[14] Ulrich Kamp, Benjamin J Growley, Ghazanfar A Khattak, and Lewis A Owen. Gis-based landslide susceptibility mapping for the 2005 kashmir earthquake region. *Geomorphology*, 101(4):631–642, 2008.

[15] Lasafam Iturrizaga. Paraglacial landform assemblages in the hindukush and karakoram mountains. *Geomorphology*, 95(1-2):27–47, 2008.

[16] Anne Replumaz and P Tapponnier. Reconstruction of the deformed collision zone between india and asia by backward motion of lithospheric blocks. *Journal of Geophysical Research: Solid Earth*, 108(B6), 2003.

[17] Wahid Ullah and Nihei Takaaki. Climate change vulnerability of pakistan towards natural disasters: A review. *International Journal of Environmental Protection and Policy*, 4(5):126, 2016.

[18] Clément Michoud, M Jaboyedoff, MH Derron, Farrokh Nadim, and Eric Leroi. Classification of landslide-inducing anthropogenic activities. In *Proceedings 5th Canadian Conference on Geotechnique and Natural Hazards, 15–17 May 2011, Kelowna*, 2011.

[19] Atta-ur Rahman, Amir Nawaz Khan, and Andrew E Collins. Analysis of landslide causes and associated damages in the kashmir himalayas of pakistan. *Natural hazards*, 71(1):803–821, 2014.

[20] Saro Lee and Jasmi Abdul Talib. Probabilistic landslide susceptibility and factor effect analysis. *Environmental Geology*, 47(7):982–990, 2005.

[21] SG Wright and Ellen M Rathje. Triggering mechanisms of slope instability and their relationship to earthquakes and tsunamis. *Pure and Applied Geophysics*, 160 (10-11):1865–1877, 2003.

[22] Harianto Rahardjo, TH Ong, RB Rezaur, and Eng Choon Leong. Factors controlling instability of homogeneous soil slopes under rainfall. *Journal of Geotechnical and Geoenvironmental Engineering*, 133(12):1532–1543, 2007.

[23] Bijan Khazai and Nicholas Sitar. Evaluation of factors controlling earthquake-induced landslides caused by chi-chi earthquake and comparison with the northridge and loma prieta events. *Engineering geology*, 71(1-2):79–95, 2004.

[24] Thomas L Saaty. Decision making—the analytic hierarchy and network processes (ahp/anp). *Journal of systems science and systems engineering*, 13(1):1–35, 2004.

[25] Jiří Franek and Aleš Kresta. Judgment scales and consistency measure in ahp. *Procedia Economics and Finance*, 12:164–173, 2014.

[26] Hamid Reza Pourghasemi, Biswajeet Pradhan, and Candan Gokceoglu. Application of fuzzy logic and analytical hierarchy process (ahp) to landslide susceptibility mapping at haraz watershed, iran. *Natural hazards*, 63(2):965–996, 2012.

[27] Prabin Kayastha, Megh Raj Dhital, and Florimond De Smedt. Application of the analytical hierarchy process (ahp) for landslide susceptibility mapping: A case study from the tinau watershed, west nepal. *Computers & Geosciences*, 52:398–408, 2013.

[28] M Ercanoglu, O Kasmer, and N Temiz. Adaptation and comparison of expert opinion to analytical hierarchy process for landslide susceptibility mapping. *Bulletin of Engineering Geology and the Environment*, 67(4):565–578, 2008.

[29] Johan AK Suykens and Joos Vandewalle. Least squares support vector machine classifiers. *Neural processing letters*, 9(3):293–300, 1999.

[30] Simon Tong and Edward Chang. Support vector machine active learning for image retrieval. In *Proceedings of the ninth ACM international conference on Multimedia*, pages 107–118, 2001.

[31] S Rasoul Safavian and David Landgrebe. A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics*, 21(3):660–674, 1991.

[32] David W Hosmer Jr, Stanley Lemeshow, and Rodney X Sturdivant. *Applied logistic regression*, volume 398. John Wiley & Sons, 2013.

[33] Sahibsingh A Dudani. The distance-weighted k-nearest-neighbor rule. *IEEE Transactions on Systems, Man, and Cybernetics*, (4):325–327, 1976.

[34] Andrew James Turner and Julian Francis Miller. The importance of topology evolution in neuroevolution: a case study using cartesian genetic programming of artificial neural networks. In *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, pages 213–226. Springer, 2013.

[35] Fausto Guzzetti, Alberto Carrara, Mauro Cardinali, and Paola Reichenbach. Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, central italy. *Geomorphology*, 31(1-4):181–216, 1999.

[36] R Dikau, D Brunsden, L Schrott, and ML Ibsen. Landslide recognition: identitication, movement and causes, lnternational association of geomorphologists, publication no. 5, john wiley and sons, chichester, 1996, 251 p. 1996.

[37] William E Dietrich, Robert Reiss, Mei-Ling Hsu, and David R Montgomery. A process-based model for colluvial soil depth and shallow landsliding using digital elevation data. *Hydrological processes*, 9(3-4):383–400, 1995.

[38] Pietro Aleotti and Robin Chowdhury. Landslide hazard assessment: summary review and new perspectives. *Bulletin of Engineering Geology and the environment*, 58(1):21–44, 1999.

[39] Kang-Tsung Chang and Shou-Hao Chiang. An integrated model for predicting rainfall-induced landslides. *Geomorphology*, 105(3-4):366–373, 2009.

[40] Jaewon Choi, Hyun-Joo Oh, Hong-Jin Lee, Changwook Lee, and Saro Lee. Combining landslide susceptibility maps obtained from frequency ratio, logistic regression, and artificial neural network models using aster images and gis. *Engineering Geology*, 124:12–23, 2012.

[41] Biswajeet Pradhan. Landslide susceptibility mapping of a catchment area using frequency ratio, fuzzy logic and multivariate logistic regression approaches. *Journal of the Indian Society of Remote Sensing*, 38(2):301–320, 2010.

[42] Aafaf El Jazouli, Ahmed Barakat, and Rida Khellouk. Gis-multicriteria evaluation using ahp for landslide susceptibility mapping in oum er rbia high basin (morocco). *Geoenvironmental Disasters*, 6(1):3, 2019.

[43] Kuan-Tsung Chang, Abdelaziz Merghadi, Ali P Yunus, Binh Thai Pham, and Jie Dou. Evaluating scale effects of topographic variables in landslide susceptibility models using gis-based machine learning techniques. *Scientific reports*, 9(1):1–21, 2019.

[44] Sajid Ali, Peter Biermanns, Rashid Haider, and Klaus Reicherter. Landslide susceptibility mapping by using a geographic information system (gis) along the china–pakistan economic corridor (karakoram highway), pakistan. *Natural Hazards & Earth System Sciences*, 19(5), 2019.

[45] Himan Shahabi, Mazlan Hashim, and Baharin Bin Ahmad. Remote sensing and gis-based landslide susceptibility mapping using frequency ratio, logistic regression, and fuzzy logic methods at the central zab basin, iran. *Environmental Earth Sciences*, 73(12):8647–8668, 2015.

[46] Earl E Brabb. The world landslide problem. *Episodes Journal of International Geoscience*, 14(1):52–61, 1991.

[47] Phillip N Flentje, Anthony Miner, Graham Whitt, and Robin Fell. Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. 2007.

[48] Gerald F Wieczorek. Preparing a detailed landslide-inventory map for hazard evaluation and reduction. *Bulletin of the Association of Engineering Geologists*, 21(3): 337–342, 1984.

[49] Robert H Wright, Russell H Campbell, and Tor H Nilsen. Preparation and use of isopleth maps of landslide deposits. *Geology*, 2(10):483–485, 1974.

[50] M Humbert. La cartographie en france des zones exposées à des risques lies aux mouvements du sol—cartes zermos. *Bulletin of the International Association of Engineering Geology-Bulletin de l'Association Internationale de Géologie de l'Ingénieur*, 16(1):80–82, 1977.

[51] JP Méneroud and A Calvino. carte zermos. *Zones exposés ades risques liés aix-mouvements du sol et du sous-sol a1*, 25, 1976.

[52] Eric Leroi. Landslide hazard-risk maps at different scales: objectives, tools and developments. In *Landslides*, pages 35–51, 1996.

[53] Himan Shahabi, Mazlan Hashim, and Baharin Bin Ahmad. Remote sensing and gis-based landslide susceptibility mapping using frequency ratio, logistic regression, and fuzzy logic methods at the central zab basin, iran. *Environmental Earth Sciences*, 73(12):8647–8668, 2015.

[54] Shamsa Kanwal, Salman Atif, and Muhammad Shafiq. Gis based landslide susceptibility mapping of northern areas of pakistan, a case study of shigar and shyok basins. *Geomatics, Natural Hazards and Risk*, 8(2):348–366, 2017.

[55] Pietro Aleotti and Robin Chowdhury. Landslide hazard assessment: summary review and new perspectives. *Bulletin of Engineering Geology and the environment*, 58(1):21–44, 1999.

[56] Shraban Sarkar, Archana K Roy, and Tapas R Martha. Landslide susceptibility assessment using information value method in parts of the darjeeling himalayas. *Journal of the Geological Society of India*, 82(4):351–362, 2013.

[57] Qianqian Ba, Yumin Chen, Susu Deng, Qianjiao Wu, Jiaxin Yang, and Jingyi Zhang. An improved information value model based on gray clustering for landslide susceptibility mapping. *ISPRS International Journal of Geo-Information*, 6(1):18, 2017.

[58] Dongli Fan, Xi-min Cui, De-bao Yuan, Jiafeng Wang, Jinlin Yang, and Shengyao Wang. Weight of evidence method and its applications and development. *Procedia Environmental Sciences*, 11:1412–1418, 2011.

[59] Mukhiddin Juliev, Martin Mergili, Ismail Mondal, Bakhtiar Nurtaev, Alim Pulatov, and Johannes Hübl. Comparative analysis of statistical methods for landslide susceptibility mapping in the bostanlik district, uzbekistan. *Science of the total environment*, 653:801–814, 2019.

[60] Ataollah Shirzadi, Dieu Tien Bui, Binh Thai Pham, Karim Solaimani, Kamran Chapi, Ataollah Kavian, Himan Shahabi, and Inge Revhaug. Shallow landslide susceptibility assessment using a novel hybrid intelligence approach. *Environmental Earth Sciences*, 76(2):60, 2017.

[61] Kuan-Tsung Chang, Abdelaziz Merghadi, Ali P Yunus, Binh Thai Pham, and Jie Dou. Evaluating scale effects of topographic variables in landslide susceptibility models using gis-based machine learning techniques. *Scientific reports*, 9(1):1–21, 2019.

[62] Jaewon Choi, Hyun-Joo Oh, Hong-Jin Lee, Changwook Lee, and Saro Lee. Combining landslide susceptibility maps obtained from frequency ratio, logistic regression, and artificial neural network models using aster images and gis. *Engineering Geology*, 124:12–23, 2012.

[63] Dieu Tien Bui, Tran Anh Tuan, Harald Klempe, Biswajeet Pradhan, and Inge Revhaug. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides*, 13(2):361–378, 2016.

[64] Huangqiong Chen and Zhigang Zeng. Deformation prediction of landslide based on improved back-propagation neural network. *Cognitive computation*, 5(1):56–62, 2013.

[65] Biswajeet Pradhan, Hyun-Joo Oh, and Manfred Buchroithner. Weights-of-evidence model applied to landslide susceptibility mapping in a tropical hilly area. *Geomatics, Natural Hazards and Risk*, 1(3):199–223, 2010.

[66] Jean Baptiste Nsengiyumva, Geping Luo, Amobichukwu Chukwudi Amanambu, Richard Mind'je, Gabriel Habiyaremye, Fidele Karamage, Friday Uchenna Ochege, and Christophe Mupenzi. Comparing probabilistic and statistical methods in landslide susceptibility modeling in rwanda/centre-eastern africa. *Science of the total environment*, 659:1457–1472, 2019.

[67] Daniel Guillermo Cavaller Riva, Antonio Sottile Bordallo, Emiliano Andrés Dueñas, Cristian Darío Ortega Yubro, Héctor Nicolás Sosa, Diego Vidal Silva, Roberto Fermín Moreno, Pablo Ariel Masier, and Matías Rodolfo Codorniú. Análisis de anomalías en los días de la semana, en precios del bitcoin, con r studio® y rapidminer studio®. 2011.

[68] Kevin Johnston, Jay M Ver Hoef, Konstantin Krivoruchko, and Neil Lucas. *Using ArcGIS geostatistical analyst*, volume 380. Esri Redlands, 2001.