Pan African University
Institute of Water
and Energy Sciences

PAN-AFRICAN UNIVERSITY

INSTITUTE FOR WATER AND ENERGY SCIENCES

(Including CLIMATE CHANGE)

Master Dissertation

Submitted in partial fulfillment of the requirements for the Master degree in

[WATER ENGINEERING]

Presented by

Roset NAMWANJE

A COMPARATIVE ANALYSIS ON GIS BASED METHODS AND MACHINE LEARNING ALGORITHM IN LANDSLIDE SUSCEPTIBILITY MODELING, A case study of Bududa, Uganda

Defended on 15th/11/2021 Before the Following Committee:

Chair	Khaldoon A. Mourad	Dr.	Lund University, Sweden
Supervisor	John M. Gathenya	Prof	JKUAT, Kenya
External Examiner	Saber Mohamed	Dr.	Kyoto University, Japan
Internal Examiner	Tayeb Boulemaiz	Dr.	University of Ghardaia, Algeria

Abstract

Landslide susceptibility modeling is of critical importance to landslide risk management, urban planning, understanding landscape evolution, and identifying landslide spatial and temporal signatures. In this study, the GIS-based weight of evidence model and the Support Vector Machine learning algorithms were compared in landslide susceptibility assessments using a case study of Bududa located in the eastern part of Uganda. The inventory of landslides applied in the study was created using satellite imagery and historical maps of the region. The causative factors were derived from the STRM DEM of 30m resolution and the Geological characteristics of the region were obtained from the Ministry of Geological Survey and Mines. The Weight of Evidence model revealed 5 factors with a positive spatial association to landslide occurrences and these were the Slope Angle, Profile Curvature, Plan Curvature, Geology and Distance to Rivers. The SVM model identified Slope Angle, Profile Curvature, Stream Power Index and Elevation as triggering factors. The comparative analysis between the two methods was conducted using the Confusion Matrix, the Receiver Operating Curves (ROC) and the developed susceptibility maps. The ROC curves gave an accuracy of 87% of the Area under Curve (AUC) for the Optimizable Support Vector Machines and a 79% (AUC) for Weight of Evidence in performance. The field validation data placed ten landslide points in the mapped susceptibility zones for both models which indicated a good overall performance for both methodologies, however the study evinced potential of more efficient and accurate results with the integration of both approaches in landslide susceptibility modeling.

Key words: Comparative Analysis, Weight of Evidence Model, Support Vector Machines, GIS, Machine learning Algorithms, landslides, Bududa, Uganda

Declaration

I, Roset NAMWANJE declare that this is the original copy of my work and no other duplicate of the same content, whole or in part is going to be found in any other degree or professional qualification apart from where works of others have been cited and properly referenced.

Signature;

Date; 17th/11/2021

Approval

I undersigned, Prof Mwangi Gathenya, a guest lecturer at the Pan African University Institute of Water and Energy Sciences including Climate Change (PAUWES), and permanent lecturer at Jomo Kenyatta University of Technology, Kenya; certify that Ms. Roset NAMWANJE conducted her Master Thesis research under my supervision. Certified further, that this master thesis entitled "A Comparative Analysis on GIS Based Methods and Machine learning algorithms in landslide susceptibility modeling, a case study of Bududa Uganda" is an authentic work of Ms. Roset NAMWANJE who carried out the research under my guidance.

Signature  Date 19th/11/2021

Acknowledgement

I extend my gratitude to the African Union through the Pan African University for granting me the opportunity to partake in this research and also financing it to its successful completion.

Sincere thanks go out to the Pan African University Institute of Water and Energy sciences (including Climate Change) fraternity for the tremendous support and guidance during the entire academic program

I would love to express my deep and sincere gratitude to my supervisor, Prof John M. Gathenya for the consistent guidance and direction during my entire duration of the research.

Finally, I am grateful for the love and support of my family and colleagues through the trying times of the pandemic that challenged the progress of this study

Abbreviation

AUC	Area under Curve
DEM	Digital Elevation Model
DDPM	Department of Disaster Preparedness and Management
OPM	Office of the Prime Minister
RBF	Radial Basis Function
ROC	Receiver Operating Curve
SA	Slope Angle
SPI	Stream Power Index
SVM	Support Vector Machine
TWI	Topographic Wetness Index
TRI	Topographic Ruggedness Index
UNMA	Uganda National Meteorological Agency
UBOS	Uganda Bureau of Statistics
USGS	United States of Geological Survey
WoFE	Weight of Evidence

Table of Contents

List of Tables

List of figures

1.0 INTRODUCTION

1.1 Background

Landslides, also known as landslips, are among the most destructive geological disasters occurring in the world today. They are part of the natural geomorphological processes that shape the earth's surface however, when they endanger humanity and the environment, they create a deleterious impact that requires alleviation. Globally, landslides are associated with loss of human lives, injury to people, infrastructure damage, degradation of farmlands, and water resources deterioration (Dang et al., 2020; Vojteková & Vojtek, 2020; Xiao et al., 2020). Their occurrences and impacts are still prevalent and present on all continents, including Africa. This has stirred further research into risk management and preparedness in most prone countries.

According to (Broeckx et al., 2018), Africa continues to be under-represented in landslide science, with little or no studies being done in the most vulnerable countries. Over the last 2 decades, Uganda has experienced more than 700 landslides, with over 400 hitting Bududa, a sub-county located on the foothills of Mt Elgon in the eastern part of the country. More than 1,000 human lives have been lost in the last 20 years as a result of the disaster and a total of 9666 people and more have been displaced (Uganda Red Cross Society, 2019). Studies adopting GIS and other related geospatial technologies in landslide susceptibility assessments have been conducted in the populated areas of the Rwenzori Mountains, Western Uganda and Pixel-based regression models, radar-based DEMs and ASTER DEMs have also been applied (Jacobs et al., 2018) to identify susceptible areas. In Bududa, the LAPSUS-LS landslide model that identifies landslide triggering sites and scenarios of landslide erosion, sediment loading and deposition pathways has been studied (K. M. G. Kitutu, 2010) however the occurrences are still prevalent and often hitting the same spot continuously.

In summary, studies that apply GIS in risk mapping are common. However, studies that apply machine learning algorithms like K-Nearest Neighbor, Support Vector Machines, and Random Forest Classifiers are still limited. To thoroughly manage disaster-related risks, an effective decision support system integrated with the state-of-art technologies that apply GIS, machine learning algorithms and their hybrid versions in susceptibility modeling is of vital importance. This study will explore the applicability of machine learning algorithms in landslide

susceptibility modeling and their comparative advantage to the widely adopted GIS methods in Bududa.

1.2 Problem Statement and Justification

The prevalence of both shallow and deep landslides in Bududa and other mountainous/hilly parts of Uganda has adversely affected the water quality in the region, this is due to the high sediment loads affiliated with the disaster. Landslides have constrained economic growth through agricultural disruption and have also claimed hundreds of human lives (Namono et al., 2019). On December 4, 2019, more than 38 people in Bududa lost their lives to landslides, and many were reported missing (Red Cross Society, 2019). The continued recurrence of landslides in the region has become a threat to community livelihood, with Bududa being hit severely in relation to other prone regions (>400 times since 2010). Occasionally, the occurrence of landslides is accompanied by severe flooding around the Manafwa River in Bududa. These floods submerge households, destroy property and cut off transport routes. The relocation policy adopted by the Office of the Prime Minister for Disaster Management and Preparedness has been ineffective due to the inadequate compensation plan accompanied with the policy (Vlaeminck et al., 2016). It neglects the social and economic implications of displacement of citizens from landslide hotspots. This makes the prominent mitigation measure in the country limited thus the need for diverse risk preventative strategies. Landslide susceptibility modeling investigates the major factors responsible for triggering landslides, identifies highly susceptible areas and the spatial or temporal probability of such occurrences in the areas. Furthermore, this study will be a step towards identifying the most reliable method for risk assessment, hazard zonation and mitigation in prone areas of the country.

1.3 Objectives

1.3.1 Main Objective

To compare the relative performance of GIS based methods to Machine learning Algorithms in landslide susceptibility modeling

1.3.2 Specific Objectives

i. To investigate landslide causative factors and conduct an extensive analysis on the current landslide inventory of the study area
ii. To establish a correlation between landslides location and the spatial distribution of the causative factors using machine learning algorithms and GIS based methods
iii. To evaluate the efficiency of the models and their relative predictive performance

1.4 Relevance of the Research

Studies applying GIS through the use of pixel based models, Digital Elevation Models and remote sensing in susceptibility mapping have been conducted on both landslides in Rwenzori and Elgon, Uganda. However, studies adopting the machine learning algorithms like the k-NN, SVM, random forest modeling, decision tree models in landslide susceptibility modeling are limited in as far as risk assessment and management is concerned. Furthermore, this study brings forth a comparative analysis between the two approaches with the sole aim of establishing the relative advantage of each approach to the other in landslide susceptibility assessment for the study area. This will be a step towards optimal regional planning, disaster preparedness and risk reduction in landslide prone areas of the country.

1.5 Working Hypothesis

i. Landslide triggering factors have a spatial association to landslide locations and Machine learning Algorithms have a higher predictive potential in landslide susceptibility modeling than the GIS based models.

1.6 Research Questions

i. What are the main triggering factors of landslides in Bududa?
ii. What is the predictive performance of the GIS Based Methods?
iii. What is the predictive performance of the Machine learning Algorithms?
iv. What are the comparative advantages and limitations of GIS Based Methods and Machine learning Methods in landslide susceptibility modeling?

1.7 Scope and limitation

This research was limited to the study region of Bududa, situated in the Eastern region of Uganda. It focused mainly on attaining the research objectives, answering the developed research questions, the data available, models accessible, and time frame allocated.

1.8 Thesis structure

This work is subdivided into five broad chapters as described below;

Chapter One: Introduction. Entails the problem statement, the objectives that specify the research goals, information to be gathered, and questions or hypotheses to be tested.

Chapter Two: Literature Review. Discusses the strengths, limitations, and, gaps in previous studies with citations from the most recent publications.

Chapter Three: Methodology. Describes the research design clearly stating the methods/procedures used to attain the proposed objectives of the research, the detailed description of Bududa region, data sources, collection and analysis,

Chapter Four: Results and Discussion. Includes the Quantitative and Qualitative narrative of the findings, it contains tables, maps, figures with their interpretative discussions

Chapter Five: Conclusion and Recommendations. Gives a brief overview of the objectives in relation to the major findings and gives a conclusion in that regard. According to the study, recommendations in policies, methods or models is given and also points to unanswered questions that require further research

2.0 LITERATURE REVIEW

2.1 Introduction

Landslides by definition refer to the movement of a mass of rock, debris, or earth down the slope under the influence of gravity(Broeckx et al., 2019; Richefeu & Villard, 2016). They are triggered by a wide range of factors, natural and man-made factors. Natural factors are the naturally occurring processes such as rainfall, water level changes, stream erosion, earthquakes, snowfall, and volcanic activity among others (Margottini et al., n.d.) and man-made factors such as land use changes, different land management practices, and settlement (Khoshtinat et al., 2019) among others. These activities accelerate the occurrences of the landslide phenomena through the continuous distortion of the soil slope and its hydrological response

2.2 Global impacts of landslides

Globally, natural disasters lead to catastrophic out-comes to the environment and all its dwellings. These disasters are on the rise due to the significant climate variations experienced worldwide. They are subdivided in two categories 1) hydrological disasters that are triggered by climate changes and 2) Geophysical disasters that are a result of land deformation changes. Hydrological disasters take up the largest percentage (51.7%) (Moresi et al., 2020) and the geophysical disasters are the least with 9.1% .According to (Dang et al., 2020) the world recorded 3876 Landslides since 1995 to 2014, 163,658 human lives were lost to the disaster and 11,689 lives affected. In 2016, 8133 deaths were reported, 569million people were affected and 154billions in US dollars was spent in damages. From 1979 to 2002, Italy lost a total of 50,593 lives to an approximate of 2580 floods and landslides and she recorded a total of more than 500,000 landslides in 2020 (Ballabio & Sterlacchini, 2012; Moresi et al., 2020; Segoni et al., 2015). In 2015, Vietnam recorded 250 landslides, 779 deaths and 426 injuries due to landslides and flashfloods (Dang et al., 2020; Thanh Noi & Kappas, 2017). In Slovakia, 4% of the country is covered in landslide prone regions (Vojteková & Vojtek, 2020). The landslide menace is recurrent on most continents, Africa inclusive

Landslide studies in Africa are still limited and underrepresented in scientific literature. This is due to data scarcity, lack of expertise in dealing with the phenomena and limited funds to sponsor further exploration into advanced methodologies (Broeckx et al., 2018). Most African countries face difficulty in the spatial analysis and modeling of the landslide disasters due to the

sophisticated technicalities involved thus limited literature review available and missing inventory databases (Namono et al., 2019). In South Africa, according to (Odhiambo et al., 2019), all attempts to integrate geo hazard data were limited by the few documentation compiled on the mass movements in the country. This was due to the fact that only Mount Curie debris in Kokstad, KwaZulu-Natal, Lake Fundudzi Paleo-landslide in Limpopo Province and the Ukhahlamba-Drankensburg foot slope landslide in Kwazulu-Natal had been documented (Singh et al., 2008) despite the high number of mass movements occurring in the country. In Table (2-1) below, an overview of landslide occurrences in most African countries is displayed. South Africa takes the lead with 4399 landslides as by 2018, followed by Morocco with 3750, Angola with 1106. Uganda had a total of 967 as at the conducting of the study by (Broeckx et al., 2018)

In Uganda, landslides are pervasive in the eastern and western parts of the country, affecting regional development, deterring community livelihood and claiming human lives. (Jacobs et al., 2018; K. M. G. Kitutu, 2010; Namono et al., 2019; Vlaeminck et al., 2016). This has been attributed towards the evolving variation in the causative factors such as change in land use, change in rainfall patterns and intensities (M. G. Kitutu et al., 2011; Namono et al., 2019). Coping with the risks associated with landslides has been through preventive resettlement (Vlaeminck et al., 2016). Landslide analysis has been through regional farmers' perception and indigenous knowledge derived from past experiences, for instance, farmers in Bulucheke, Bududa submitted 3 factors; steep slopes, flow of underground water and, concavities as the main triggering factors (M. G. Kitutu et al., 2011) while other studies stressed, it was the prolonged low intensity rainfalls in the region (Namono et al., 2019).

Table 2-1: Overview of all African mapped landslides by country (Broeckx et al., 2018)

Country	**Total # LS**	**Sources**
Algeria	588	Bouhadad et al., 2010; Bourenane et al., 2015; Busche, 2001; Capot-Rey, 1954; Gabert, 1984;
Angola	1308	Dinis et al., 2013
Benin	8	
Burkina Faso	3	
Burundi	204	Nibigira et al., 2013

Cameroon	585	Afungang, 2015; Ayonghe et al., 2004, 1999; Che, 2011;
Central African	6	
Chad	87	
Comoros	33	Audru et al., 2010
Congo Republic	77	
Cote d'Ivoire	2	
Djibouti	33	
DR Congo	916	Maki Mateso and Dewitte, 2014; Moeyersons et al., 2010, Moeyersons et al., 2004;
Egypt	120	Arnous, 2011; Emam et al., 2010
Eritrea	74	
Ethiopia	801	Asfaw, 2010; Asma, 2013; Broothaerts et al., 2012; Girma, 2010; Hagos, 2012; Ismail, 2013;
Gabon	31	
Ghana	16	
Guinea	23	
Guinea-Bissau	0	
Kenya	325	Maina-Gichaba et al., 2013; Ngecu and Mathu, 1999
Lesotho	35	
Liberia	6	
Libya	174	Busche, 2001; Ostaficzuk, 1973
Madagascar	411	Ramasiarinoro et al., 2012
Malawi	177	Msilimba, 2010; Msilimba and Holmes, 2010
Mali	35	
Mauritania	129	Busche, 2001
Mauritius	5	
Morocco	3750	Choubert and Ennadifi, 1970; Fonseca, 2014
Mozambique	334	Bomans, 2005

Namibia	50	
Niger	146	Busche, 2001
Nigeria	191	Agbor et al., 2014; Igwe and Fukuoka, 2014; Okagbue, 1994
Reunion	21	
Rwanda	72	Moeyersons, 2003; Moeyersons et al., 2004
Sao Tome and Principe	5	
Senegal	2	
Sierra Leone	13	
Somalia	73	
South Africa	4399	Bijker, 2001; Chiliza and Richardson, 2008; De Lemos, 2013; Gupta, 2001; Hardwick, 2012;
South Sudan	45	
Sudan	41	
Swaziland	10	
Tanzania	958	Bomans, 2005; Kimaro et al., 2010; Temple and Rapp, 1972
Togo	11	
Tunisia	588	Bonvallot, 1984; Busche, 2001; Dimanche and Hamza, 1978
Uganda	967	Broeckx et al., in prep; Jacobs et al., 2016, 2017; Knapen et al., 2006; Maertens, 2016
Western Sahara	29	
Zambia	36	
Zimbabwe	7	
Total	**18,053**	

2.3 Types of Landslides

There exists 2 major categories of landslides and these include A) rotational slides and B) translational slides and these are identified according to the causes, location, where they occur and long term movement however, some individual landslides are often a result of one or more types of mass movement which still makes it difficult to classify. The dominant type of movement (i.e. topple, fall, spread, slide, flow, creep or slide) and the rock composition of where they fall (i.e. consist of bedrock or regolith) mostly is used to classify the landslide type as rotational or translational. In fig (2-1) the existing types of landslides and the existing types of movements are displayed (*Types of Landslide | NatureScot*, n.d. 2020)

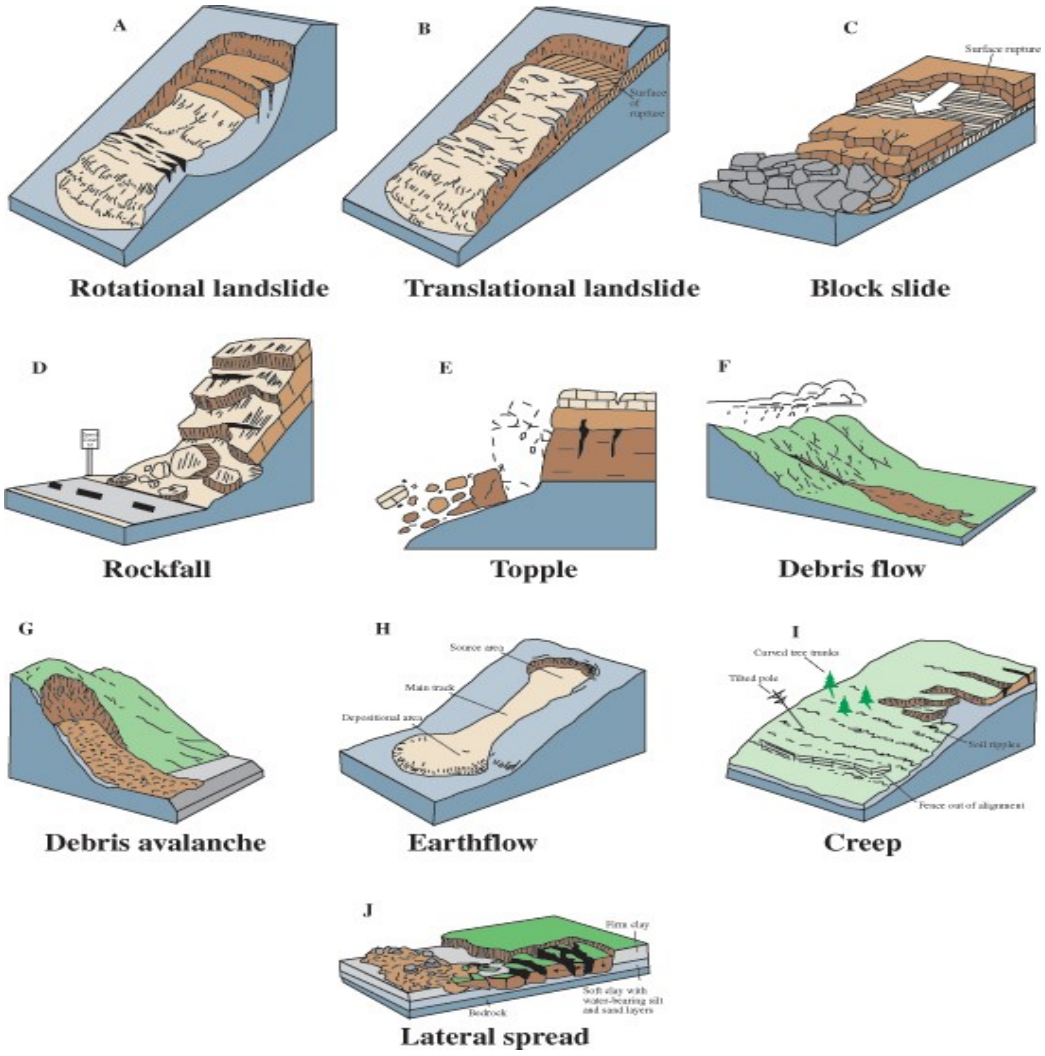


Figure 2-1: Landslide types

2.4 Landslide Inventory Mapping

A landslide inventory is a record of the dates, types, number, magnitude, and location of previous landslides, as well as the resulting damages (death toll, missing persons, infrastructure damage, etc. (Broeckx et al., 2019; Nakileza & Nedala, 2020; Vojteková & Vojtek, 2020; Xiao et al., 2019). It serves as the foundation for susceptibility studies and hazard zonation. These studies are critical tools in risk mapping and assessments to establish highly prone areas or regions. Susceptibility models that depend on the input variables of past landslides to make future predictions largely depend on Inventory datasets (Broeckx et al., 2019; Yao & Dai, 2006). Thus, each country prone to natural disasters has created an inventory database containing all necessary information on past landslides or other related disasters with their respective data collection techniques(Gorokhovich et al., 2013).

There are several methods for developing landslide inventory datasets, including 1) Geomorphological field mappings, which involve physical site inspections and data collection on location, extent, and damages of landslides(Jacobs et al., 2017). 2) Scientific literature, which entails a thorough examination of historical documents as well as published reports, articles, and newspapers,(M. G. Kitutu et al., 2011) 3) aerial photographs and airborne LiDAR generated DEMs which are conducted with the aid of remote sensing techniques that include the use of pixel and object automatic extraction of landslides using passive or active optical imagery (Herrera, 2016; Jacobs et al., 2018; Thanh Noi & Kappas, 2017; Xiao et al., 2019). This technology allows for the identification of landslide locations, features and sites.

LiDAR derivatives are still widely adopted in developing new inventories and updating existing ones, however, Satellite SAR Interferometry techniques plus new data analysis algorithms are increasingly being applied for the same purpose (Jacobs et al., 2018; Xiao et al., 2020). Studies emphasize efficiency in integrating both remote sensing techniques and the traditional identification methods in landslide inventory creation and updating. In India, the National Remote sensing data center integrated the semi-automatic image analysis with satellite DTMs plus historical records to develop an event based landslide inventory database (Pourghasemi et al., 2013). Event-based landslide inventories aid in the temporal probability predictions, size distribution and the actual density of landslides which simplifies susceptibility assessments and hazard zonation (Western et al, 2013)

2.5 Landslide Susceptibility Modeling and Assessments

Landslide susceptibility assessments entail proper planning, encompassing geotechnical measures (Gorokhovich et al., 2013) to landslide data inventory inference and update, hazard zonation through susceptibility modeling (Jacobs et al., 2018). The occurrence of landslides pose a deleterious effect on people's livelihood and the environment hence need to identify regions that are susceptible to landslides to aid in optimal planning and lessen the impact associated with their occurrence. Landslide susceptible areas can be identified through the application of scientific methods or approaches to study the triggering factors such as slope, climate, soil texture, vegetation cover. Other factors such as human settlement, population, and land management practices in region also contribute to the magnitude or extent of the damage of the landslide.

Landslide triggering factors greatly influence landslide hazard zonation. Landslide Hazard refers to the probability of a damaging slope failure occurring in a given area in a given time (An et al., 2018). Therefore, for a conclusive landslide susceptibility assessments to be conducted, accurate collection of data on the landslide triggering factors on all past events and current events should be made. The data is analyzed for input geo-environmental parameters that are modeled into either statistical approaches or deterministic methods that apply mechanical laws to predict the probability of future landslide occurrences, slope failures, and evaluate the zones for the hazard, risks and the magnitude of the disaster (An et al., 2018; Ballabio & Sterlacchini, 2012; Dang et al., 2020; Westen, 2002). The nature of the hazard caused by landslides is still complex to quantify unlike other disasters like earthquakes, hurricanes where measurement units have been agreed upon. The damage caused by landslides is currently associated to death tolls, missing persons, infrastructure damage and environmental impact in terms of water resources quality deterioration.

Landslide susceptibility modeling has evinced its potential in alleviating risks associated with mass movement of soil, debris (Xiao et al., 2020). It brings to focus; the interdependence relationship between landslide locations and their proclivity triggering factors such as threshold rainfalls, vegetation/land cover use (Reichenbach et al., 2018), the slope angle and the hydrological covariates at the study area (Xiao et al., 2020). Different approaches have been applied to aid landslide susceptibility modeling; Machine Learning Algorithms, Statistical Approaches, Frequency Factor, Certainty Factor and the integrated hybrid versions of several

approaches to improve the efficiency of the results obtained from the study (Brenning, 2005; Canavesi et al., 2020; Costache et al., 2020; Dang et al., 2020; Marjanović et al., 2009; Wu et al., 2016; Xiao et al., 2020). The pixel by pixel logistic regression approach has been applied to evaluate the performance of different Digital Elevation Models (DEMs) derived from SRTM and TanDEM-X (Broeckx et al., 2019; Jacobs et al., 2018; Xiao et al., 2019)

The investigation of hybrid approaches has been a common practice globally in landslide prone areas and the comparisons of different model performances to establish errors, shortcomings associated with one model in relation to the other or to the study site conditions (Wu et al., 2016). However, this has been an arduous attempt in most African countries, Uganda inclusive due to the technicalities involved and data scarcity in regions prone to the landslides (Broeckx et al., 2019; Jacobs et al., 2018). Furthermore, the foot slopes in Bududa are extremely fertile, boosting agricultural production and income generation for farmers in the region (Broeckx et al., 2019; Gorokhovich et al., 2013; K. M. G. Kitutu, 2010; M. G. Kitutu et al., 2011). This often attracts resettlement in the afore-warned landslide prone regions by civilians which makes attempts to alleviate landslide risk futile.

2.6 GIS-Based Methods in landslide Susceptibility Modeling

GIS-based methods involve physically-based predictors such as the 4Slide model (Moresi et al., 2020), expert-based models such as the Multi-Criteria Analysis System (Vojteková & Vojtek, 2020) that apply the Analytical Hierarchy Process, a decision-making tool, and the Entropy Model to influence the weighting of the various input parameters (Xiao et al., 2019, 2020). GIS interfaces such as ArcGIS, QGIS with GRASS have the functions to analyze the various input parameters of the landslide triggering factors and apply methods like the weighted factor, geospatial modeling, the weight of Evidence among so many to aid susceptibility assessments.

The 4Slide is an ArcGIS add-on that integrates the TOPMODEL (a Topography-based hydrological MODEL), an infinite slope model applied to estimate saturated water levels and the vegetation root strength model that aids in identifying locations of shallow landslides through assessment of the effect of plant roots on slope stability (Moresi et al., 2020). The 4Slide tool allows integration of numerical data derived from the study site and the spatially distributed data using raster maps, the main function is to calculate the Factor of Safety (FS) index through analyzing slope values extracted from Digital Terrain Models at the study site. However, the

Weight of Evidence and information values is the most commonly used method in susceptibility assessments as it aids in causative factor investigation and susceptibility mapping

2.6.1 Weight of Evidence Approach

The concept of weight-of-evidence refers to the process of gathering and weighing multiple pieces of evidence to come up with a conclusion about an issue. It was proposed in 1906 by F.P.Agterberg for mineral potential exploration and later proved applicable in environment risk assessments (Elmoulat & Brahim, 2018; SETAC, 2018; Westen, 2002). It involves three main steps and these are displaced in the diagram below

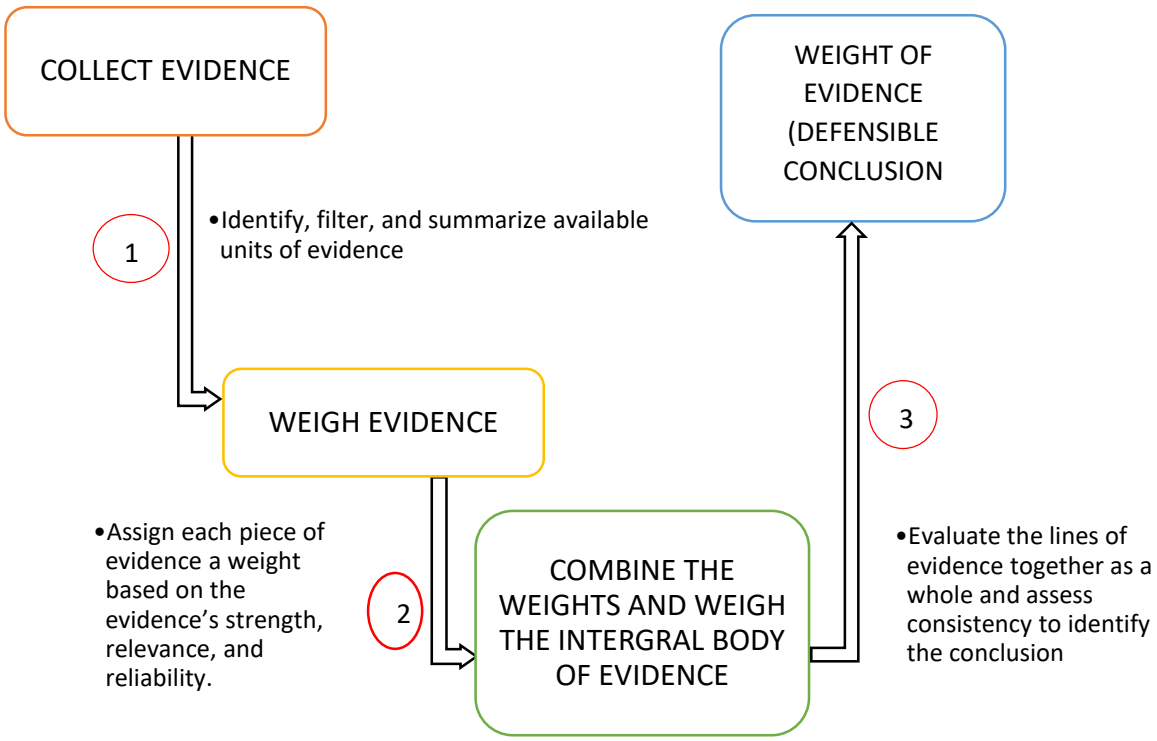


Figure 2-2: Weight of Evidence flowchart

Different units of evidence can be gathered through various methods of data collection, this could be through literature books, records, consulting stakeholders or field surveys. Evidence indicating positive correlation to the subject under study is identified, filtered out. Where possible and necessary, new studies can be conducted to fill gaps in the available data. Once salient evidence is obtained, it is assessed and assigned a weight. Finally, it is integrated and the whole body of evidence is weighed to establish a defensible conclusion

After the relevant classes of evidence are collected, weights to each class of evidence is assigned. This is to rank the outcome or the impact/importance of the evidence to the subject understudy or the scientific question being explored. Relevant evidence could identify which factors trigger the events under study through establishing which units are stronger or reliable to produce accurate assessments.

Weights can be represented in a variety of ways, ranging from more qualitative methods based on professional judgments to complex quantitative methods. The most important factor is that these weights are assigned and used in a transparent manner. Weights can be numerical values on a scale of words like "highly" or "weakly," positive or negative, or symbols like +, -, and 0.It is sometimes useful to construct a weight-of-evidence matrix in a weight-of-evidence assessment.

Table 2-2: Weight of Evidence Assessment criterion

Line of Evidence	**Relevance**	**Strength**	**Reliability**
1	+	++	**-**
2	++	**0**	+++
3	**-**	++	+

Where; + represents relevant/strong/reliable line of Evidence, - represents not relevant/strong/reliable line of evidence and 0 for neutral.

Finally, the entire body of evidence is evaluated using its corresponding weighted line of evidence in order to generate a balanced, weight-of-evidence interpretation for the problem at hand. This step can be difficult, however the weight-of-evidence framework provides a systematic approach to accomplishing it. It starts by integrating all of the evidence that has been gathered and weighted for a particular question. The integrated evidence is then interpreted.

Logic is applied to the comparisons of the evidence computed or gathered to the problem at hand and determine whether or not there is a probable solution or an answer to the developed working hypothesis. A body of evidence may appear inconsistent at times. The inconsistency in values may be explained by an expertise or a professional based on knowledge and experience, which in this case maybe be applied by a group of scientists with diverse skills on the matter

If the available evidence cannot answer the question, the assessors must reconsider other ways to apply information in the risk assessment. This entails to revising the original question in

attempt to investigate if there is a possibility that the wrong hypothesis has been developed or the wrong factors are being considered

2.8 Machine learning Algorithms

Machine learning Algorithms applied in risk analysis are the supervised learning algorithms that involve classifiers such as the k-Nearest Neighbor (k-NN), Support Vector Machine (SVM), and Random Forest (RF) in landslide susceptibility pattern recognition (Ballabio & Sterlacchini, 2012; Brenning, 2005; Bzdok et al., 2018; Dang et al., 2020; Rabby et al., 2020; Thanh Noi & Kappas, 2017). The Classifiers have the potential to conduct probabilistic inferences that are used for susceptibility indices (Dang et al., 2020). The k-NN is among the simpler algorithms, frequently used with Support Vector Machines, they are fitted to training and testing geo-data sets (input parameters and coordinates) in a GIS Environment (Potential, 2020). The confusion matrix or the Receiver Operating Characteristics Curve (Area under Curve AUC) are applied to establish the performance accuracy of the model results expressed as a percentage. (Ballabio & Sterlacchini, 2012)

2.8.1 K-NN Algorithm

K-NN is one of the simplest machine learning classifier, also referred to as the lazy learning algorithm. This is due to the simplicity of its performance by placing similar objects close to each other. In the k-NN classifier, k can be any integer, and if X_1 is the test object, the classifier finds all the k numbers that are close to X_1.Three distances are used to find the K numbers and these include; Euclidean distance, Manhattan, and Murkowski distances. The most commonly applied distance in landslide susceptibility assessment is the Euclidean distance using the formula in (2.1)

$$D(X1, x) = \sqrt{\sum} \tag{2.1}$$

KNN has a number of types and these include Fine KNN, Weighted KNN, Medium KNN, Coarse KNN, Cosine KNN and Cubic KNN

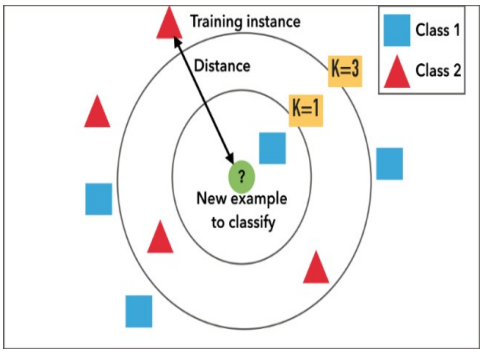


Figure 2-3: KNN training diagram

2.8.2 SVM Algorithm

SVM is a statistical learning theory- based algorithm founded by Vapnik in 1995. It works by separating classes using an optimal hyper-plane between class margins (Ballabio & Sterlacchini, 2012; Dang et al., 2020; Xiaowen & Min, 2009). The data points close to this hyper-plane also referred to as the decision surface are called ***support vectors***. SVMs can solve issues involving both linear and nonlinear classification and regression problems (Thanh Noi & Kappas, 2017). The support vectors close to the decision surface are part of the training dataset being used for the algorithm and the principle behind its functionality is to solve quadratic optimization problems.

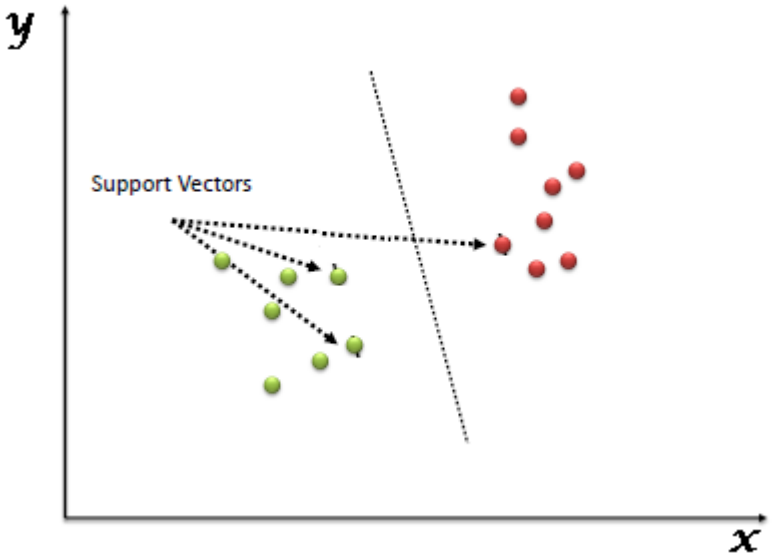


Figure 2-4: Support Vectors Diagram

If the problem is linearly separable, SVMs use the optimal separation hyper plane by maximizing the margin, which is the perpendicular distance between the support vectors close to the decision surface however, if the problem is nonlinear, SVMs use the kernel function, which transforms the input data into a higher dimensional space, converts it to linear, and solves it using the hyper-plane. (Dougherty, 2012). There are several kernel functions that are applied for this transformation and these include; Linear, Gaussian Kernel, Polynomial, Exponential, Laplacian, Bessel, Circular, Hyperbolic Tangent (Sigmoid) Kernel, Multiquadric, Inverse Multiquadric, Spherical, Rational Quadratic, Wave, Power, Spline Kernel, B-Spline (Radial Basis Function), Cauchy, and Chi-Square Kernel (Herrera, 2016). The most commonly used kernel functions are the Gaussian Kernel, Polynomial Kernel and Sigmoid Kernels

Two primary concerns can be recognized when dealing with landslide susceptibility mapping using statistical methodologies such as SVMs. The first concern is seen in the long - established parametric models' strong assumptions, which often presume data invariance. These models will perform badly if the relationship between the dependent and independent variables is not linear (Ballabio & Sterlacchini, 2012). Furthermore, the kernel machines in SVMs are quite sensitive to over-fitting the model selection criterion. Even in light of this drawback, SVMs remain one of the most common machine learning algorithms and are widely regarded as the preferred way for producing a high-performing algorithm with minimal adjustment.(Lee et al., 2017) This is due to their ability to tune model capacity to fit data complexity thus minimizing error tests and model complexity.

2.8.4 Random Forest

The Random Forest Classifier algorithm was developed in 2001 by Breiman, it integrates several algorithms to solve a regression or classification problem (Canavesi et al., 2020). It involves classifying a bunch of decision trees and also considered an improvement to both the decision tree and bagging models. Bagging is a type of ensemble model that combines decision trees. It uses the bootstrap sampling method to generate several decision trees. A large variance makes decision trees unstable thus enabling the algorithm unreliable too. Random forests have characteristics such as decision nodes, leaf nodes, and branches. The decision tree asks a series of questions about a training observation to classify it as a predictive class. The first step in the

decision tree is to divide the data set into two, with the goal of finding the most important predictor. The key parameters that determine the decision tree's performance are;

a) Feature selection and
b) Use of predictors during splitting,
c) Deciding when to stop

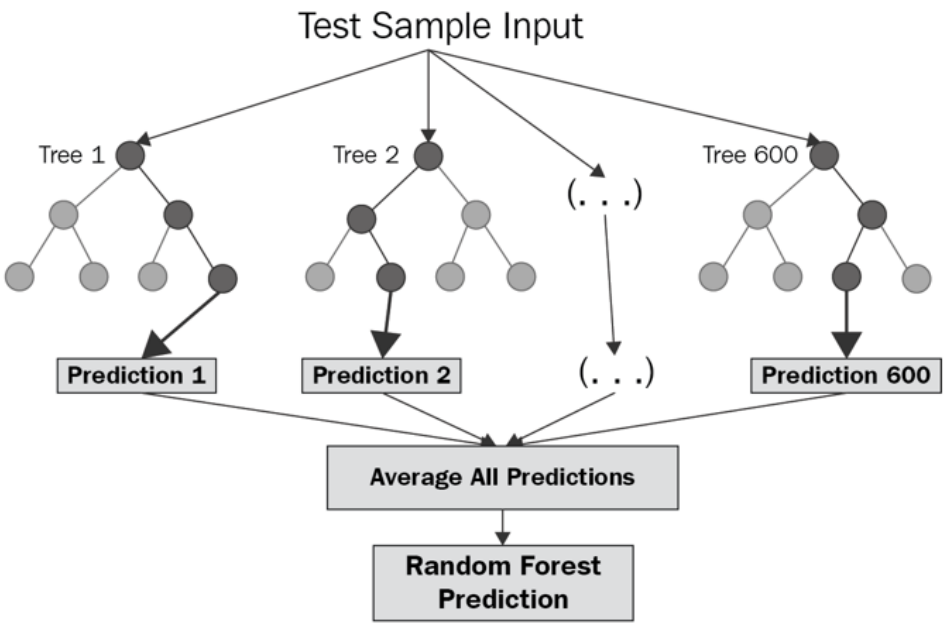


Figure 2-5; Random Forest flow Chart

In summary, models mimic reality, however, they are subjected to varying conditions at varying study sites, and a model with a high percentage of area under the receiver operating curves may unnecessarily indicate high predictive capacity (Xiao et al. 2020). Therefore, an extensive analysis in causative factors, model configurations, accurate spatial resolution, and data collection strategies possibly may enhance the predictive accuracy of any approach applied (Brenning, 2005). Results of the model selected have a critical influence on regional planning and risk disaster management for which they are applied.

3.0 METHODOLOGY

This chapter includes the research design clearly stating the methods/procedures used to attain the objectives of the research, the detailed description of Bududa region, data sources, collection and how it was analyzed to obtain results. The tools described are the GIS WoFE based models and Machine learning SVM Models applied in landslide susceptibility assessments. The ROC and confusion matrix is used to establish the predictive accuracy of the models.

3.1 Description of the Study Area

Bududa is a district bound at latitudes 1º04'N, 1º00'N and longitude of 34º15'E and 34º26'E in the eastern part of Uganda. It has an average elevation of 2050m above sea level (a.sl) with altitudes ranging from 1206m to 4317m. It lies west of Mt Elgon, a former active volcano during the Miocene epoch and has steep slopes of 60-70º. The slopes are made up of colluvial deposits that formed as a result of Mt Elgon's volcanic eruptions in the early middle Neogene, making them highly susceptible to mud landslides. The natives of this region are the Bagishu tribe; they cultivate the steep slopes of Bududa to grow food crops and cash crops that provide a source of income to enable their livelihoods (education for their children, health care and other necessities of life).

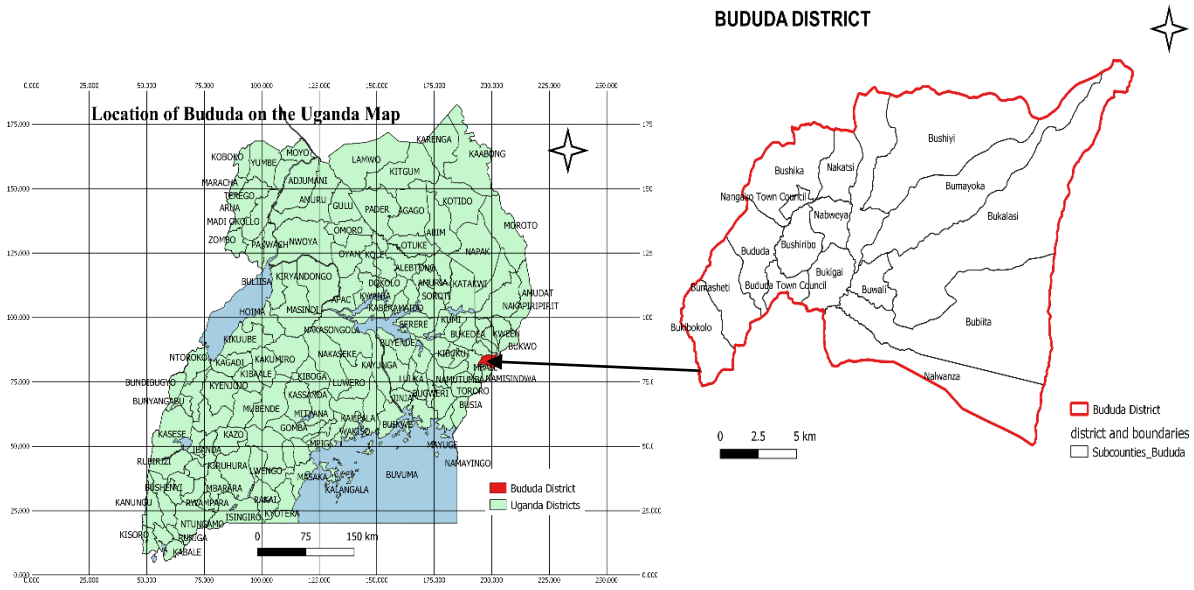


Figure 3-1: Location of the Study Area on the Ugandan Map

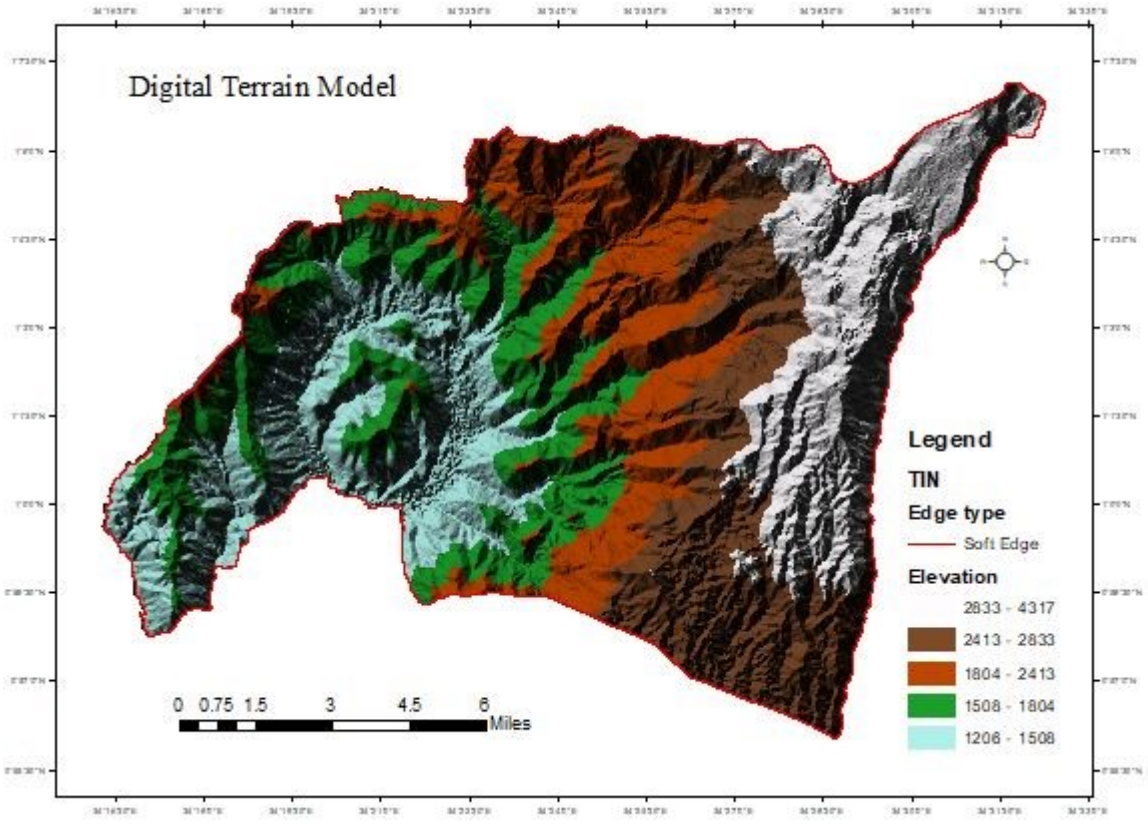


Figure 3-2: Overview of Bududa's Terrain

3.1.1 Population and Economic activities

According to the Uganda National Bureau of Statics, Bududa had a population of 79,100 persons in 1991 which grew to 123,100 in a span of 10years and continued to rise steadily at a rate of 4% annually to 210,173 persons in 2014 and was estimated at 271,100 in 2020. Human settlements in the region are attracted by the highly fertile soils that are suitable for farming food crops such as maize, beans, cassava and cash crops such as Arabica Coffee, Sugar Canes

With the surge in the coffee prices placing 1kg at 8000 Ugandan Shillings and a total area of 34,000ha used to cultivate coffee, the income generated from crop growing alone is a considerable economic asset for both the country Uganda and the local farmers in Bududa. The economic activities, are however frequently disrupted by the prevalent landslides which impedes the region’s steady economic growth

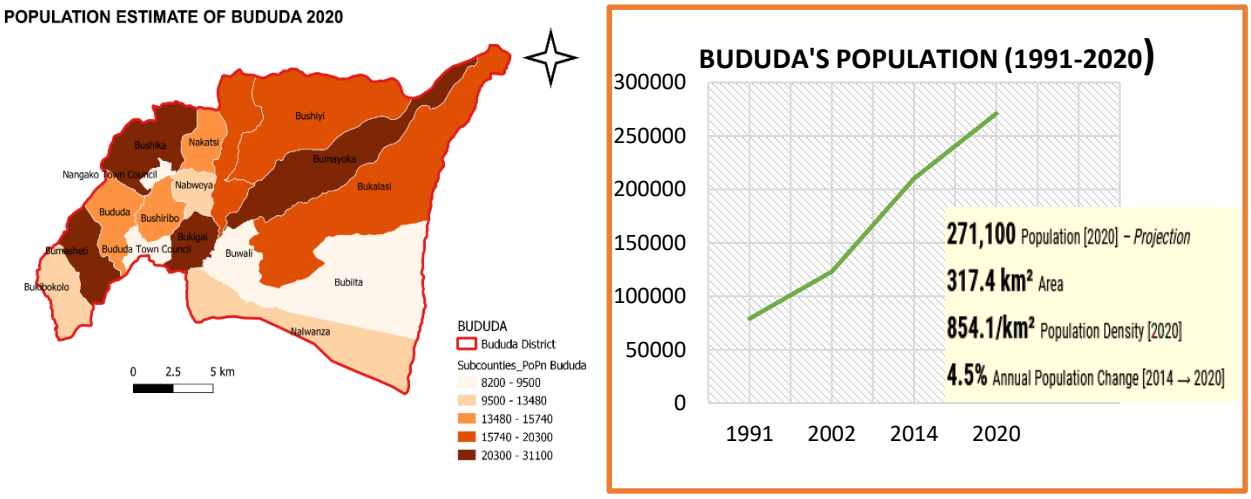


Figure 3-3: Population Estimate of Bududa 2020

3.1.2 Geology and Soils of Bududa

Bududa has three main lithology and these include 1) Butiriku carbonatites covering the central part which corresponds with the Sub-county of Bukigai, 2) a zone of fenitised basement rocks of Precambrian age surrounding this central carbonatite outcrop and 3) Mount Elgon agglomerates and tuffs situated in the north east of the district and falling within the borders of Mount Elgon National Park. These highly weathered rocks are composed of extremely fine pyroclasts of potash feldspar and are referred to as potash ultra-fenites.

The Geology around the Butiriku carbonatites is the coarse grained intrusive igneous rocks mainly composed of 50% carbonatites and syenite, a grayish rock composed of alkali feldspar. The intrusive rocks are surrounded with the metamorphic rocks of undifferentiated gneisses in the north, schist, quartzite and marble and lastly the volcanic rocks and their associated sediments along the slopes of the Elgon.

The soils weathered from these rocks and their distinguished classes are summarized in table 3

Table 3-1: Geology and Soils of Bududa

No	Soil Classes	Parent Rock	Soil Type
1	Humose Red Sandy Clay Loams	Elgon volcanics	Organic Soils, Non-Hydromorphic
2	Black Humose Sandy Clay Loams	Elgon volcanics	Organic Soils Non-Hydromorphic
3	Red Clay Loams And Sandy Clay Loams	Elgon volcanics and Basement Complex granites	Ferrallitic Soils - Not Differentiated
4	Yellowish Brown Sandy Clay Loams	Basement Complex granites	Ferrallitic Soils - Not Differentiated
5	Red Clay Loams & Sandy Clay Loams,	Elgon volcanics and Basement Complex granites	Ferrallitic Soils - Not Differentiated
6	Red Sandy Clay Loams Occasionally Lateritized,	Elgon volcanics	Ferrisols
7	Red Sandy - Sandy Clay Loams,	Basement Complex granites and amphibolites	Ferrisols
8	Black And Grey Clays Often Calcareous,	River alluvium	Ferrisols

3.1.3 Climatic Characteristics

Rainfall

Bududa experiences a bi-modal rainfall pattern with the first season coming in March to May and the second season in August to September as displayed in the 2020 monthly totals in fig (3-4). A brief dry month is experienced in mid-June to mid-July while the dry spell from November to March is extensively longer with December experiencing the least amount of rainfall. The average rainfall is 1872mm per annum. These high amounts of rainfall in Bududa are key to agriculture production thus revenue for the farmers in the region.

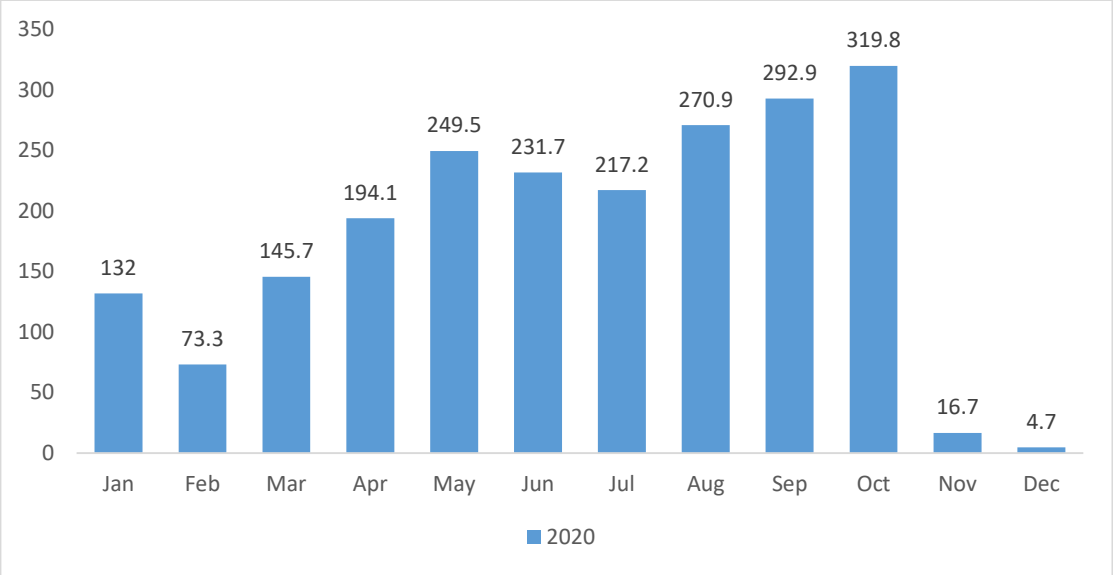


Figure 3-4: Bududa Monthly rainfall (2020) (Source: UNMA)

Temperature

The average monthly maximum temperature in the district ranges between 25°C and 29°C according to UNMA. This results in a long-term dry spell that lasts from December to March. During the months of January and May, relative humidity ranges from 80% to 90% in the mornings and from 61% to 66% in the afternoons. The pixel map in fig (3-5) shows the variability of temperature over the sub counties in Bududa for the year 2018

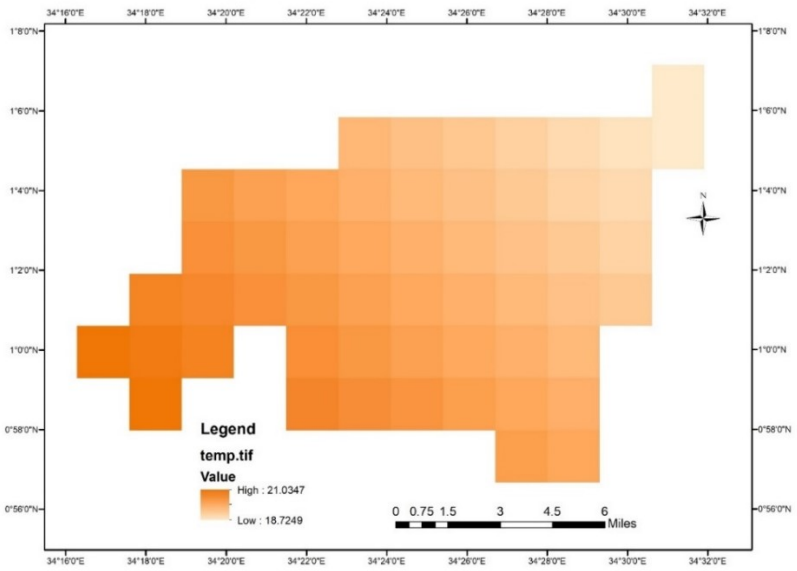


Figure 3-5: Bududa's Temperature pixel map (Source: Earth Consult)

Wind

Due to its proximity to the equator, the district is swept by the South East and North East monsoon winds. It has a mean annual vapor pressure of 18-20 mm hg and the peak potential evaporation occurs in March.

3.1.4 Ground water and the Hydrological Characteristics

Despite the district having plenty of streams and rivers, the overall area under water is insignificant in comparison to the entire land area. This is due to the lack of large water bodies, such as lakes, and the steepness of the terrain which allows for immediate runoffs when the rain comes falling down. As a result, wetlands that would have retained some water are constantly being drained by rivers and streams. When compared to the rest of the country, which is covered with water to the tune of 25%, the district has 12 km^2 of land under water/wetlands, or 4.3 percent of the total land area. The main rivers in the district include Manafwa, Uukha, Tsutsu, Liisi, Malabasi, Kasuuni, Suume and Khokhobi as shown in fig (11) below

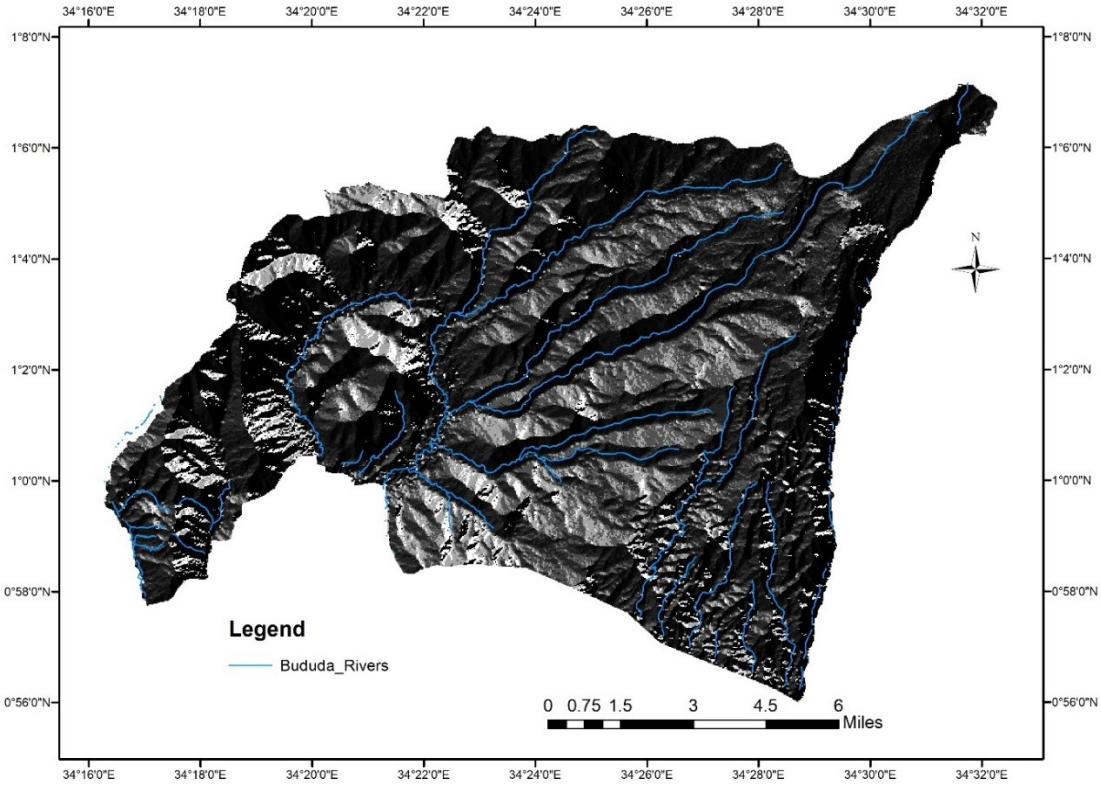


Figure 3-6: Map showing Bududa's rivers

3.1.5 Vegetation and land use

Due to the mountainous terrain of the region and the high population growth rate, there is barely any consistent natural vegetation in the district as it is cleared for human settlement and other developments. (40%) of the vegetation is the well-stocked high tropical forest and alpine vegetation mixed with a copious part of the bush towards the peak of Mountain Elgon. Bamboo forests, moors and fern also exist on the mountain summit. Other vegetation zones constitute of grasses and swamps. At the foothills of the mountain is the subsistence farmland surrounded with human settlement and other economic developments like roads, schools, hospitals etc. The map in fig (12) below shows the land cover types as distributed in the district.

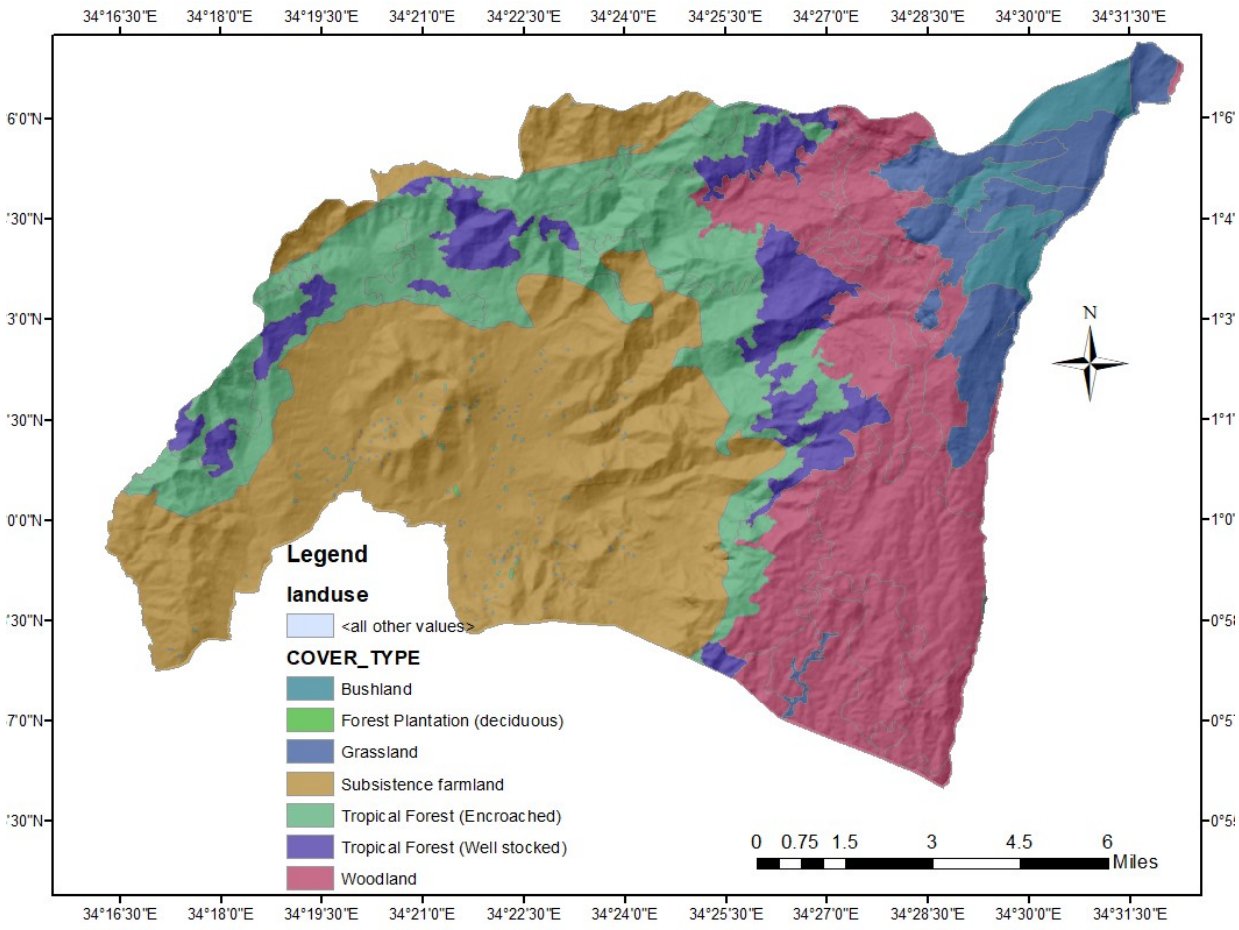


Figure 3-7: Map Showing the Land cover of Bududa

3.2 Research Design (Methodology Flow Chart)

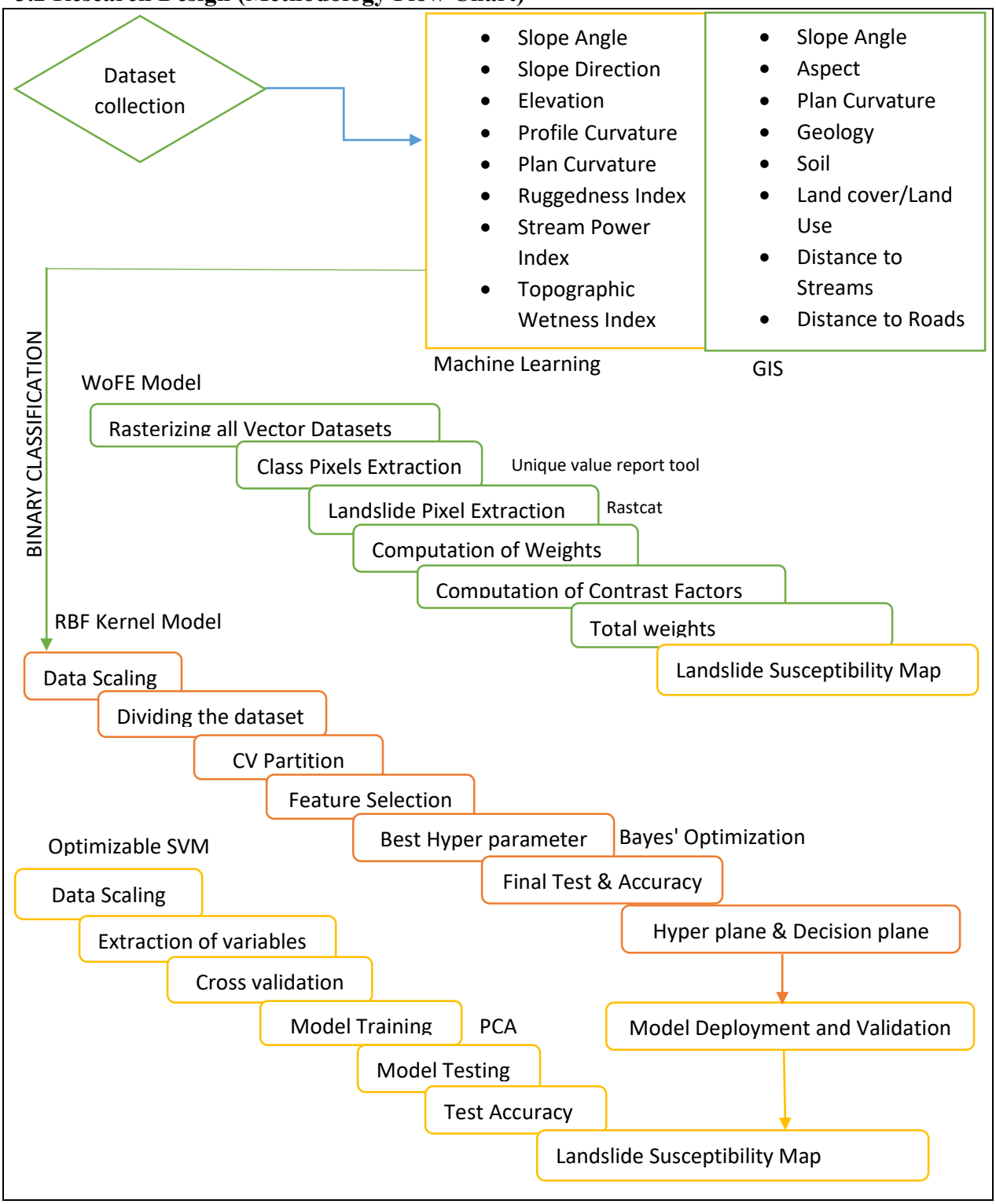


Figure 3-8: Methodology Flow Chart

3.3 Data Collection

The datasets were collected from the respective government ministries responsible for the particular discipline of the study. The rainfall data was collected from the Uganda National Meteorological Authority (UNMA), under the Ministry of Water and the Environment. The Geological datasets and lithological data was collected from the Directorate of Geological Survey and Mines, under the Ministry of Energy and Mineral Development. The landslide inventory data was retrieved from the Department of the Disaster Preparedness and Management under the Office of the Prime Minister (OPM). The Digital Elevation Models were downloaded from the United States Geological Survey earth explorer website https://earthexplorer.usgs.gov . Other datasets like the recent Uganda administrative boundary shape files were downloaded from the Uganda Bureau of Statics website (UBOS). Table 3-2 gives a summary of the sources of Data used in this study

Table 3-2: Data sources and description

No.	Data	Description	Source
1	Uganda Subnational Administrative Boundary Layers, Population Data.	Districts, County, Sub County, Parishes, Roads, Rivers, Streams, buildings, waterways.	Uganda Bureau of Statics website (UBOS)
2.	Climatic Data.	Rainfall(annual precipitation, mm and maximum rainfall, mm) and Temperature	Uganda National Meteorological Authority (UNMA)
3.	Landslide Inventory Data	Number of landslides, Spatial distribution, location, Damages, deaths.	Department of Disaster preparedness and Management (DDPM)
4	DEM	Slope Angle, Aspect, Curvature, TRI	Earth Explorer (United States

Geological Survey)

website

3.4 Data Preparation and Analysis

3.4.1 Terrain Analysis

Upon downloading the 30m resolution Digital Elevation Model of Bududa using the SRTM Downloader, a terrain analysis was conducted using the raster tool in QGIS/ArcGIS to extract the different data layers of the study which include Aspect (Slope direction), Elevation, Slope Angles, Hillshade, Plan Curvature, Topographic wetness index and the Stream network, as shown in fig (3-9) ;

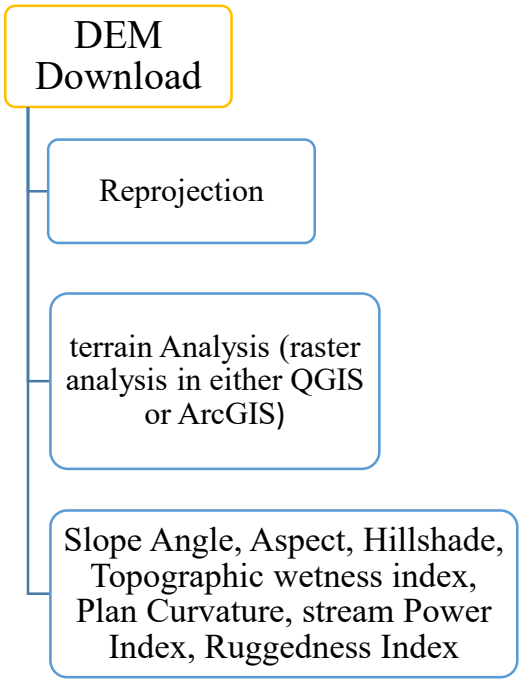


Figure 3-9: Terrain analysis flow diagram

3.4.2 Reclassification

The data layers extracted were then reclassified by table in ArcGIS under the spatial analyst tab to assign new values of preference to the raster datasets. The new classified layers were then applied in further analysis and modeling. The layers were reclassified in groups shown in table (3-3) below

Table 3-3: Classification of Factor layers

Data Layers	Classes
Slope	0-9, 10-15, 16-21, 22-28, 29-35, 36-46, >47
Aspect	North(Flat), North East, East, South East, South, South West, West, North West, North
Plan Curvature	-2.2—5, -4.9—2.0, -1.9—1.0, -9.3—4.6, -4.5—1.0, 1.1—6.7, 6.8-1.8, 1.9-4.8, 4.9-2.6
Elevation	1205-1605, 1605-1962, 1962-2379, 2379-2779, 2779-3223, 3223-3729 >3729
Land use	Bush land, Forest Plantation, Grassland, Subsistence farming, Tropical Forest (encroached), Tropical Forest (well stocked), Woodland
Soil	RS-SCL, YBSCL, BHSCL, HRSCL, RSCL(lateritized), RCL&SCL,BGCL(Calcareous)
Geology	(carbonatite and syenite),(volcanic rocks and associated sediments, (Undifferentiated gneisses)
TWI, SPI, TRI	1,2,3,4,5
Distance to Streams & Roads	1,2,3,4,5

The reclassified groups were exported as images (img.) to an external file where the Rastcat program was used to count the landslide pixels within each class group. The pixels within each class group were counted using the unique value table in QGIS or in the attribute table of the rasterized layers in ArcGIS. The number of pixels counted were further modeled using the weight of evidence formulas as discussed in the subsequent steps

3.5 Landslide Inventory Mapping

The landslide Inventory data retrieved from the Disaster preparedness constituted landslide information from 1933 to 2014, with data from 2015 to 2020 missing or pending updating in the database, therefore, the Google satellite map was applied in extracting the spatial locations of landslide sites, relict and recent ones. These were overlaid on the Bududa slope map to create the

inventory map that is later used for modeling as the causative factors at these landslide locations are extracted for analysis

3.6 Weight of Evidence Modeling

Upon extraction of class pixels and the landslide pixels using the Rastcat program, the weights of evidence for landslide occurrence were calculated using the posterior and prior probability formulas below

$$\text{Prior Probability} \quad \frac{\text{Total Number of Landslide Pixels}}{\text{Total Number of Class Pixels}} \tag{3.2}$$

Posterior Probability = Prior Probability {with evidence} also called the conditional probability due to its ability to include landslide conditioning factors. The computation entailed the following terms;

N Map; total number of pixels in the entire map

N slide; landslide pixels in the entire map

N Class; number of pixels in a given class

NSL class; number of pixels having landslides in that class

Weight formulas that were used include

NPix1=NSL class

NPix2=N slide-NSL class

Npix3=N Class-NSL class

Npix4=N Map-N slide-N Class +NSL class

Using the weight formulas above, the positive (W+) and negative weights (W-) of each factor were computed using the formulas below;

$$W+ \quad = \quad \text{In}\left[\frac{\frac{\text{Npix1}}{(\text{Npix1}+\text{Npix2})}}{\frac{\text{Npix3}}{\text{Npix3}+\text{Npix4}}}\right. \tag{3.3}$$

$$W- \quad = \quad \text{In}[\frac{\text{Npix2} * (\text{Npix3} + \text{Npix4})}{(\text{Npix1} + \text{Npix2}) * \text{Npix4}}] \tag{3.4}$$

The total weights of each factor map were then obtained by summing up the positive weights of the factor itself and the negative weights of other factors using the equation below

$$\sum \text{W}_{\text{map}} = \text{Wplus} + \text{Wmintotal} - \text{Wmin} \tag{3.5}$$

Where; $W_{mintotal}$= total of all negative weights.

After the final weights were computed, the contrast factors were also computed using the formula below;

$$\text{Cw} = (\text{W}^{+} - \text{W}^{-}) \tag{3.6}$$

Where C= 0 represented overlapping between the landslide pattern and Map class pattern , positive contrast values represented positive association between landslides and class patterns while negative contrast values represented absence of correlation between landslide occurrence and the factor

3.6.1 Landslide Susceptibility mapping using WoFE

The total summation of weights computed for each respective factor map were multiplied with the corresponding layers using the raster calculator of the spatial analyst tool in the GIS interface to generate a map. The resultant map is the landslide susceptibility map obtained through the use of the WoFE model which is further reclassified as low, moderate, or high to indicate highly susceptible zones from lowly susceptible zones. The raster calculator formula used is as follows;

$$\begin{aligned} &Landslide\ Susceptibility\ Map \\ &\quad = \{Slope * Ws + Aspect * Wa + Elevation * We \\ &\quad + Plan\ Curvature * Wp + Geology * Wg + Soil * Ws \\ &\quad + Land\ use * Wlu + Distance\ to\ Roads * Wdr \\ &\quad + Distance\ to\ Streams * Wds\} \end{aligned} \tag{3.7}$$

To validate the model's performance, the ROC Curve showing area under curve of the sensitivity values to the false discovery rates was generated

3.6.2 Field Validation of the Model

In additional to the ROC Curve, field visits were conducted to collect 10 spatial coordinates of the recent landslide sites from 2015 to the most recent using the global positioning coordinate receiver. They were uploaded into the GIS interface and overlaid onto the landslide susceptible map to further validate the susceptibility zones generated on the map

3.7 Machine Learning Modeling

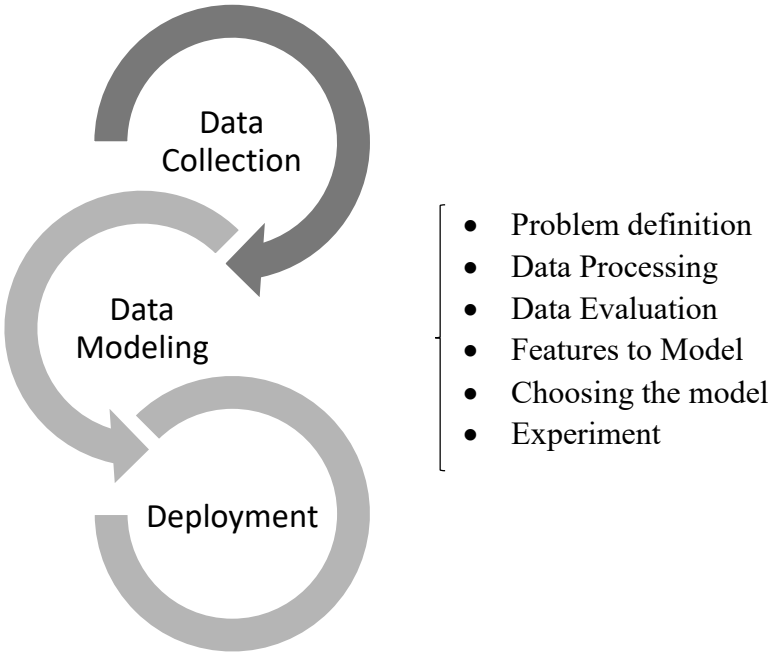


Figure 3-10: Machine learning Modeling

3.7.1 Data Collection

The data layers extracted from the Digital Elevation model are used in the algorithm training and modeling. The layers applied for machine learning were raster images and these include; slope angle, slope direction, elevation, profile curvature, plan curvature, ruggedness index, stream power index and the topographic wetness index. Using the research tool in QGIS, random points over the study area were generated. A point sampling plugin was added to extract the respective variable values existing at the random points. The attribute table of the new layer containing the values of the 8 variables was exported as a comma delimited file with XY coordinates. In the

table, the random points are assigned a ‘present’ or ‘absent’ value to indicate the presence or absence of landslides at the chosen point

After data scaling, feature selection and the principal component analysis, the dataset was then modeled using the Support Vector Machines in MATLAB. The models selected for the study are the Optimizable and radial basis function Kernel SVM

3.7.2 Optimizable SVM

The structured dataset containing the different environmental variables was arranged setting the predictor variables to correspond to the 8 environmental variables and the response variable assigned to landslide spatial locations as earlier identified from the random points

The 5 fold cross validation scheme was chosen to split the dataset into 5 subsets. The first subset was designated as the validation set and the remaining four used for the training session. The accuracy of each of the different datasets was computed using the validation subset and the average values obtained are used to get the final cross validation accuracy of the model.

Feature selection is also another important aspect in machine learning modeling as some features are of less importance in the training than others. Therefore in this study the correlation and regression analysis was applied to select what variables have a significant impact or strong spatial relationship to the occurrence of landslides and the variables with a strong positive or negative correlation were chosen for the model to train on.

The Principal Component Analysis was also deployed on the data before the model training session to reduce the dimensionality of the dataset and increase interpretability while minimizing information loss.

After selecting a cross validation type, feature selection and adding a PCA to the dataset, the model was trained in the MATLAB classification learner workspace. The results in form of a scatterplot, a confusion matrix, ROC, parallel coordinate plot and the misclassification error plot are evaluated and assessed for the model’s performance. Furthermore, the trained model was exported into a new session and fitted on a new dataset to further test its predictive performance using the code below;

$$Yfit = OptimizableSVM.PredictFcn\ (‘New\ Data’) \tag{3.8}$$

Where; *Yfit* is the new predicted values of the model, *OptimizableSVM* is the trained classifier model, *PredictFcn* is the prediction function in the Matlab workspace and the *New Data* is the new table with similar variables like the training dataset

3.7.3 Radial Basis Function Kernel Model

The radial basis function kernel model is developed and trained in the Matlab workspace too using the following steps;

The dataset with the 8 environmental variables and the predictor variable is loaded into the workspace and split with a ratio of 80:20 for the training and testing dataset respectively. X is assigned the independent variables and Y the dependent variable which in this study is the landslide spatial locations. The division was performed using the steps below;

```
rand_num = randperm(size(X,1));
X_train = X(rand_num(1:round(0.8*length(rand_num))),:);
y_train = y(rand_num(1:round(0.8*length(rand_num))),:);
X_test = X(rand_num(round(0.8*length(rand_num))+1:end),:);
y_test = y(rand_num(round(0.8*length(rand_num))+1:end),:);
```

The CV partition function was deployed on the dataset to make a random partition that defined the training and validation sets as follows;

c = cvpartition(y_train,'k',5);

Feature Selection is conducted using the wrapper methods to choose parameters that are important or key contributors to the model's performance and accuracy. This was necessary to remove all features that have no strong relation to the outcome of the model and are likely to slow it down while reducing its efficiency in learning. The forward sequential selection was adopted in this study and it evaluates each feature separately for the model number of iterations in order to obtain the optimum kernel function. The code processed to achieve this was;

[fs, history] = sequentialfs(classf, X_train, y_train, 'cv', c, 'options', opts,'nfeatures',2);

The training sets with the best features was obtained and used in the model fitting as shown below;

Md1=fitcsvm(X_train_w_best_feature,y_train,'KernelFunction','rbf','OptimizeHyperparameters','auto',...'HyperparameterOptimizationOptions',struct('AcquisitionFunctionName',...'expected-improvement-plus','ShowPlots',true));

The Objective Function Model was developed to establish the sequential feature selection function evaluations along the kernel scales and the established box constraints during the iterations.

The model was optimized using the Bayes' optimization algorithm that attempts to minimize the X scalar objective function in a bounded domain. After the optimization, functional evaluations, and training, the model was deployed on the test set with the forward sequential feature selection and its performance accuracy computed

X test w best feature = X test (:,fs);

*Test accuracy = sum ((predict (Md1, X test w best feature) == y test))/length(y test)*100*

From the models performance, hg scatters and g scatters giving an overview of the hyper plane or decision boundary by show of the support vectors selected in the training was generated.

The Model feature importance is used in developing the landslide prediction and susceptibility maps to indicate areas that are lowly and highly susceptible to landslides occurrence

3.8 Comparative Analysis

To compare the results of the two methodologies, the confusion matrix and the receiver operating curves generated from the two approaches was used. The confusion matrix displays the false positives, false negatives, the true positives and true negatives thereby providing a visual on the performance of an algorithm on a test data set where true values are known. The ROC curve gives the accuracy of the model by giving the percentage value of the area under the curve and also shows the sensitivity analysis which is the success rate versus the false positive rates on a scale of 0-1. The two susceptibility maps and the second inventory map were also used in identifying limitations and strengths of the two methods.

4.0 RESULTS AND DISCUSSIONS

4.1 Landslide Inventory Database

The inventory constituted data on the various natural disasters occurring in the region, with a record of 77 landslide hotspots recorded from 1933 to 2014. The landslides had no exact spatial location however they had sub county locations and the affiliated data of damages and reported triggering cause. Bududa as a region at the foot slopes of mountain Elgon has been hit by various natural disasters and these include earthquakes, mudslides, flashfloods, forest fires among others as shown in the pie chart in fig (4-1)

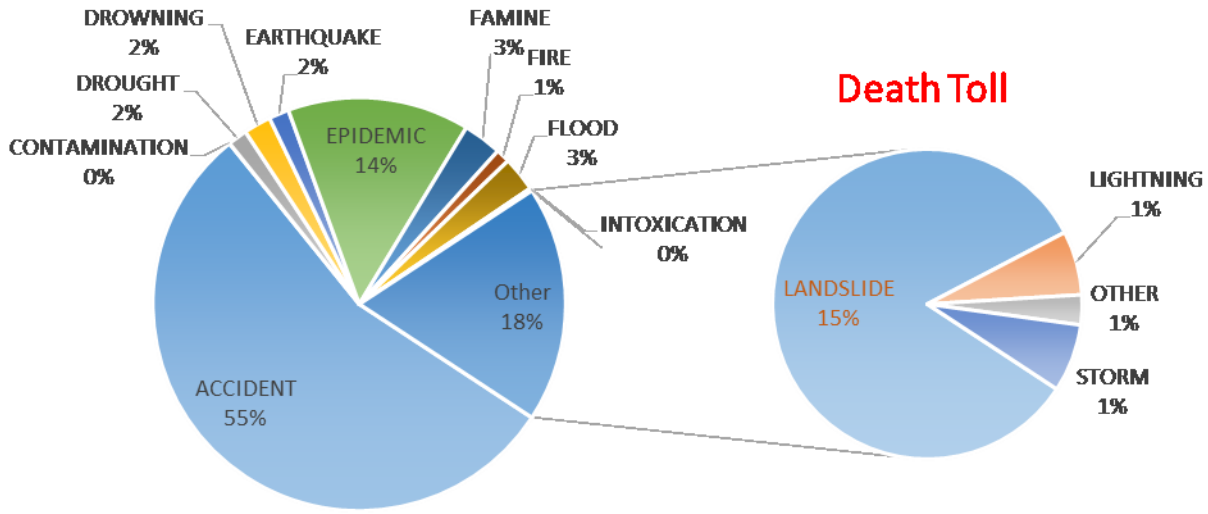


Figure 4-1; Natural disasters that have occurred in Bududa since1933 Source: DDPM

After accidents with a 55% death toll, landslides come second with a toll of 15% of total deaths in the region. Landslides are associated to other damages other than death such as injury, displacement, cropland destruction, livestock killings, and road destructions among so many. The graph in fig (4-2) displays the overall impacts of landslides on Bududa since 1933

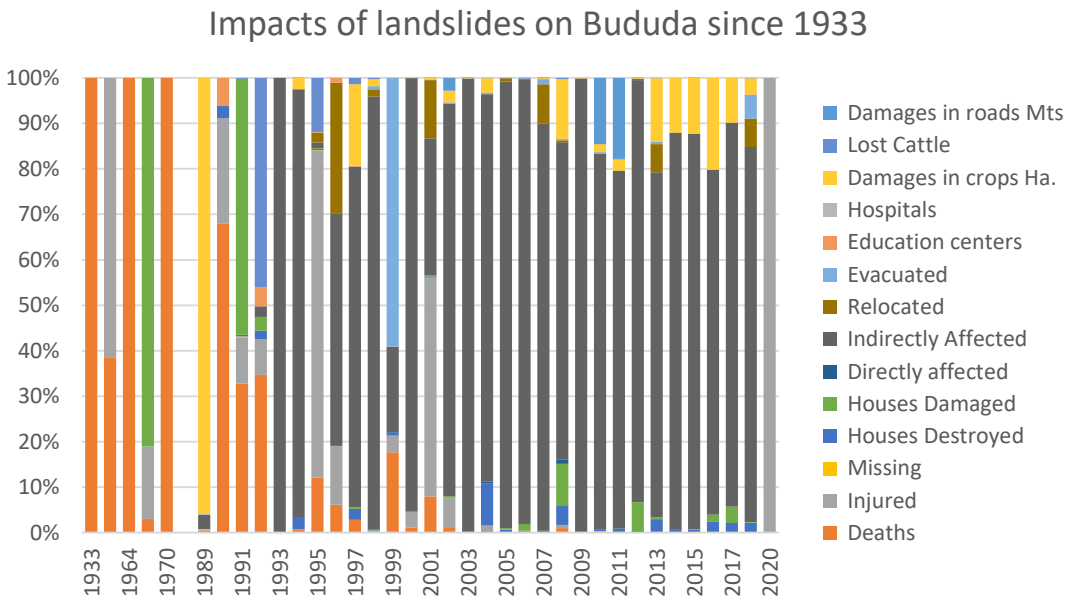
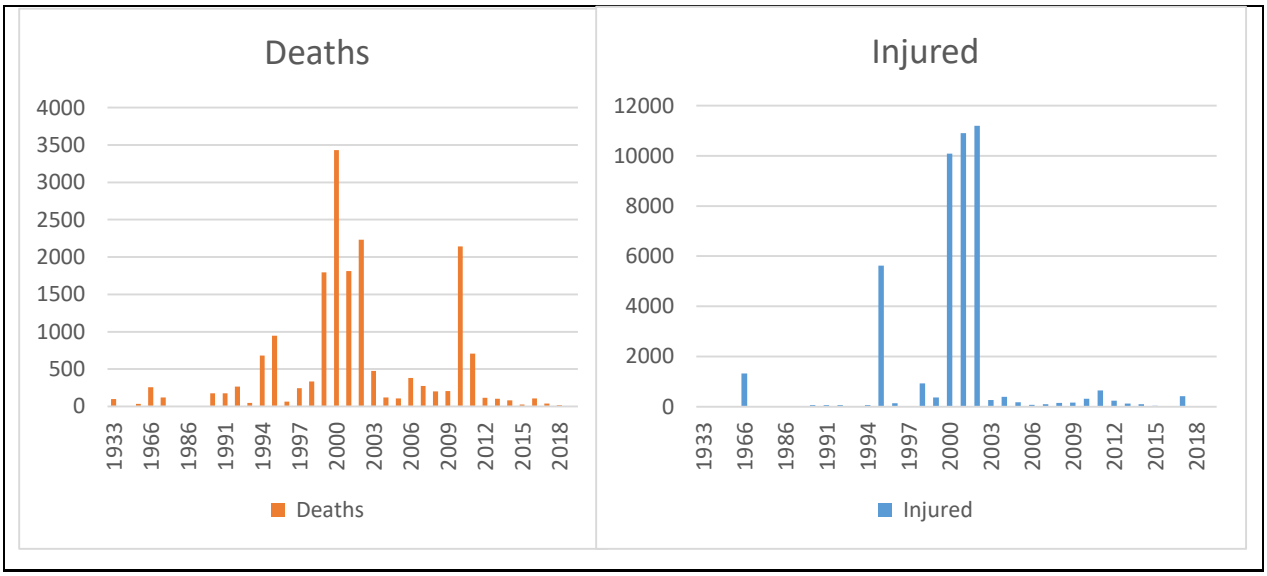


Figure 4-2: Impacts of landslides on Bududa since 1933

In the year 2010, Bududa was hit with a massive landslide that claimed a total of 1905 lives, injured 950 persons, damaged several houses and indirectly affected a sum of 305,797 persons. This raised awareness of the prevalent disaster of landslides and initiated a number of research and investigations into the causative factors and existing policies to help in resettlement, relocation and relief to the affected citizens of Bududa. The graphs in fig (4-3) display the adverse effects of landslides over the years to the people of Bududa



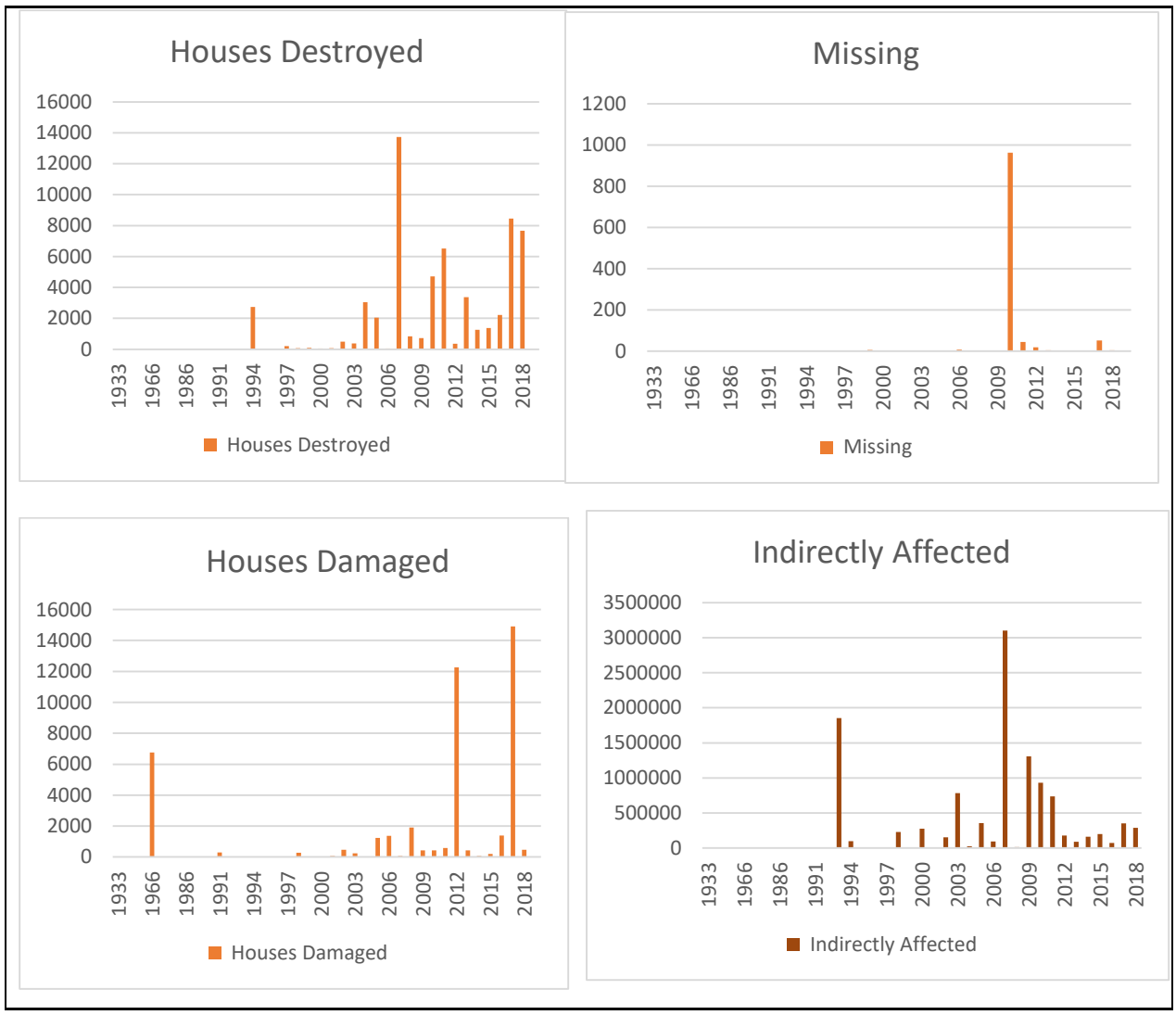


Figure 4-3; Bar graphs showing the death tolls, injured persons, missing, number of houses damaged and those that were indirectly affected by landslides over the years.

4.1.1 Landslide Inventory Map

The inventory map showing the spatial locations of landslides in the Bududa district is shown in fig (4-4). This spatial distribution of landslides is later applied in the causative factor pixel value extraction (i.e. elevation pixels occupied by the landslides) and all values were extracted for the 13 variables on a 30X30m resolution. These were further modeled using the weight of Evidence approach to establish their association to landslide occurrences as discussed in the subsequent steps. The landslide inventory comprised of 77 landslides, recorded up to 2014. For more locations, investigative approaches were conducted through the use of Google satellite imagery to identify scars and pin more locations. A total of I35 spots was located and added to the

inventory map. A Google capture of one of the landslide scars taken during the investigation is displayed in fig (4-4)

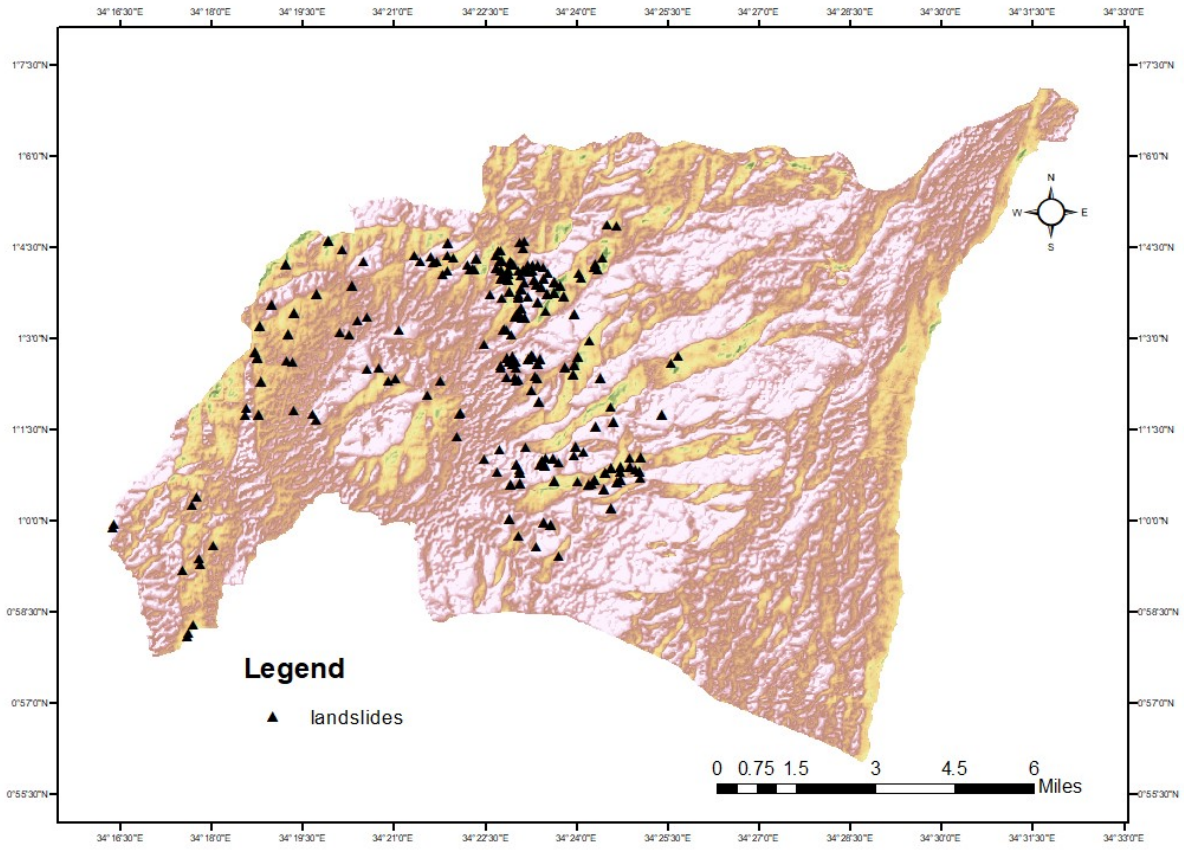


Figure 4-4: Landslide Inventory Map of Bududa

Figure 4-5: Google Satellite capture of one of the landslide Scars on 20th September, 2021

4.2 Landslide Triggering Factors

There are several factors explored in this study as triggering factors of landslides in the region. As stated by several authors, rainfall plays a significant role in causing landslides. This is due to its direct influence on the soil slope stability however, to trigger an occurrence a combination of two or more factors has to take place. There exits two categories of factors 1) internal factors such as slope angle, slope shape, direction, soil, geology, curvature and 2) external factors (anthropogenic activities) such as vegetation removal, infrastructure development, mining activities that excavate deep in the ground and underground leaks due to poor drainage systems. These are generally responsible for triggering factors. The 14 causative factors investigated in this study are a) Slope Angle, b) Slope Direction (Aspect), c) Elevation, d) Profile Curvature, e) Plan Curvature, f) Topographic Ruggedness Index, g) Stream Power Index, h) Topographic Wetness Index, i) Soil, j) Geology, k) Land cover, l) Distance to roads, m) Distance to Streams and finally Rainfall. The respective thematic maps of these triggering factors are displayed in fig (4-7)

The Aspect gives the direction the slope angles are facing (North, East, South or West) and the Slope angle map distinguishes very steep slopes that are susceptible to landslides, the moderately steep slopes and the flat areas in the region. For instance, the slope angle map in a) displays areas in dark blue with steep slopes of angles >31degrees and those in maroon have a 0 to 8° degree which implies relatively flat areas. Generally, according to the slope angle map in a), Bududa is an overly steep slope region which implies high shear stress acting on the hill slopes posing a greater risk to mass displacement in the presence of other favorable conditions like heavy rainfall or prolonging low intensity rainfalls.

The Plan Curvature in e) displays the convergence and divergence of the flows across the surfaces of the region. When the plan curvature is zero, it implies the surface is flat, however, when the values are between -4 and 4, it implies extreme relief with very steep slopes. The Plan and Profile Curvature gives an overview of the drainage characteristics of the region

The Geological composition of the region also has a significant role in landslide occurrences. This is due to the fact that different rocks have different permeability properties which instigate geomorphological processes that are often responsible for the type of landslide occurring in the region (rotational or translational). The lithology at the surface area also represents the strength

rating of the existing parameters that influence the detachability of debris from the slope mass. The rocks existing in the region are subdivided into 3 broad groups as shown in Table (4-1) below;

Table 4-1: Geological characteristic of Bududa

Age	Rock Type	Rock	Rock Description
Tertiary (& Mesozoic)	Intrusive	Carbonatite and Syenite center	Tertiary To Cretaceous: Carbonatite And Syenite Centers
Tertiary (And Mesozoic)	Volcanic	Volcanic rocks and associated sediments	Tertiary: Volcanic Rocks And Associated Sediments
Precambrian	Metamorphic	Undifferentiated gneisses and in the north granulite facies rocks	Undifferentiated Gneisses And, In The North, Granulites Facies Rocks (Basement Complex)

The stream power index represents the erosive capacity of streams in the region and the topographic wetness index represents areas where flow accumulates and it's the function of upslope area. It also shows the moisture content of the soil at a given location thus high positive topographic wetness indexes indicate zones that tend to saturate first making the slope susceptible to failure and lastly the ruggedness index represents the surface erosive capacity by expressing the elevation differences between contiguous cells of a DEM. All these causative factors are further investigated with the susceptibility modeling in the proceeding sections.

Rainfall

Bududa experiences two wet seasons throughout the year and the dates of several landslide occurrences happen during these seasons of prolonged rainfall intensities with reference to the inventory database that is in August to November and occasionally in March and April. Rainfall has a direct influence over the entire hydrological systems in the region. It contributes to the weathering of the rock strata, increases pore water pressure through the increment in the infiltration of rainwater or the rise of the groundwater surfaces. Fig 4-6(a) displays the Average monthly rainfall totals in mm for Bududa from 2010 to 2020 plus the maximum 24hr rainfall

received during these years in 4-6(b). The overall evaluation indicated Bududa receives higher amounts of rainfall annually in comparison to most districts in Uganda as depicted in the Ugandan map in fig 4-6(c). The inventory recorded August as the month during which majority of landslides occur thus rainfall is identified as the main triggering factor of landslides in the region

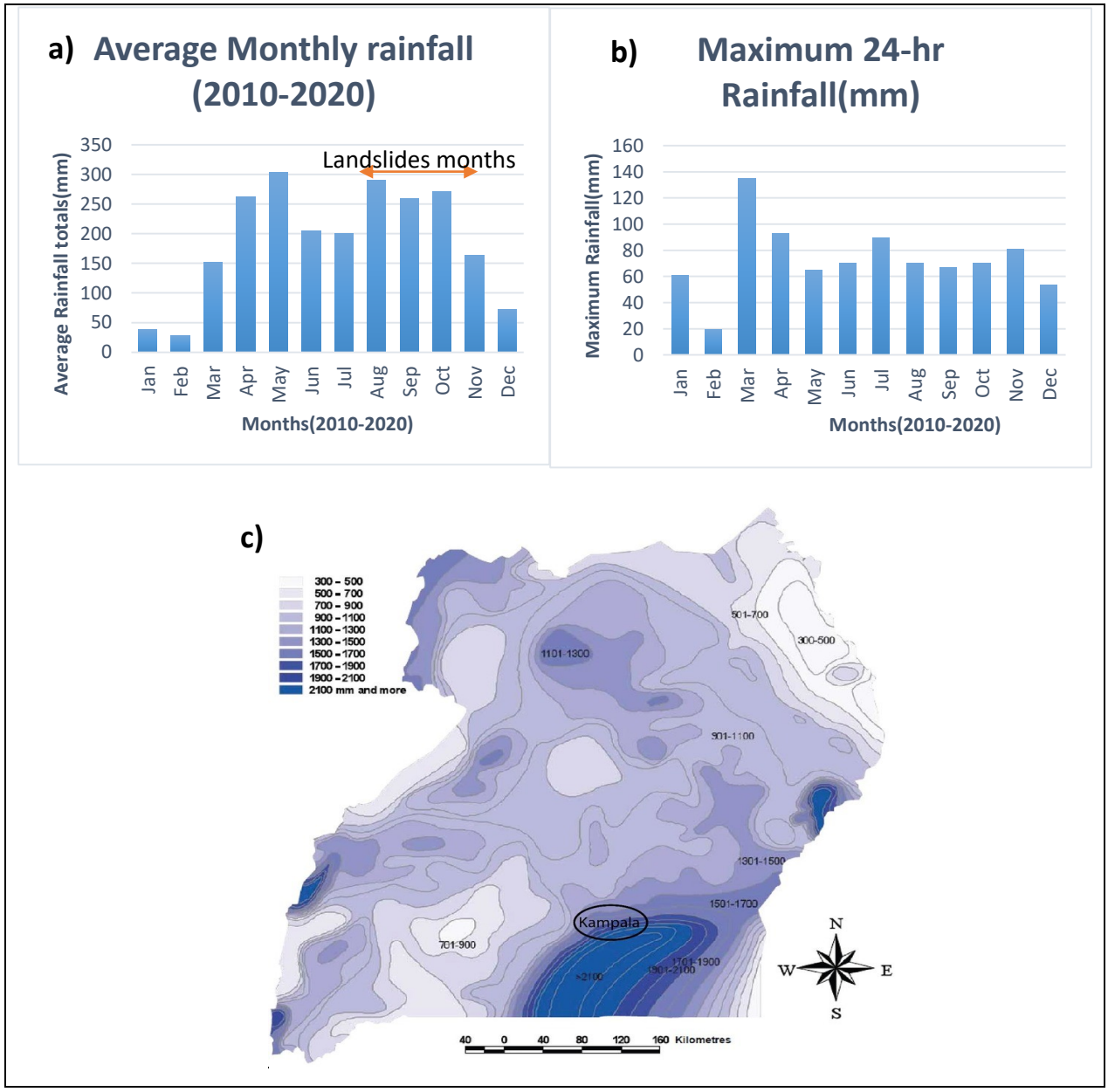
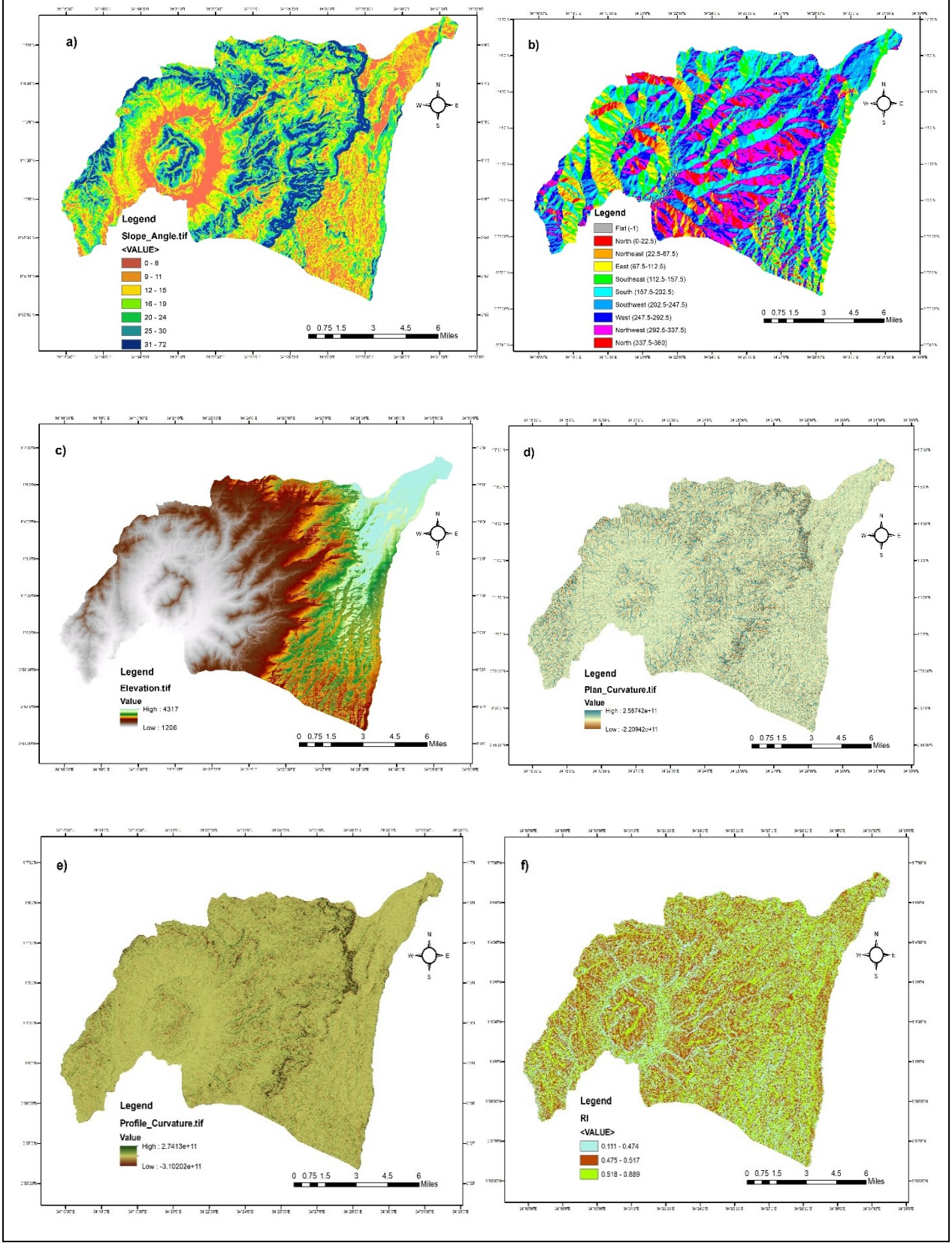
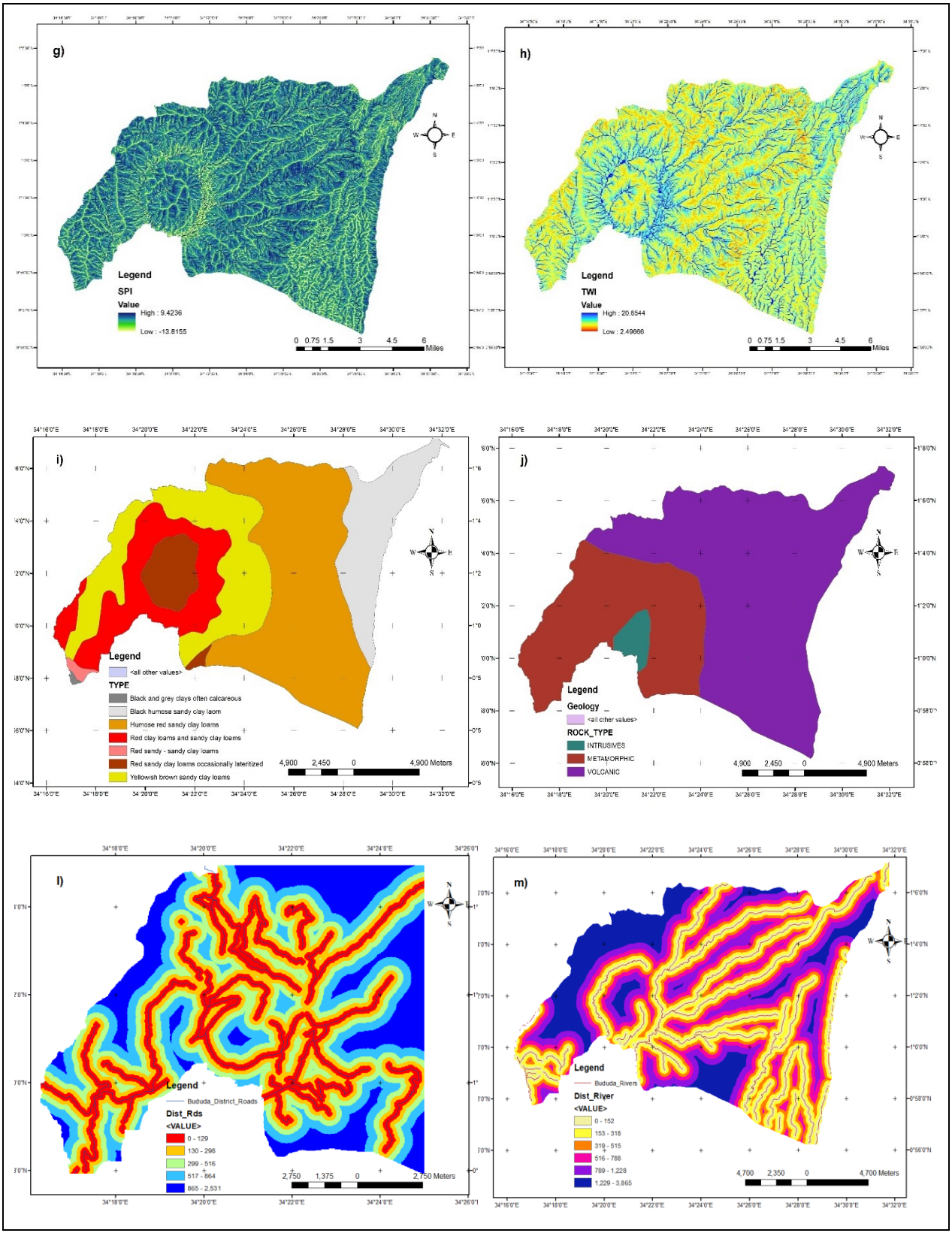


Figure 4-6: a) Average Monthly Rainfall Totals for Bududa, b) Maximum 24-hr Rainfall and c) Uganda annual rainfall map showing Bududa's annual Rainfall in comparison to other regions Source: (Habonimana, 2014.)


a)
Legend
Slope_Angle.tif
<VALUE>
0 - 8
9 - 11
12 - 15
16 - 19
20 - 24
25 - 30
31 - 72
0 0.75 1.5 3 4.5 6
Miles
b)
Legend
Flat (-1)
North (0-22.5)
Northeast (22.5-67.5)
East (67.5-112.5)
Southeast (112.5-157.5)
South (157.5-202.5)
Southwest (202.5-247.5)
West (247.5-292.5)
Northwest (292.5-337.5)
North (337.5-360)
0 0.75 1.5 3 4.5 6
Miles
c)
Legend
Elevation.tif
Value
High : 4317
Low : 1206
0 0.75 1.5 3 4.5 6
Miles
d)
Legend
Plan_Curvature.tif
Value
High : 2.58742e+11
Low : -2.20942e+11
0 0.75 1.5 3 4.5 6
Miles
e)
Legend
Profile_Curvature.tif
Value
High : 2.7413e+11
Low : -3.10202e+11
0 0.75 1.5 3 4.5 6
Miles
f)
Legend
RI
<VALUE>
0.111 - 0.474
0.475 - 0.517
0.518 - 0.889
0 0.75 1.5 3 4.5 6
Miles


g)
Legend
SPI
Value
High : 9.4236
Low : -13.8155
0 0.75 1.5 3 4.5 6
Miles
h)
Legend
TWI
Value
High : 20.6544
Low : 2.49666
0 0.75 1.5 3 4.5 6
Miles
i)
Legend
<all other values>
TYPE
Black and grey clays often calcareous
Black humose sandy clay loam
Humose red sandy clay loams
Red clay loams and sandy clay loams
Red sandy - sandy clay loams
Red sandy clay loams occasionally lateritized
Yellowish brown sandy clay loams
4,900 2,450 0 4,900 Meters
j)
Legend
Geology
<all other values>
ROCK_TYPE
INTRUSIVES
METAMORPHIC
VOLCANIC
4,900 2,450 0 4,900 Meters
l)
Legend
Bududa_District_Roads
Dist_Rds
<VALUE>
0 - 129
130 - 298
299 - 516
517 - 864
865 - 2,531
2,750 1,375 0 2,750 Meters
m)
Legend
Bududa_Rivers
Dist_River
<VALUE>
0 - 152
153 - 318
319 - 515
516 - 788
789 - 1,228
1,229 - 3,865
4,700 2,350 0 4,700 Meters

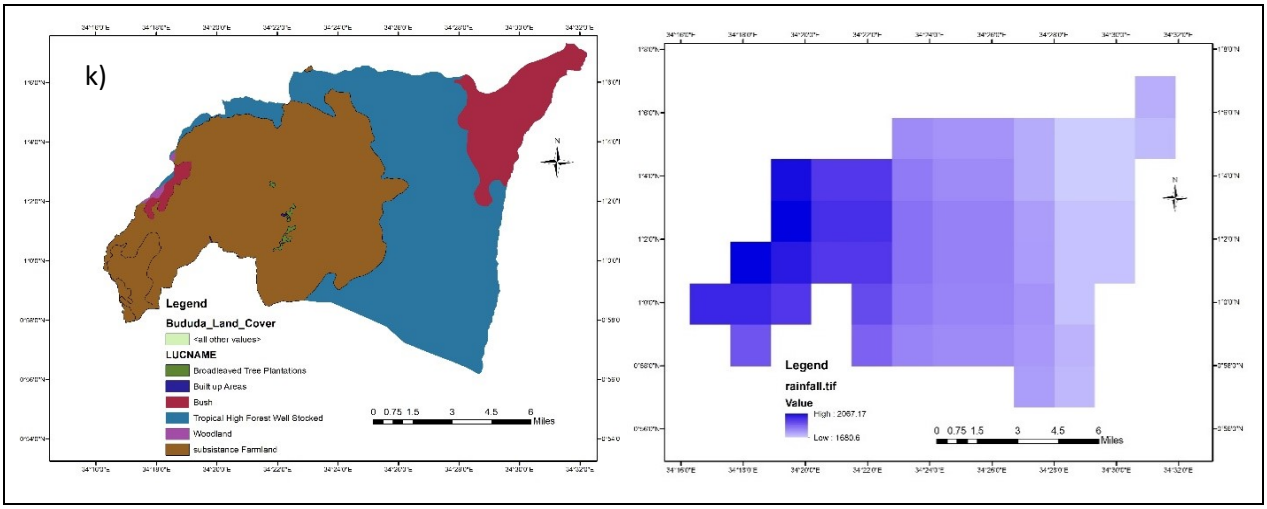


Figure 4-7: Landslide Triggering factors; a) slope angle, b) slope direction, c) elevation, d) plan curvature, e) profile curvature, f) RI, g) SPI, h)TWI, i) Distance to rivers, j) Distance to roads, l) geology, m) soil, K) land cover, rainfall: Data Source; UNMA)

After all the datasets for the 13 triggering factors have been obtained, the weight of evidence approach and support vector machines algorithms are applied on the data to establish a correlation between landslides location and the spatial distribution of the causative factors

4.3 Weight of Evidence Modeling

Using the unique value report and the rastcat program, the prior probabilities of each class of the factor maps was computed. The formulas of the model as explained in the methodology were applied on the data to compute the various weights (positive and negative weights) and the contrast values as displayed in table (4-2). Total weights of each factor map were computed and analyzed to develop the landslide susceptibility map in fig (4-9)

Table 4-2: Spatial relationships between Landslide locations and causative factors

Data Layers	Class Interval	Class (%)	Landslide (%)	W+	W-	Ww	Cw
Slope	0-9	17.4	4.5	-1.37	0.15	-1.22	-1.52
	10_15	24.2	8.3	-1.08	0.19	-0.89	-1.27
	16-21	22.0	18.2	-0.19	0.05	-0.14	-0.24
	22-28	17.5	26.4	0.42	-0.11	0.30	0.53
	29-35	11.9	24.4	0.72	-0.15	0.57	0.87
	36-46	5.7	15.2	0.99	-0.11	0.88	1.10
	>47	1.2	3.1	0.95	-0.02	0.93	0.97

Aspect	North(Flat)	4.6	3.5	-0.26	0.01	-0.25	-0.27
	North East	4.9	3.0	-0.50	0.02	-0.48	-0.52
	East	9.3	14.3	0.43	-0.06	0.37	0.48
	South East	13.1	18.8	0.36	-0.07	0.29	0.43
	South	13.6	14.7	0.08	-0.01	0.07	0.09
	South West	14.9	10.7	-0.33	0.05	-0.28	-0.38
	West	15.6	12.0	-0.26	0.04	-0.22	-0.30
	North West	13.7	14.3	0.04	-0.01	0.03	0.04
	North	10.2	8.7	-0.17	0.02	-0.15	-0.18
	Plan Curvature						
Plan Curvature	1	0.0	0.0	0.00	0.00	0.00	0.00
	2	0.4	0.8	0.71	0.00	0.71	0.72
	3	3.2	4.9	0.42	-0.02	0.40	0.44
	4	10.6	12.3	0.15	-0.02	0.13	0.17
	5	36.6	30.2	-0.19	0.10	-0.10	-0.29
	6	35.0	31.3	-0.11	0.06	-0.06	-0.17
	7	13.5	19.7	0.38	-0.08	0.31	0.46
	8	0.6	0.7	0.22	0.00	0.22	0.22
	9	0.0	0.0	0.26	0.00	0.26	0.26
Elevation	1205-1605	27.3	19.9	-0.32	0.10	-0.22	-0.42
	1605-1962	20.8	77.8	1.34	-1.28	0.07	2.62
	1962-2379	11.2	2.3	-1.60	0.10	-1.50	-1.70
	2379-2779	18.4	0.0	0.00	0.21	0.21	-0.21
	2779-3223	11.8	0.0	0.00	0.13	0.13	-0.13
	3223-3729	7.4	0.0	0.00	0.08	0.08	-0.08
	>3729	3.1	0.0	0.00	0.03	0.03	-0.03
Profile Curvature	1	0.0	0.0	0.32	0.00	0.32	0.32
	2	0.2	0.5	1.10	0.00	1.09	1.10
	3	1.0	2.3	0.84	-0.01	0.83	0.85
	4	5.2	7.5	0.36	-0.02	0.34	0.38
	5	25.6	26.1	0.02	-0.01	0.01	0.03
	6	49.7	40.9	-0.20	0.16	-0.03	-0.36
	7	17.7	21.6	0.20	-0.05	0.15	0.25
	8	0.5	1.0	0.60	0.00	0.60	0.61
	9	0.1	0.1	0.86	0.00	0.86	0.86
RI	0.11-0.45	6.8	4.7	-0.38	0.02	-0.36	-0.41
	0.45-0.48	23.4	21.6	-0.08	0.02	-0.06	-0.10
	0.48-0.50	38.1	45.9	0.19	-0.14	0.05	0.32

	0.50-0.54	24.6	23.1	-0.07	0.02	-0.05	-0.09
	0.54-0.89	7.1	4.9	-0.39	0.02	-0.36	-0.41
SPI	-7	6.6	1.5	-1.50	0.05	-1.44	-1.55
	-0.6	13.6	16.3	0.18	-0.03	0.15	0.21
	0.345	30.8	18.7	-0.50	0.16	-0.34	-0.66
	1.34	40.9	55.4	0.31	-0.28	0.02	0.59
	9.42	8.1	8.1	0.00	0.00	0.00	0.00
TWI	2.49-5.05	30.8	42.2	0.07	-0.18	-0.10	0.25
	5.05-5.62	42.2	18.5	-0.83	0.35	-0.48	-1.18
	5.62-6.25	19.0	15.0	-0.24	0.05	-0.19	-0.29
	6.25-7.25	6.0	12.9	0.78	-0.08	0.71	0.86
	7.25-20.65	2.0	11.4	1.76	-0.10	1.66	1.86
Land use	Bush land	3.0	0.0	0.00	0.03	0.03	-0.03
	Forest Plantation	0.2	0.1	-1.05	0.00	-1.05	-1.05
	Grassland	6.8	0.0	0.00	0.07	0.07	-0.07
	Subsistence farming	38.1	58.4	0.43	-0.40	0.03	0.83
	Tropical Forest (Well Stacked)	7.8	9.3	0.18	-0.02	0.16	0.20
	Tropical Forest (encroached)	19.1	32.3	0.53	-0.18	0.35	0.71
	Woodland	24.9	0.0	0.00	0.29	0.29	-0.29
Distance to Roads	0-131	37.2	3.4	-2.40	0.44	-1.96	-2.84
	131-304	27.4	7.8	-1.27	0.24	-1.03	-1.51
	304-517	18.6	18.3	-0.02	0.00	-0.01	-0.02
	517-864	11.8	32.0	1.03	-0.26	0.76	1.29
	864-2531	5.0	38.4	2.16	-0.44	1.72	2.60
Distance to Rivers	0-180	39.9	34.4	-0.15	0.09	-0.06	-0.24
	180-390	30.1	39.7	0.28	-0.15	0.13	0.43
	390-655	18.5	10.3	-0.59	0.10	-0.50	-0.69
	655-1093	8.0	9.3	0.14	-0.01	0.13	0.16
	1093-3865	3.5	6.3	0.61	-0.03	0.58	0.64
Soil	RS-SCL	0.5	0.0	0.00	0.00	0.00	0.00
	YBSCL	21.5	87.8	1.43	-1.87	-0.43	3.30
	BHSCL	14.2	0.0	0.00	0.15	0.15	-0.15
	HRSCL	41.5	1.8	-3.15	0.52	-2.62	-3.67

	RSCL(lateritized)	5.9	0.0	0.00	0.06	0.06	-0.06
	RCL&SCL	16.3	10.4	-0.45	0.07	-0.38	-0.52
	BGCL(Calcareous)	0.1	0.0	0.00	0.00	0.00	0.00
Geology	Intrusive	2.6	0.0	0.00	0.03	0.03	-0.03
	Volcanic	64.5	57.0	-0.13	0.19	0.07	-0.32
	Metamorphic	32.8	43.0	0.27	-0.17	0.11	0.44

The weight formulas are applied on both the class pixels and landslide pixels to obtain values for the NPix1, 2, 3, 4 respectively. The positive and Negative weights for each class are computed and are used to compute the total weights for each factor map. The contrast factor for each class are also derived using the formulas discussed in the methodology

According to the analysis, 5 factors had a positive overall weight Σ (Ww) which was obtained by adding all weights Ww in table (4-2). A positive Ww indicates a direct relationship to the occurrence of landslides and these factors were 1) Slope Angle with 0.43, 2) Plan Curvature with 1.88, 3) Distance to rivers with 0.28 4), Profile Curvature with 4.16 and the Topographic Wetness Index with 1.59. These weights were obtained using the WoFE equation in (3.5). The factors with extreme negative weights were Elevation with -1.22., Soil with -3.21, and Stream Power Index with -1.61. Negative W^+ values indicate the significance of the absence of the factor to the occurrence of landslides. Total weights with extreme values were used in susceptibility mapping however weights around zero were ignored as evidence suggested no or limited influence on the occurrence of landslides. Table (4-3) displays the classes of each factors with a positive association to landslide occurrence.

Table 4-3: Factor classes with a positive spatial association to the occurrence of landslides

No	**Data layer**	**C^+**	***Cw***
1	Slope	$>22^0$	*0.53-0.97*
2	Aspect	East and South-East	*0.43-0.48*
3	Elevation	1605-1962m	*2.62*
4	Plan Curvature	2, 3, 7,	*0.44-0.72*
5	Land use	Subsistence farming, Tropical Forest (encroached)	*0.7-0.83*
6	Geology	Metamorphic	*0.44*

7	Soil	YBSCL	*3.3*
8	Distance to Roads	>517m	*1.29-2.6*
9	Distance to Streams	180-390m, 1093-3865m	*0.43-0.64*
10	Ruggedness Index	0.48-0.5	*0.32*
11	Stream Power Index	-0.6, 1.34	*0.21-0.59*
12	Topographic Wetness	6.25-20.65	*0.86-1.86*
13	Profile Curvature	All values except 5 & 6	*0.25-1.10*

Slope Angles greater than 22^0 showed a positive spatial correlations with landslide locations, giving a contrast value of 0.53 to 0.97. When the parallel component of the Gravity force by which the land mass slides down increases, the slope angle increases, thus overcoming resistance to downward movement. This resistance is referred to as friction, and it is determined by the soil slope's surfaces, as well as the perpendicular component of gravity. The land mass or soil debris will slide down the slope when the parallel component of the gravity force exceeds the perpendicular component. In other words, the angle of repose, which is the greatest angle at which an object will remain at rest has been surpassed. Therefore, the angle of repose in Bududa according to the weight of evidence model is approximately greater or equal to 22^0 thus regions with slopes greater than the angle of repose are highly susceptible to landslides

The Altitude of 1605-1962m was also associated to landslide occurrence with 78% of the landslides occurring in this elevation range. The spatial association for this range is 2.62 as displayed in table (4-3) which is an extremely high value. The highest altitudes in the region are around the Elgon national park and are densely covered in thick forests thus no landslides have been reported or identified in these regions. The model allocates a negative spatial relationship of -0.03 to -0.13 which is close to the zero mark indicating no evidence present to support high elevations to contribute to landslide occurrences in the region. Therefore, Elevation obtained an overall negative weight of -1.22.

Class 4 of the plan curvature occupied 31% of the total landslides in the entire map, obtaining a strong and reliable weight value of 1.88, the second highest in all factor weights after profile curvature with 4.16. This indicates that curvature has the highest correlation to causing landslides

in the region. This evidence validates a study by (Kitutu et al., 2011) where farmers associated steep slopes, rising water tables and concavities as triggering factors of landslides.

The Subsistence farmland class in land use obtained a 58% of total landslides, higher than other forms of land use in the area such as infrastructure development, human settlement, and forests among several anthropogenic activities. It attained a contrast value of 0.83 which implies a positive spatial association between regions covered in subsistence farming and landslide occurrences. Other spatial relationships of factors and landslide occurrences are summarized in fig (4-8) below clearly indicating what classes have no influence on landslides and what classes have a direct influence on the occurrences of landslides.

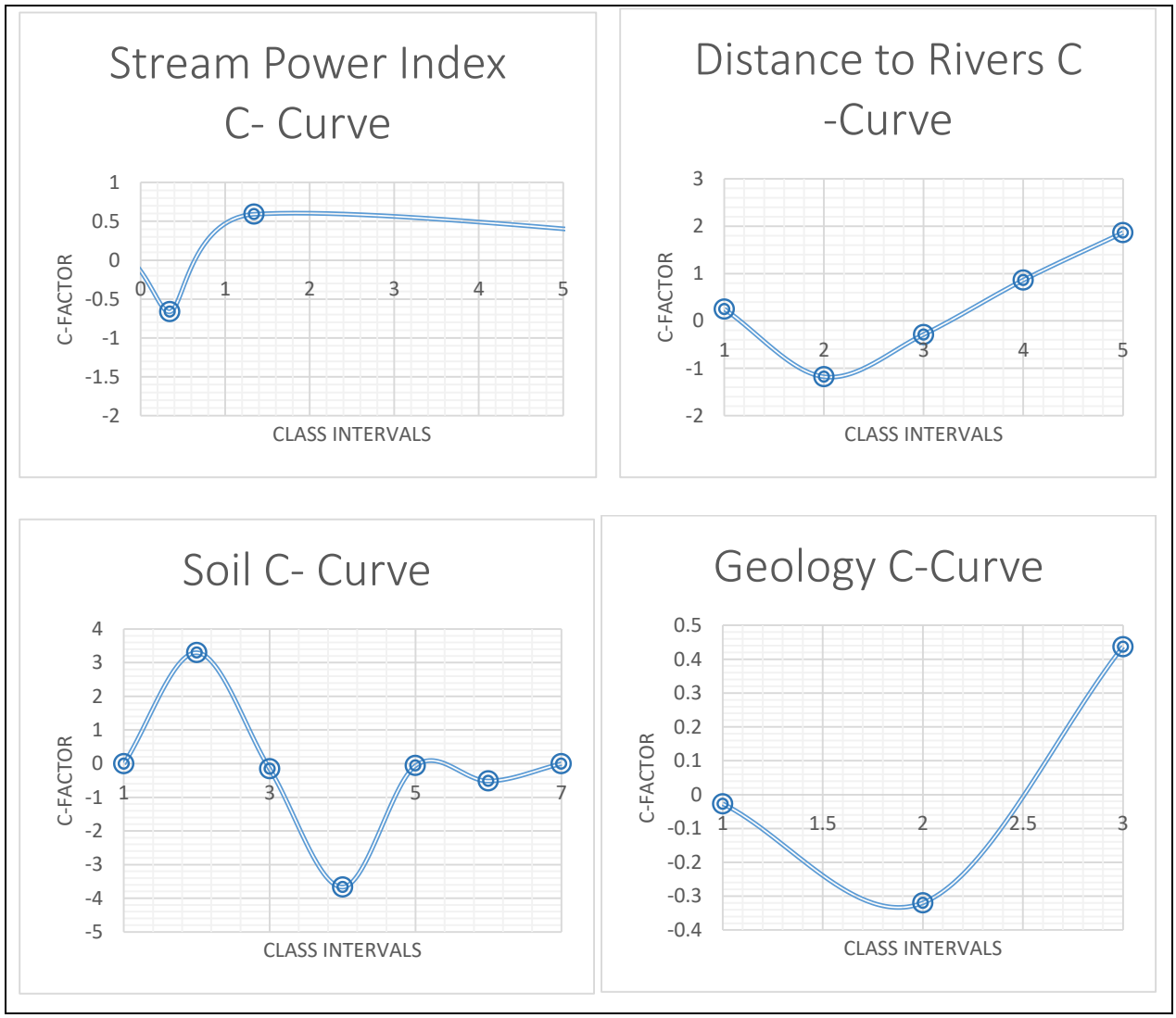


Figure 4-8: Graphs illustrating the spatial relationship between the soil, Geology, SPI and Distance to rivers factors and landslide occurrences

4.3.1 Landslide Susceptibity Mapping using the WoFE Model

After the spatial relationships between the 13 factors and the landslide locations have been investigated , the computed weights are used to develop a susceptibity map which is classified into 3 classes; Low, Moderate and High susceptible zones as shown in Fig (4-9) below

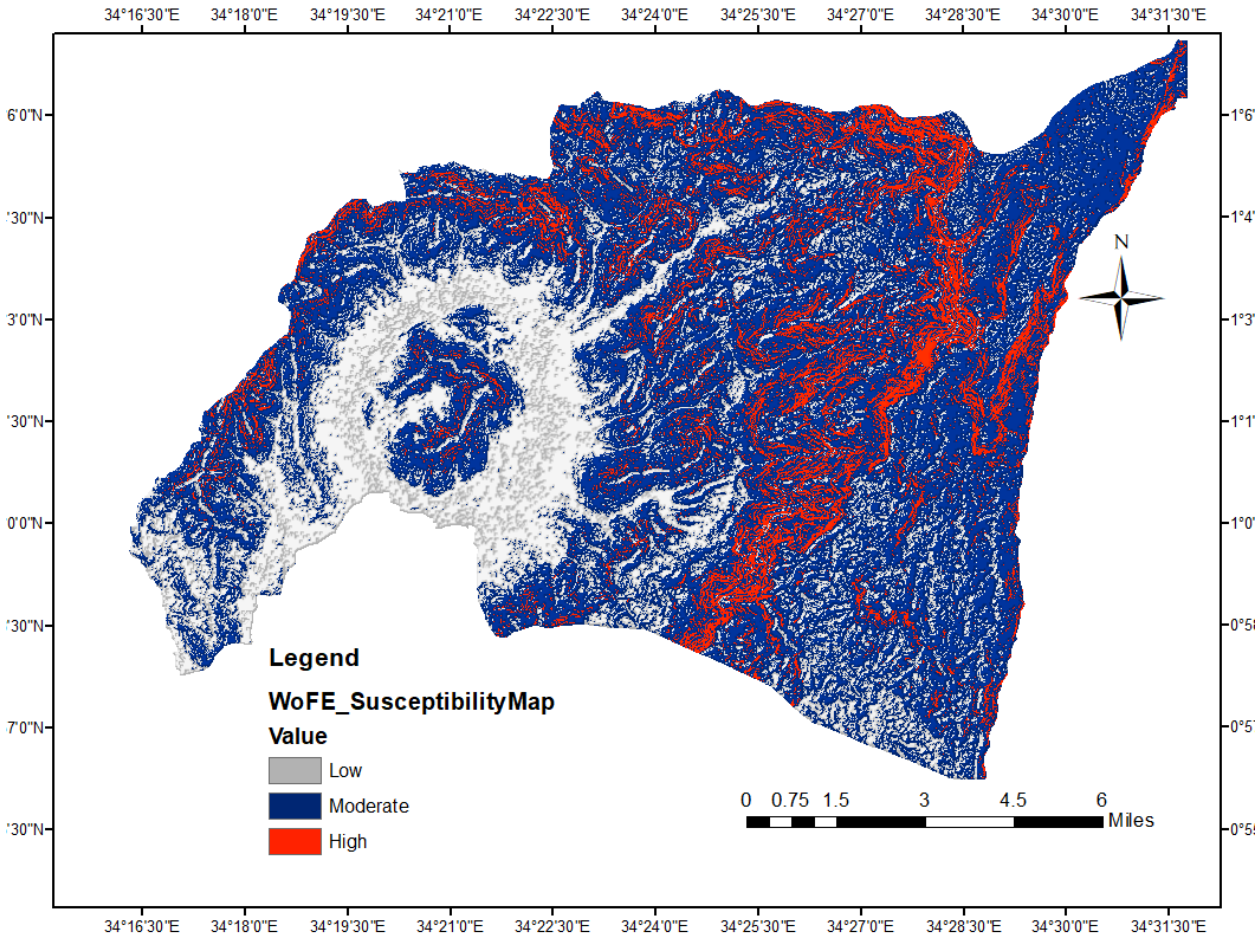


Figure 4-9: Landslide Susceptibility Map using the Weight of Evidence approach

Using the landslide true spatial locations fitted on the susceptibility map, the ROC Curve in fig (23) giving the true positive rates to the false positive rates was generated and used in the evaluation of the Model's performance.

4.3.2 AUC/ROC Curves

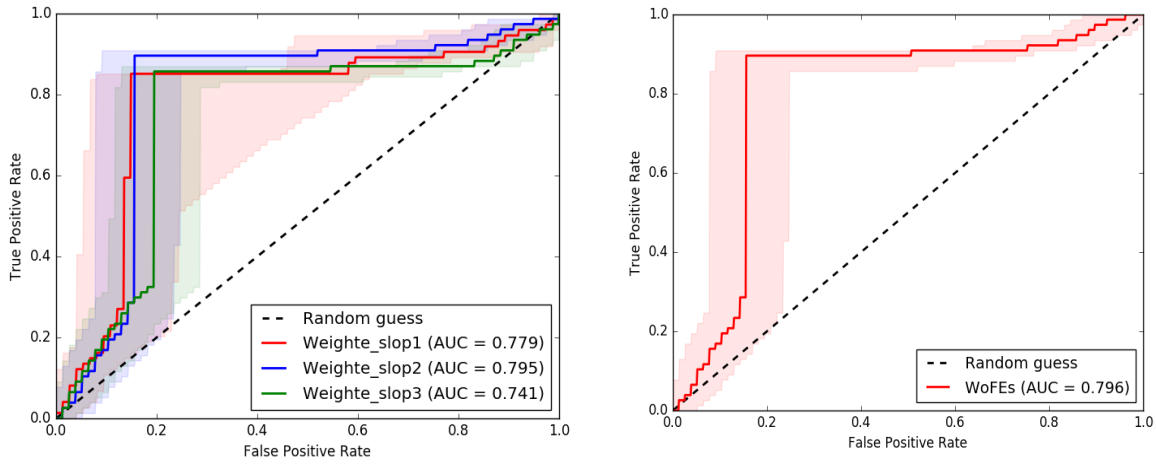


Figure 4-10: ROC Curves for the WoFE Model

The receiver operating curves gave an accuracy of 74-80% for area under curve which indicates a good performance by the weight of evidence model in landslide susceptibility mapping

4.3.3 Field Validation

Table 4-4: Landslide locations identified during fieldwork for validation purposes

No.	Latitude	Longitude	Elevation (m)
1	1°0’36.03”N	34°19’52.53”E	1306.82
2	1°4’6.03”N	34°20’17.11”E	1483.52
3	1°3’54.84”N	34°20’32.77”E	1572.64
4	1°3’57.48”N	34°20’33.33”E	1599.83
5	1°3’58.91”N	34°20’35.46”E	1627.01
6	1°4’3.48”N	34°20’30.57”E	1612.20
7	1°3’48.88”N	34°20’32.29”E	1511.41
8	1°3’40.96”N	34°22’41.84”E	1624.06
9	1°2’59.08”N	34°23’12.24”E	1529.03
10	1°3’3.97”N	34°23’14.56”E	1531.00

The coordinates were overlaid on top of the susceptibility map as shown in fig (4-11) to further validate the generated susceptibility classes in additional to sensitivity analysis using the ROC Curve

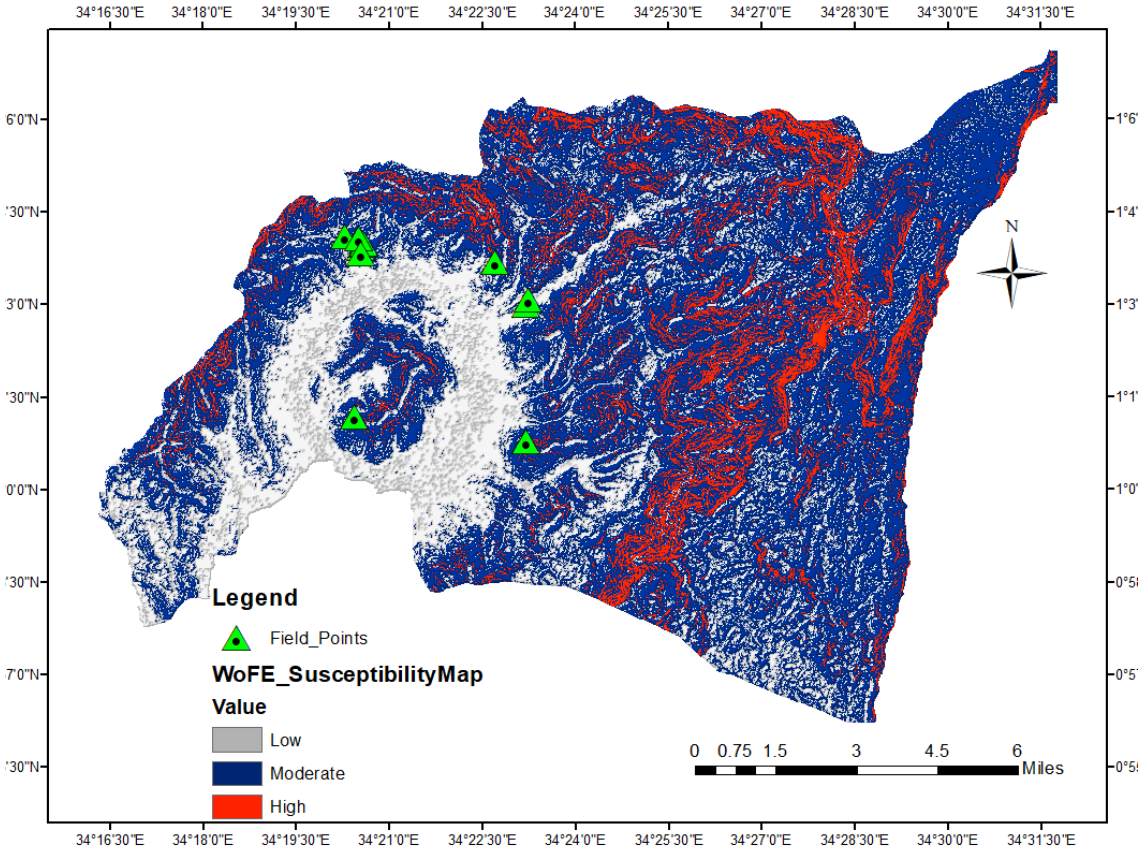


Figure 4-11: Field Validation Map (WoFE)

Figure 4-12: Landslide hotspots identified during the field visit

4.4 Machine Learning Modeling

Training points were randomly created with a 50:50 coverage on both landslide spatial locations and locations with no recorded landslides. 270 training points were taken with 135 points for landslide locations and the other 135 points for absence of landslides as shown in fig (4-13).

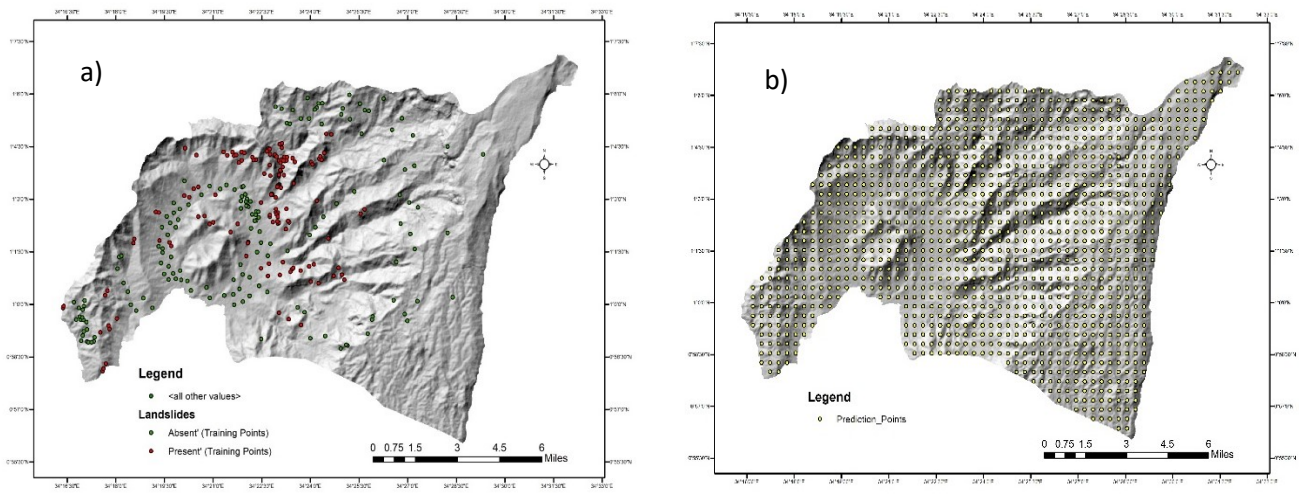


Figure 4-13; a) SVM training Points and b) Prediction Points

Figure 4-14: 3D View of the Bududa Terrain

Random points as shown in fig 4-13(b) were also extracted for fitting the model after it had trained and learnt on the training dataset. These points were to further validate the accuracy of the machine learning model. The training dataset was then subjected to scaling, feature selection and Principal component analysis.

4.4.1 Feature Importance

The feature importance task is to eliminate variables that have less or zero correlation to the occurrence of the landslides in the region. According to table (4-5), only 4 features showed a

strong correlation to the occurrences thus were chosen for modeling. The 4 principal component analysis was turned on to reduce on the dimensionality of the entire dataset while maintaining all the key information necessary for the training. The 4 selected features for the Support Vector Machines were 1) Slope Angle, 2) Profile Curvature, 3) Stream Power Index and 4) Finally, the elevation.

Table 4-5; Feature Importance

	Slope Angle	Aspect	Elevation	Plan Curvature	Profile Curvature	RI	SPI	TWI	SLIDES
Slope Angle	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Slope Direction	-0.07	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Elevation	0.29	0.08	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Plan Curvature	0.04	-0.01	0.04	1.00	0.00	0.00	0.00	0.00	0.00
Profile Curvature	-0.02	0.02	-0.10	-0.55	1.00	0.00	0.00	0.00	0.00
RI	0.08	-0.03	0.15	-0.06	-0.39	1.00	0.00	0.00	0.00
SPI	0.21	0.01	-0.06	-0.52	0.30	-0.04	1.00	0.00	0.00
TWI	-0.44	0.10	-0.17	-0.44	0.25	-0.03	0.60	1.00	0.00
Landslides	0.40	-0.12	-0.19	-0.10	0.15	-0.06	0.19	-0.12	1.00

Using regression analysis, the P values of the coefficients of the first 6 variables were computed, upon which 4 were selected for fitting the model, these P values are later used as indexes for susceptibility mapping as they indicate the key factors that influence landslide occurrences. The regression statistics used for computing are displayed in table (4-6) and the P values shown in table (4-7) with their respective upper and lower 95% values, standard error and t stats. The P values indicate the statistical significance of these variables in landslide occurrence

Table 4-6: Regression Statistics

Regression Statistics

Multiple R	0.529
R Square	0.280
Adjusted R Square	0.264
Standard Error	0.429
Observations	284

Table 4-7: P values

	Coefficients	**Standard Error**	**t Stat**	**P-value**	**Lower 95%**	**Upper 95%**
Intercept	0.746	0.229	3.265	0.001	0.296	1.197
Slope Angle	0.025	0.004	6.016	0.000	0.017	0.034
Slope Direction	0.000	0.000	-1.266	0.207	-0.001	0.000
Elevation	0.000	0.000	-5.630	0.000	0.000	0.000
Profile Curvature	0.000	0.000	2.264	0.024	0.000	0.000
SPI	0.005	0.011	0.466	0.642	-0.016	0.027
TWI	-0.002	0.022	-0.085	0.933	-0.046	0.042

4.4.2 Radial Basis Function Kernel SVM

The test data was loaded into the workspace and the RBF SVM code applied to the data to generate the results in fig (4-15). The properties of the model show the response name, categorical predictors, class names as assigned by the algorithm, number of observations, hyper parameter optimization results, kernel parameters, box constraints and other relevant properties

Model 1 =

RBF SVM

Response Name: 'Y'

Categorical Predictors: []

Class Names: [1 2]

Score Transform: 'none'

Num Observations: 152

HyperparameterOptimizationResults: [1×1 Bayesian Optimization]

Alpha: [84×1 double]

Bias: -1.4118

Kernel Parameters: [1×1 struct]

Box Constraints: [152×1 double]

Convergence Info: [1×1 struct]

Is Support Vector: [152×1 logical]

Solver: 'SMO'

The class order taken by the model to represent true for present and false for absence of landslides is;

Class Order =

1 (negative class; absent)

2 (positive; present)

CV partition

The dataset was partitioned using the 5 k fold cross validation and the values used are;

Type;	['k-fold']				
NumTestSets	5				
TrainSize	[122	121	121	122	122]
TestSize	[30	31	31	30	30]
NumObservation	152				

Feature Selection

The forward sequential feature selection was used and the results are displayed below;

Initial columns included: none

Columns that cannot be included: none

Step 1, added column 1, criterion value 0.256579

Step 2, added column 4, criterion value 0.243421

Final columns included: 1, 4

Several approaches can also be applied other than the forward sequential feature selection such as the ‘backward’ ‘Direction’ and the accuracy of the results may vary according to the method applied

Objective Function Model

The Objective function model conducts iterations while pairing two features along the kernel scales and the box constraints to choose the best outcome or the features with the best predictive accuracy. The model uses the criterion values set in sequential feature selection to conduct iterations until the most feasible points are attained. In this study, 30 iterations were conducted and the results of the process are displayed in fig (4-15) and table (4-8). Upon completion, the Objective function Model is plotted showing the observed data points, the model mean, next points and the model feasible minimum. The corresponding plot of the minimum objective functions alongside function evaluations was also plotted. The minimum kernel scale and BoxConstraint values were attained and the values displayed below the model

Table 4-8: Objective function iteration values

Iter	Eval result	Objective	Objective runtime	BestSoFar (observed)	BestSoFar (estim.)	BoxConstraint	KernelScale
1	Best	0.27632	3.322	0.27632	0.27632	26.311	7.2612
2	Accept	0.31579	0.34824	0.27632	0.27848	0.0015471	0.25194
3	Accept	0.31579	0.28819	0.27632	0.27978	34.098	0.011685
4	Accept	0.31579	0.15614	0.27632	0.27887	196.33	976.91
5	Accept	0.28947	0.74383	0.27632	0.27883	92.984	8.1849
6	Accept	0.27632	0.16098	0.27632	0.27631	13.085	6.7073
7	Accept	0.28289	0.21224	0.27632	0.27632	14.278	10.58
8	Best	0.25658	0.16963	0.25658	0.25659	22.682	4.306
9	Best	0.19737	0.16434	0.19737	0.19759	26.307	3.0619
10	Accept	0.25	0.12383	0.19737	0.23714	37.158	2.6016
11	Accept	0.26974	0.13502	0.19737	0.24082	0.90742	2.9516
12	Accept	0.21711	0.12937	0.19737	0.202	42.791	2.9674
13	Accept	0.21053	0.14051	0.19737	0.20671	14.422	3.2099
14	Accept	0.19737	0.13726	0.19737	0.20057	34.029	3.2473
15	Accept	0.20395	0.22371	0.19737	0.19735	45.11	3.2952
16	Accept	0.20395	0.14637	0.19737	0.20114	24.696	3.2466
17	Accept	0.19737	0.15804	0.19737	0.19877	27.86	3.1847
18	Accept	0.20395	0.1289	0.19737	0.20036	25.86	3.158
19	Accept	0.23684	0.13486	0.19737	0.20021	1.1458	1.3465
20	Accept	0.32895	0.13545	0.19737	0.20023	7.5301	0.77092

Iter	Eval result	Objective	Objective runtime	BestSoFar (observed)	BestSoFar (estim.)	BoxConstraint	KernelScale
21	Accept	0.20395	0.14541	0.19737	0.20006	34.543	3.4003
22	Accept	0.31579	0.10685	0.19737	0.2	0.0048902	1.2949
23	Accept	0.31579	0.10978	0.19737	0.20001	0.010958	8.2481
24	Accept	0.26974	0.10891	0.19737	0.19998	21.13	27.708
25	Accept	0.27632	0.22351	0.19737	0.19996	798.25	52.608
26	Accept	0.31579	0.1174	0.19737	0.19997	0.94016	53.27
27	Accept	0.30263	0.11985	0.19737	0.19998	742.7	126.53
28	Accept	0.31579	0.11706	0.19737	0.19998	0.065998	21.409
29	Accept	0.31579	0.13861	0.19737	0.20001	0.0010223	0.0010006
30	Accept	0.20395	0.16003	0.19737	0.20018	34.948	3.3088

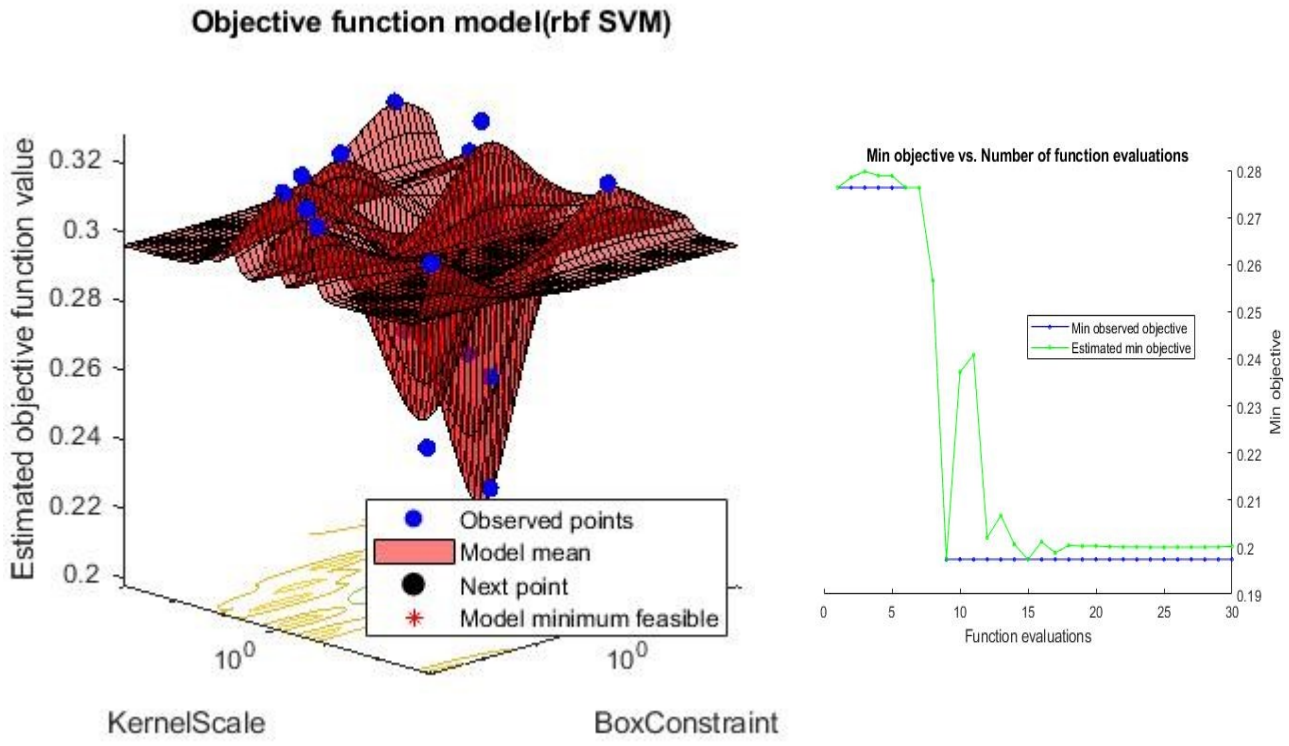


Figure 4-15: Objective function Model and the function evaluations vs. minimum objectives

Optimization completed.

MaxObjectiveEvaluations of 30 reached.

Total function evaluations: 30

Total elapsed time: 135.8085 seconds.

Total objective function evaluation time: 8.5064

Best observed feasible point:

BoxConstraint	KernelScale
26.307	3.0619

Observed objective function value = 0.19737

Estimated objective function value = 0.2027

Function evaluation time = 0.16434

Best estimated feasible point (according to models):

BoxConstraint	KernelScale
27.86	3.1847

Estimated objective function value = 0.20018

Estimated function evaluation time = 0.1772

After the training was completed, the model was tested on the other partitions created for testing and validation. Its accuracy was computed using the code line below;

[Test for accuracy = sum ((predict (Md1, X-test with best feature) == y-test))/length (y-test)*100]

Obtaining a performance accuracy of 78.9474 ~ [79%]. The scatter plot showing the predicted values and the support vectors used in the plane transformation is also displayed in fig (4-16)

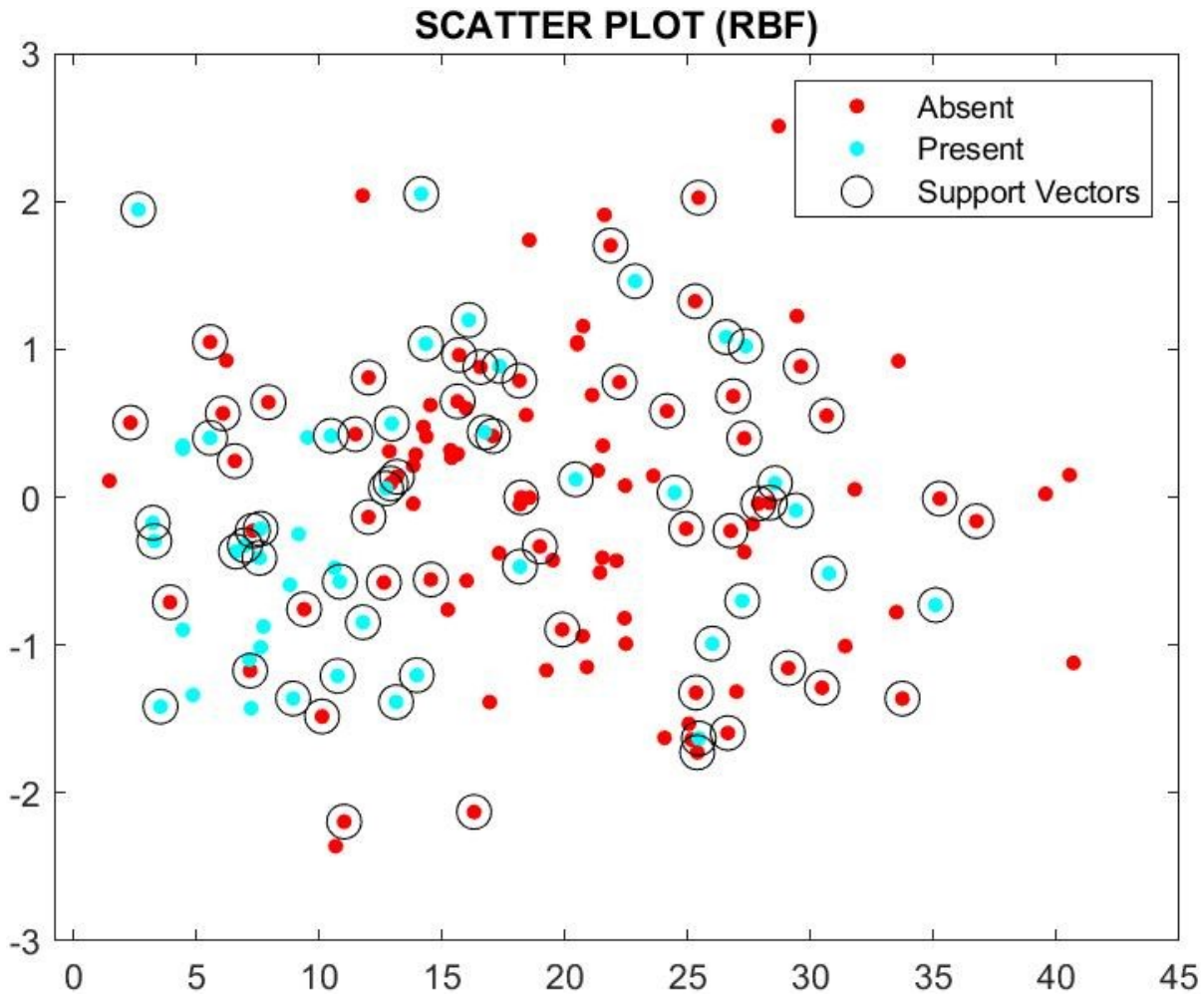


Figure 4-16: RBF Scatter plot

4.4.3 Optimizable Support Vector Machines

The Optimizable support vector machine model integrates all SVMS (Gaussian, Linear and Quadratic) and the same steps like the RBF model training were followed. 1) Data scaling, 2) feature selection and a 4 Principal component feature analysis was conducted. It was optimized using the Bayesian probability and the results displayed the best point hyper-parameters between 0.17 and 0.18 along the minimum classification errors. The parallel coordinate plots is also generated which indicates the standard deviations of correctly predicted values and incorrectly predicted values for both the presence and absence of landslides using all 8 factors. The scatter plot illustrating the results of the model is displayed in fig (4-17). Finally, the ROC curve showing the performance accuracy of 87% of the Combined SVMs is generated and shown in fig (4-21)

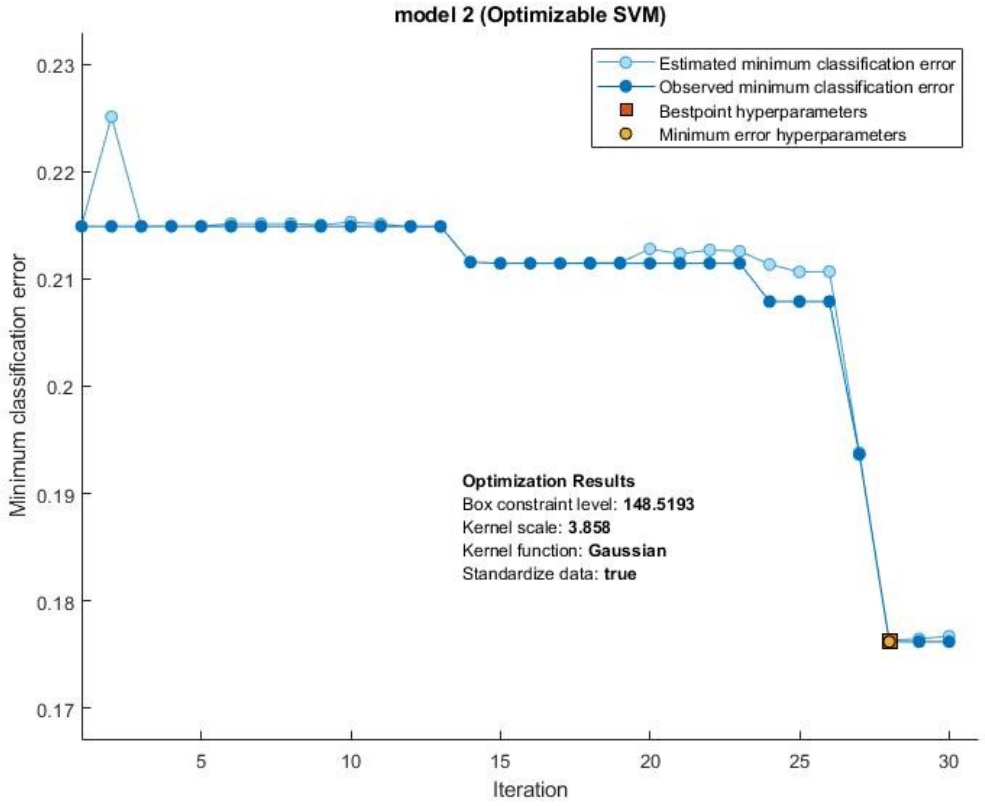


Figure 4-17: Best point hyper parameters and minimum classification errors for the OSVM Model

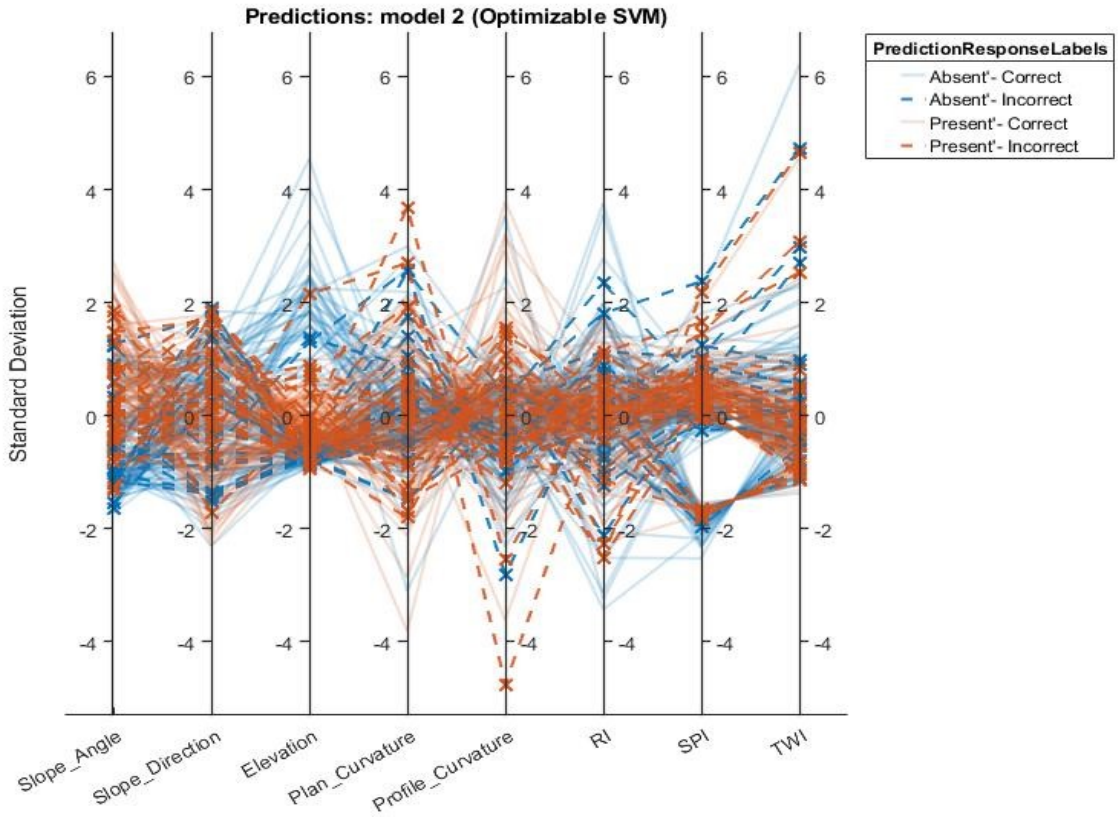


Figure 4-18: Parallel Coordinates plot for the Predictor Variables

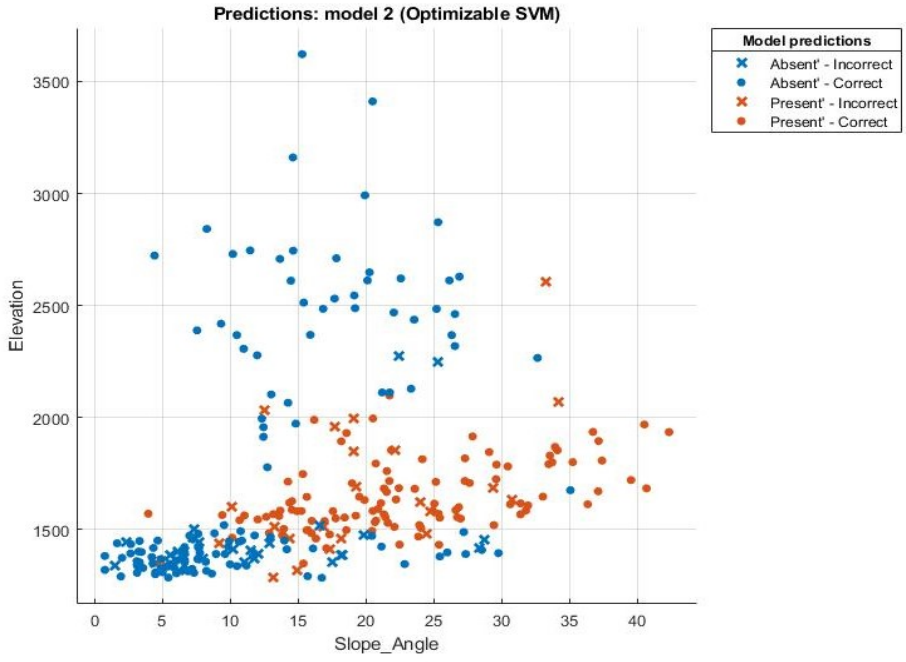


Figure 4-19: Scatter plot for the OSVM Model

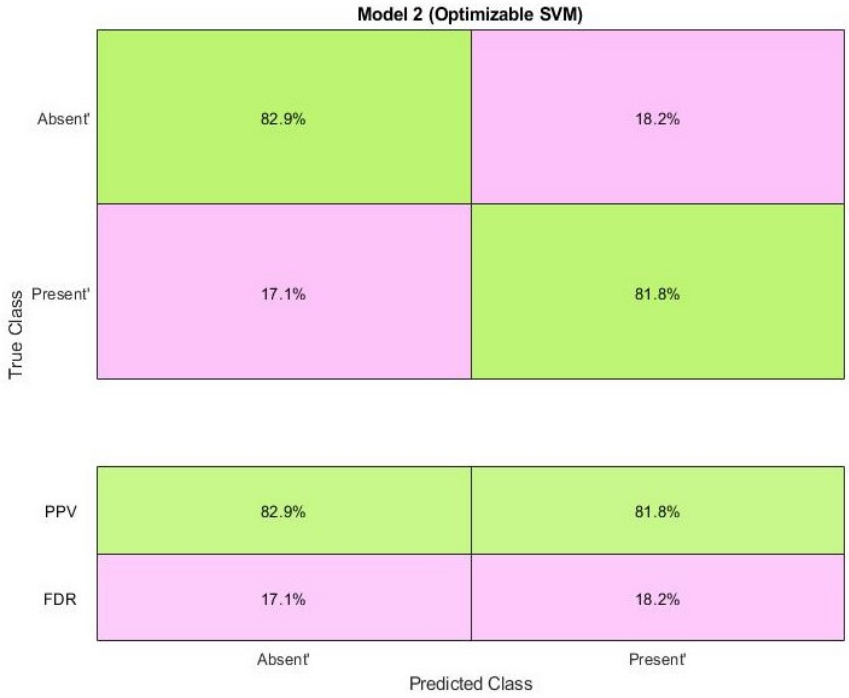


Figure 4-20: OSVM Confusion Matrix

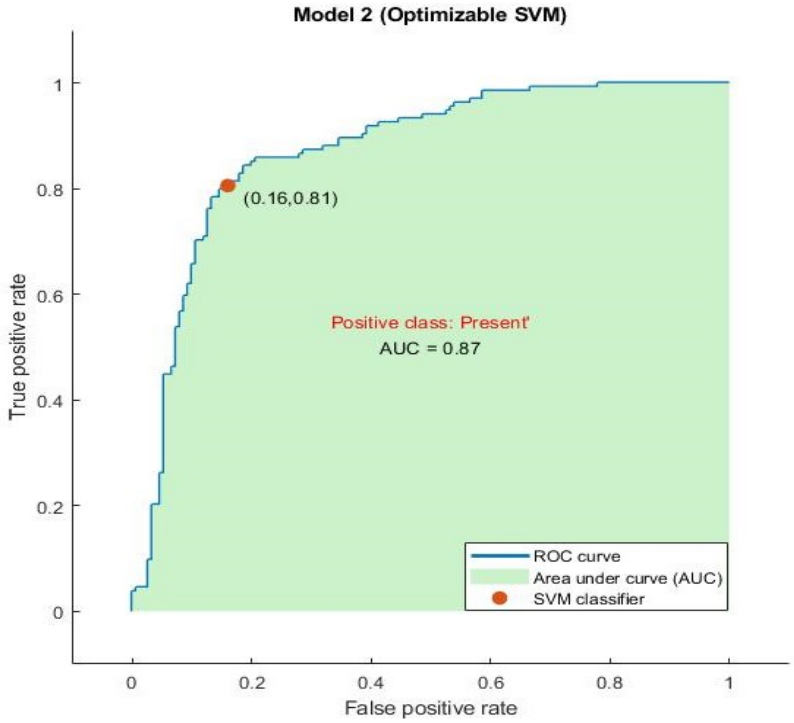


Figure 4-21: ROC Curve for the OSVM Model

The trained models are then applied to a new set of data with similar variables to further validate their performance and the values predicted are plotted in fig (4-22), the 4 features used in the modeling are further used to develop the susceptibility map in fig (4-23). The Landslide Susceptibility Map is further subdivided into classes; i.e. low, medium, high to indicate landslide susceptibility zones. When overlaid with the spatial location of landslides sites collected during the field trip, they coincide with regions of high susceptibility thus validating the developed susceptible regions of the map

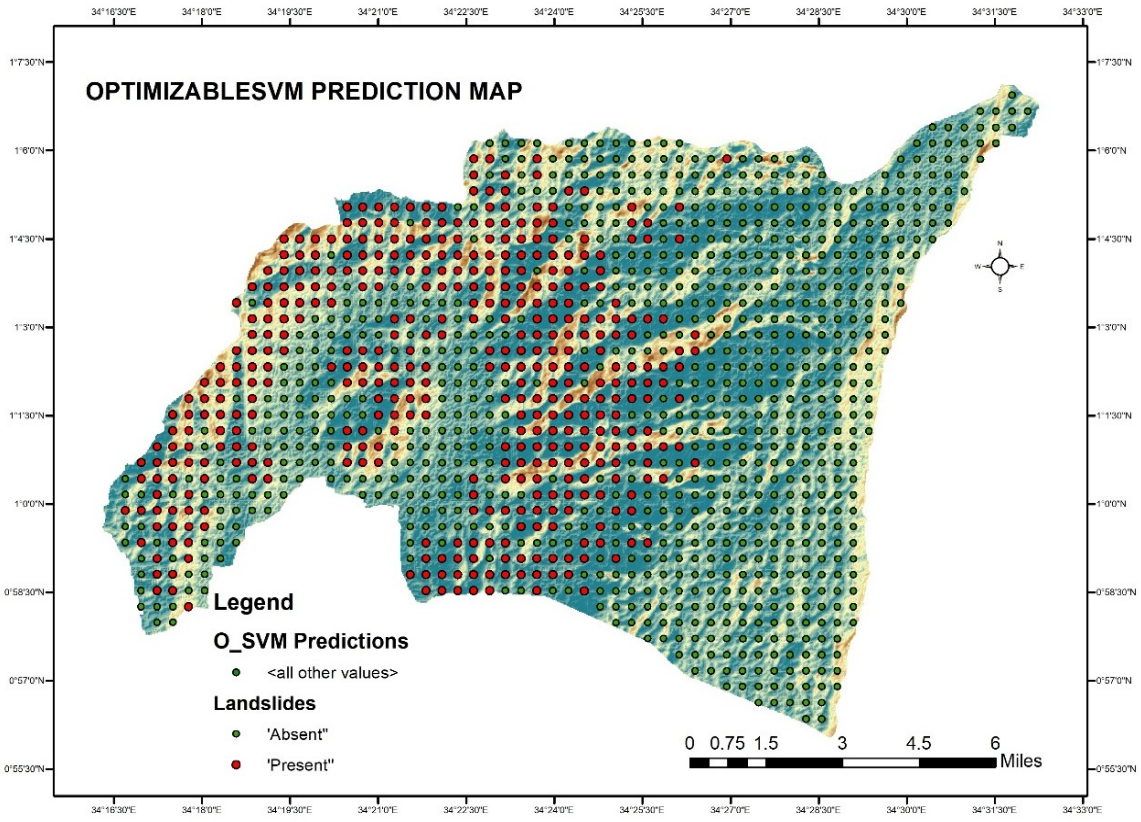


Figure 4-22: OSVM Landslide Prediction Map

Landslide Susceptibility Map using the SVM

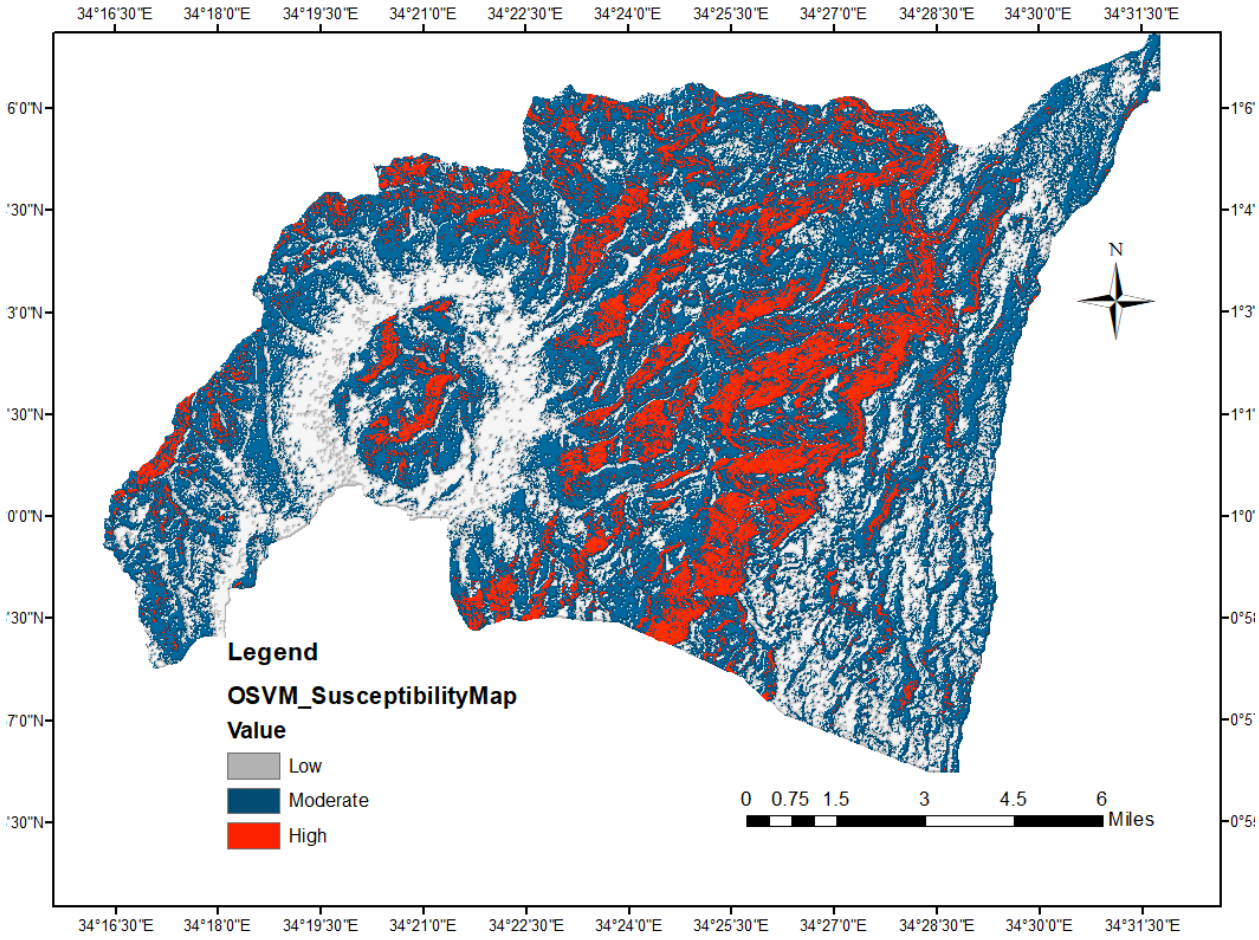


Figure 4-23: Landslide Susceptibility Map using SVM Models

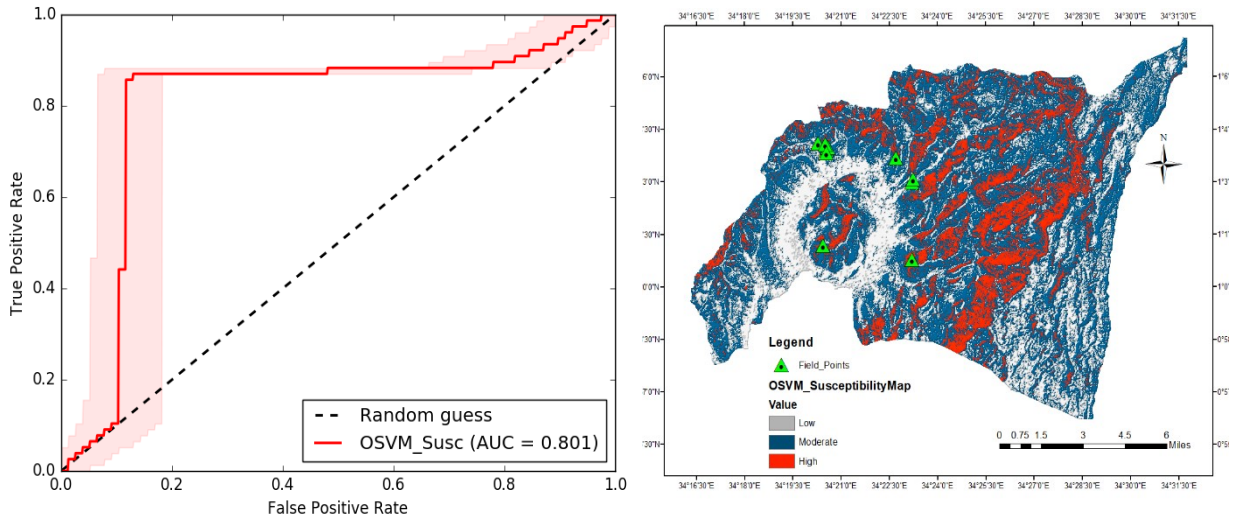


Figure 4-24: ROC Curve for the SVM susceptibility map

The ROC Curve of the developed susceptibility map using support vector machine models was generated and it showed a good performance of 80% as shown in fig (4-24). Generally, the machine learning algorithms obtained the accuracy of 79%, 87% and 80% in landslide susceptibility pattern recognition.

4.5 Comparative Analysis of the WoFE and SVMs

4.5.1 Performance evaluation using the ROC Curves

Table 4-9: ROC Performance Indicator

AUC Value	Test Quality
0.9-1.0	Excellent
0.8-0.9	Very Good
0.7-0.8	Good
0.6-0.7	Satisfactory
0.5-0.6	Unsatisfactory

The Optimizable Support Vector Model gave an accuracy of 87% of the area under curve which is a very good performance indicator according to the AUC quality criterion in table 4-9. The GIS based Model (Weight of Evidence) gave an accuracy of 79% which is a good performance indicator too. The confusion matrix for the SVM model showed a PPV accuracy of 81.8% which

represents presence of landslides and a PPV of 82.9% for absence. Generally, using the ROC curves/AUC the SVM machine learning models performed better in landslide susceptibility modeling than the weight of evidence.

4.5.2 Performance on causative factor investigations

The weight of evidence evinced its strength in integrating all environmental variables including Geology, Soil, and Land use unlike machine learning modeling that required raster datasets with a large variation for modeling. In the Weight of Evidence approach, the classes for each factors were extensively analyzed for conditional independence and correlations to landslide spatial locations which is crucial in landslide susceptibility assessments.

On the other hand, SVMs demonstrated efficiency in selecting appropriate input features and generating functions that can be applied on new data samples to make new predictions on landslide occurrences

Despite the variation in factor selections both models identified slope and curvature as a dominant cause to the occurrence of Landslides in the region. Other causative factors identified by the models are Stream Power Index, Topographic Wetness Index, Geology and Elevation

4.5.3 Performance evaluations using the susceptibility Maps

Table 4-10; Pixel Based Statistics of the models

Model	Susceptibility Level	P(Slides)	C(Pixels)	L(Pixels)	PL(Classes)
WoFE	Low	0.26-0.57	94575	467	0.49
	Moderate	0.57-0.71	198414	1739	0.88
	High	0.71-1.00	35133	353	1.00
OSVM	Low	0.26-0.57	95642.00	316.00	0.33
	Moderate	0.57-0.71	183223.00	1580.00	0.86
	High	0.71-1.00	49220.00	663.00	1.35

The probability of landslides occurring in the low susceptibility level is 0.26 to 0.57, moderate, 0.57 to 0.71 and high, 0.17 to 1.00 according to the classified susceptibility intervals. However, the probability of landslides occurring in the moderately susceptible regions was beyond the threshold in both models giving a range of 0.86 to 0.88. This is still lower in comparison to 1.00

to 1.35 range for the probability of landslides happening in the high susceptible zones by both models.

In summary the mapped thresholds for landslide susceptibility in the region of Bududa is approximately close to real time data thus a good overall performance for both the weight of Evidence and Machine learning Algorithms in susceptibility modeling

4.5.4 Strength and Limitations of the WoFE approach

The weight of evidence method is based on data. It has the advantage of avoiding the subjectivity of weight to choose, and it is simple to evaluate to compose multiple maps. It incorporates all evidence that can be accumulated for the study thereby establishing positive or negative associations between the factors and the disaster.

The limitation is that the factor maps to be modeled must be independent from each other, and the target layer must contain known points of interest, in our case study the landslide true locations.

4.5.5 Strength and Limitations of the SVMs approach

The support vector machines work best when there is a high dimensional space in the dataset. This is because of their kernel functions that are highly optimal in non-linearly separable datasets. The SVMs are more efficient when there is a clear line of separation between classes of variables

The limitations are the SVM algorithms will underperform when the dataset has target classes that are overlapping and also in cases where the existing features for each data point exceeds the number of training data samples. As the SVM models work by putting data points, above and below the decision boundary, there is no probabilistic inferences for the classified features which makes other approaches like Weight of Evidence more powerful in causative factor investigations

5.0 CONCLUSION AND RECOMMENDATION

5.1 Conclusion

Landslide susceptibility mapping is vital for landslide risk management and urban planning. In this study, landslide susceptibility was evaluated for Bududa district by analyzing fourteen causal factors with two models (A GIS Based model-WoFE) and one machine learning model (Support Vector Machines-RBF and an Optimizable SVM). The SVM model showed that 4 features trigger landslides and those were Slope Angle, Elevation, Profile Curvature and Stream Power Index. The WoFE model identified 5 factors and these were Slope Angle, Plan Curvature, Profile Curvature, Geology and the Distance to rivers. These factors are in additional to the contributing factor of rainfall as depicted in the inventory and further collaborated by natives in the region (Kitutu, 2010). The Confusion Matrix and the ROC curves gave an accuracy of 87% (AUC) for SVM and a 79% (AUC) for WoFE in performance. The field validation data placed 10 landslide points in the mapped susceptibility zones for both models which indicates a good overall performance for both methodologies, however more efficiency and accuracy could be attained with the integration of both approaches in landslide susceptibility modeling.

The machine learning models performed better than the GIS based approaches as they trained on the features present at the landslide sites and pinned all locations with similar conditions thus identifying all locations where landslides can probably occur. On the other hand, Weight of Evidence model conducted an extensive analysis on landslide causative factors establishing their spatial correlations to landslide occurrences

5.2 Recommendation

In additional to landslide susceptibility modeling to identify prone regions, early warning signs of landslides should also be identified and used to lessen the impacts of landslides in the region. For example, settlements on slopes with cracks and dykes and are greater than 22^0 should be avoided according to this study.

Natives in Bududa and other landslide prone regions should avert settlements in areas towered by aggregated cliffs or those within an approximate of 1000m from ridges allowing exceptions from ridges covered in thick forest. According to (Moresi et al., 2020), forest roots contribute to the mechanical reinforcement of the soil, improves the soil structure and also facilitate soil water

removal through evapotranspiration . Therefore areas with landslide cracks should be averted and rehabilitated with forestry.

Finally, more research integrating GIS and advanced machine learning approaches should be conducted to gain a deeper insight into diverse landslide risk reduction strategies in all prone regions of the country.

6.0 REFERENCE

An, K., Kim, S., Chae, T., & Park, D. (2018). Developing an accessible landslide susceptibility model using open-source resources. *Sustainability (Switzerland)*, *10*(2), 1–13. https://doi.org/10.3390/su10020293

Ballabio, C., & Sterlacchini, S. (2012). Support Vector Machines for Landslide Susceptibility Mapping: The Staffora River Basin Case Study, Italy. *Mathematical Geosciences*, *44*(1), 47–70. https://doi.org/10.1007/s11004-011-9379-9

Brenning, A. (2005). Spatial prediction models for landslide hazards: Review, comparison and evaluation. *Natural Hazards and Earth System Science*, *5*(6), 853–862. https://doi.org/10.5194/nhess-5-853-2005

Broeckx, J., Maertens, M., Isabirye, M., Vanmaercke, M., Namazzi, B., Deckers, J., Tamale, J., Jacobs, L., Thiery, W., Kervyn, M., Vranken, L., & Poesen, J. (2019). Landslide susceptibility and mobilization rates in the Mount Elgon region, Uganda. *Landslides*, *16*(3), 571–584. https://doi.org/10.1007/s10346-018-1085-y

Broeckx, J., Vanmaercke, M., Duchateau, R., & Poesen, J. (2018). A data-based landslide susceptibility map of Africa. *Earth-Science Reviews*, *185*(October 2017), 102–121. https://doi.org/10.1016/j.earscirev.2018.05.002

Bzdok, D., Krzywinski, M., & Altman, N. (2018). Machine learning: supervised methods. *Nature Methods*, *15*(1), 5–6. https://doi.org/10.1038/nmeth.4551

Canavesi, V., Segoni, S., Rosi, A., Ting, X., Nery, T., Catani, F., & Casagli, N. (2020). Different approaches to use morphometric attributes in landslide susceptibility mapping based on meso-scale spatial units: A case study in Rio de Janeiro (Brazil). *Remote Sensing*, *12*(11), 1–26. https://doi.org/10.3390/rs12111826

Costache, R., Pham, Q. B., Corodescu-Roşca, E., Cîmpianu, C., Hong, H., Thuy Linh, N. T., Fai, C. M., Ahmed, A. N., Vojtek, M., Pandhiani, S. M., Minea, G., Ciobotaru, N., Popa, M. C., Diaconu, D. C., & Pham, B. T. (2020). Using GIS, remote sensing, and machine learning to highlight the correlation between the land-use/land-cover changes and flash-flood potential. *Remote Sensing*, *12*(9). https://doi.org/10.3390/RS12091422

Dang, V. H., Hoang, N. D., Nguyen, L. M. D., Bui, D. T., & Samui, P. (2020). A novel GIS-Based random forest machine algorithm for the spatial prediction of shallow landslide susceptibility. *Forests*, *11*(1). https://doi.org/10.3390/f11010118

Elmoulat, M., & Brahim, L. A. (2018). Landslides susceptibility mapping using GIS and weights of evidence model in Tetouan-ras-Mazari area (Northern Morocco). *Geomatics, Natural Hazards and Risk*, *9*(1), 1306–1325. https://doi.org/10.1080/19475705.2018.1505666

Gorokhovich, Y., Doocy, S., Walyawula, F., Muwanga, A., & Nardi, F. (2013). Landslide Science and Practice. *Landslide Science and Practice*, *4*(2002), 145–149. https://doi.org/10.1007/978-3-642-31337-0

Herrera, M. (2016). *Landslide Detection using Random Forest Classifier*.

Jacobs, L., Dewitte, O., Poesen, J., Maes, J., Mertens, K., Sekajugo, J., & Kervyn, M. (2017). Landslide characteristics and spatial distribution in the Rwenzori Mountains, Uganda. *Journal of African Earth Sciences*, *134*, 917–930. https://doi.org/10.1016/j.jafrearsci.2016.05.013

Jacobs, L., Dewitte, O., Poesen, J., Sekajugo, J., Nobile, A., Rossi, M., Thiery, W., & Kervyn, M. (2018). Field-based landslide susceptibility assessment in a data-scarce environment: The populated areas of the Rwenzori Mountains. *Natural Hazards and Earth System Sciences*, *18*(1), 105–124. https://doi.org/10.5194/nhess-18-105-2018

Khoshtinat, S., Aminnejad, B., Hassanzadeh, Y., & Ahmadi, H. (2019). Application of GIS-based models of weights of evidence, weighting factor, and statistical index in spatial modeling of groundwater. *Journal of Hydroinformatics*, *21*(5), 745–760. https://doi.org/10.2166/hydro.2019.127

Kitutu, K. M. G. (2010). *Landslide Occurrences in the Hilly Areas of Bududa District in Eastern Uganda and Their Causes.* (Issue December).

Kitutu, M. G., Muwanga, A., Poesen, J., & Deckers, J. A. (2011). Farmer's perception on landslide occurrences in Bududa district, Eastern Uganda. *African Journal of Agricultural Research*, *6*(1), 7–18. https://doi.org/10.5897/AJAR09.158

Lee, S., Hong, S. M., & Jung, H. S. (2017). A support vector machine for landslide susceptibility mapping in Gangwon Province, Korea. *Sustainability (Switzerland)*, *9*(1), 15–19. https://doi.org/10.3390/su9010048

Margottini, C., Canuti, P., & Sassa, K. (n.d.). *Cees J. van Westen, Saibal Ghosh, Pankaj Jaiswal, Tapas Ranjan Martha (auth.), Claudio Margottini, Paolo Canuti, Kyoji Sassa (eds.)-Landslide Science and Practice_ Volume 1_ Landslide Inventory and Su.pdf.crdownload* (Vol. 1).

Marjanović, M., Bajat, B., & Kovačević, M. (2009). Landslide susceptibility assessment with machine learning algorithms. *International Conference on Intelligent Networking and Collaborative Systems, INCoS 2009*, *November*, 273–278. https://doi.org/10.1109/INCOS.2009.25

Moresi, F. V., Maesano, M., Collalti, A., Sidle, R. C., Matteucci, G., & Mugnozza, G. S. (2020). Mapping landslide prediction through a GIS-based model: A case study in a catchment in southern Italy. *Geosciences (Switzerland)*, *10*(8), 1–22. https://doi.org/10.3390/geosciences10080309

Nakileza, B. R., & Nedala, S. (2020). Topographic influence on landslides characteristics and implication for risk management in upper Manafwa catchment, Mt Elgon Uganda. *Geoenvironmental Disasters*, *7*(1). https://doi.org/10.1186/s40677-020-00160-0

Namono, M., Mugume, I., Negru, R., Mujuni, G., Tao, S., Racaud, S., & Oyama, S. (2019). The Barriers to Landslide Responses over the Mt. Elgon in Bududa District, Uganda. *Preprints*, *September*, 1–13. https://doi.org/10.20944/preprints201909.0022.v1

Odhiambo, B. D. O., Kataka, M. O., & Mashudu, M. (2019). *The Use of Remote Sensing to Map Landslide Prone Areas in Makhado Municipality of Limpopo Province , South Africa*. 1–28. https://www.preventionweb.net/publications/view/65868

Pourghasemi, H. R., Jirandeh, A. G., Pradhan, B., Xu, C., & Gokceoglu, C. (2013). Landslide susceptibility mapping using support vector machine and GIS at the Golestan province, Iran. *Journal of Earth System Science*, *122*(2), 349–369. https://doi.org/10.1007/s12040-013-0282-2

Rabby, Y. W., Hossain, M. B., & Abedin, J. (2020). Landslide susceptibility mapping in three Upazilas of Rangamati hill district Bangladesh: application and comparison of GIS-based machine learning methods. *Geocarto International*, *0*(0), 000. https://doi.org/10.1080/10106049.2020.1864026

Richefeu, V., & Villard, P. (2016). Modeling Gravity Hazards from Rockfalls to Landslides. In *Modeling Gravity Hazards from Rockfalls to Landslides*. https://doi.org/10.1016/c2015-0-01294-1

Segoni, S., Lagomarsino, D., Fanti, R., Moretti, S., & Casagli, N. (2015). Integration of rainfall thresholds and susceptibility maps in the Emilia Romagna (Italy) regional-scale landslide warning system. *Landslides*, *12*(4), 773–785. https://doi.org/10.1007/s10346-014-0502-0

SETAC. (2018). *Weight-of-Evidence in environmental risk assessment of chemicals*. 1–8.

Thanh Noi, P., & Kappas, M. (2017). Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors (Basel, Switzerland)*, *18*(1). https://doi.org/10.3390/s18010018

Types of landslide | NatureScot. (n.d.). Retrieved August 31, 2021, from https://www.nature.scot/landforms-and-geology/scotlands-rocks-landforms-and-soils/landforms/landslides/types-landslide

Vlaeminck, P., Maertens, M., Isabirye, M., Vanderhoydonks, F., Poesen, J., Deckers, S., & Vranken, L. (2016). Coping with landslide risk through preventive resettlement. Designing optimal strategies through choice experiments for the Mount Elgon region, Uganda. *Land Use Policy*, *51*, 301–311. https://doi.org/10.1016/j.landusepol.2015.11.023

Vojteková, J., & Vojtek, M. (2020). Assessment of landslide susceptibility at a local spatial scale applying the multi-criteria analysis and GIS: a case study from Slovakia. *Geomatics, Natural Hazards and Risk*, *11*(1), 131–148. https://doi.org/10.1080/19475705.2020.1713233

Westen, C. J. van. (2002). Use of weights of evidence modeling for landslide susceptibility mapping. *International Institute for Geoinformation Science and Earth Observation: Enschede, The Netherlands*, *January*, 23.

Wu, Y., Li, W., Wang, Q., Liu, Q., Yang, D., Xing, M., Pei, Y., & Yan, S. (2016). Landslide susceptibility assessment using frequency ratio, statistical index and certainty factor models for the Gangu County, China. *Arabian Journal of Geosciences*, *9*(2), 1–16. https://doi.org/10.1007/s12517-015-2112-0

Xiao, T., Segoni, S., Chen, L., Yin, K., & Casagli, N. (2020). A step beyond landslide susceptibility maps: a simple method to investigate and explain the different outcomes obtained by different approaches. *Landslides*, *17*(3), 627–640. https://doi.org/10.1007/s10346-019-01299-0

Xiao, T., Yin, K., Yao, T., & Liu, S. (2019). Spatial prediction of landslide susceptibility using GIS-based statistical and machine learning models in Wanzhou County, Three Gorges Reservoir, China. *Acta Geochimica*, *38*(5), 654–669. https://doi.org/10.1007/s11631-019-00341-1

Xiaowen, Z., & Min, J. (2009). Research on Landslide Prediction Model. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *38*(1), 406–410.

Yao, X., & Dai, F. (2006). Support vector machine modeling of landslide susceptibility using a GIS: A case study. *Iaeg2006*, *793*, 1–12. http://www.iaeg.info/iaeg2006/PAPERS/IAEG_793.PDF