LANDSLIDE SUSCEPTIBILITY PREDICTION AND WARNING SYSTEM USING MACHINE LEARNING AND GIS

FINAL YEAR PROJECT UG 2020

By

Sharjeel

Maria Mushtaq

Muhammad Ahad Khan

Muhammad Abdullah Ameen

Institute of Geographical Information Systems

School of Civil and Environmental Engineering

National University of Sciences and Technology, Islamabad, Pakistan

2024

This is to certify that the ‘Final Year Project’ titled:

LANDSLIDE SUSCEPTIBILITY PREDICTION AND WARNING SYSTEM USING MACHINE LEARNING AND GIS

submitted by

Sharjeel - 338752

Maria Mushtaq – 334712

Muhammad Ahad Khan - 336088

Muhammad Abdullah Ameen - 345894

has been accepted towards the requirements for an undergraduate degree in

GEO-INFORMATICS ENGINEERING

____

Quratulain Shafi

Lecturer

IGIS,SCEE

NUST, Islamabad, Pakistan

SUSTAINABLE DEVELOPMENT GOALS (SDGS)

This project can have a significant impact on life and infrastructure. This work shows the application of advanced technologies and a strong commitment to advancing global efforts towards sustainable development and climate resilience. This project is mainly comprised of three SDGs. (United Nations, 2020; Bui et al., 2018; Pal & Prasad, 2020).

ABSTRACT

Pakistan is one of the most prone regions to natural disaster such as landslides, floods, and droughts etc. Landslides are one of the most catastrophic events which occur frequently and cause numerous causalities and damage to infrastructure as well. The primary cause of landslides is high-intensity rainfall as we know it is the most focused factor in our research which we consider in predicting landslides. We also consider other influence factors such as elevation, altitude, aspect, plan curvature, profile curvature, distance from the river, roads and faults, lithology, soil, Normalized difference vegetation Index (NDVI), topographical wetness index (TWI) and soil moisture. In our research, we estimate the potential impact of these factors on our future landslide susceptibility. We generate an early warning system which alerts the government and public to take precautionary measures to save lives and maximum damage which can be caused by this hazard. We consider a machine learning model of a Random Forest Classifier to predict landslide susceptibility in Astore District, Gilgit Baltistan, Pakistan. We offer a detailed analysis and visual representation of landslide prone areas. Subsequently, We also highlights low, moderate and high-risk areas based on various factors. Additionally, we provide a customization platform, where user can identify high-risk areas of landslides, based on the provided data sets. The present study makes better and precise prediction and decisions to avoid most of the hazards. Predicting landslides is helpful in preventing most of the damage and helps in making better decisions to avoid major loss in future.

ACKNOWLEDGEMENTS

We would like to start by giving thanks to Allah, the Most Gracious the Most Merciful, who has supported us throughout this project.

We are also grateful to our supervisor, Quratulain Shafi, and all other instructors, and alumni who have consistently served as a source of guidance for us throughout.

Finally, we would like to express our gratitude to our friends and classmates who helped make the journey simpler by offering guidance and unwavering support throughout.

DEDICATIONS

Dedicated to our parents and everyone who has supported us from the start.

PREFACE

Our efforts have been directed toward two goals primarily.

First and foremost, we aim to test the constantly evolving waters of machine learning and effectively apply the information and abilities we have acquired at the Institute of Geographical Information Systems.

Second, to use our intellectual influence to support Pakistan's technological development. Our findings should provide a solid foundation for the future advancement of machine learning in the nation's present disaster management system.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Overview

One of the deadliest kinds of natural disasters is a landslide, which has the potential to do significant damage quickly. A portion of the land may slip, shift, or travel from a higher height to a lower elevation during this event. It is quite harmful because it happens quickly and leaves little time for taking preventative action. Its mobility allows for different varieties to be distinguished:

Rotational Landslide: This kind of landslide occurs when material slides over a curved surface in a rotating motion, moving below and outward. Another name for it is a slump.

Translational Landslide: This kind of landslide usually happens on extremely steep slopes when large rocks or other debris slide quickly downward along a nearly level surface. These landslides travel comparatively along a linear course, in contrast to circular landslides.

Block Landslide: Block landslides occur when a single, massive block slides down the hill. Because of their potential for large impacts, these blocks can cause significant damage. The displaced mass is not disturbed and is not broken during this activity.

Rock Fall: It features individual boulders or small rock fragments falling freely, bouncing, or rolling downslope. Rock falls at an increased velocity that can do massive damage (Farooq Ahmed & Rogers, 2016).

Topple: This kind of landslide occurs when rock shards shoot outward and forward from their source before collapsing below. Unless the situation worsens, it usually results in a tilt without collapsing (Youssef & Pourghasemi, 2021).

Debris Flow: This kind of landslide occurs when a mixture of rocks, water, and other debris slides into the water channel from the slope. It is extremely dangerous as it accelerates swiftly and damages surrounding areas and bridges across the water channel (Al-Najjar & Pradhan, 2021).

Earthflow: It is a landslide that moves slowly and is brought on by the sluggish movement of rocks and soil that have been saturated with water. Like how a thick liquid flows, it demonstrates continuously but moves slowly (Pham et al., 2022).

Creep Landslide: This kind of landslide is brought on by the rock and dirt-moving downslope very slowly. Since it moves more slowly than Earthflow, detection may need close observation. Weeks, months, or even years may pass as it happens. Although it usually cannot seriously endanger infrastructure or human life, it can be hazardous over an extended amount of time (Nguyen et al., 2019).

Lateral Spread Landslide: In this landslide, rocks and other materials are moved in a horizontal direction. It goes in a slope-parallel manner.

1.1.1 Triggering Factors

A critical first step in determining the landslide risk and putting preventive measures in place is the investigation and analysis of triggering causes. Gravitational force, which causes the mass to shift from higher to lower altitude, is the main force behind landslides. The factors can be divided into two groups: naturally occurring and man-made.

Landslide danger can be decreased by controlling human-induced elements at the local, state, and personal levels. A significant human-caused element that reduces the impact of vegetation on the surface is deforestation. The roots of plants prevent soil from moving by strengthening the bond of soil particles. Deforestation also increases surface runoff. When forests are removed, less water is absorbed, increasing surface runoff. Trees slow down the flow of water as they act as a barrier to stop the flow of water and decrease the chances of heavy destruction. Deforestation has a significant impact on climate change as deforestation raises atmospheric carbon dioxide levels and causes abrupt climate change which is a major contributor to landslides (Hussain et al., 2022). Other common human activities that raise the danger of landslides in steep terrain are building and excavation.

Excavation is generally done on mountains to develop structures and roads, which can alter the slope and natural topography of an area. Moreover, it leads to the loss of flora, which reduces the bonding between soil particles. Soil compaction, blasting, and excavating cause disturbance in the top layers of the soil and may result in mass displacement (Tien Bui et al., 2019). On the other hand, the implementation of the right engineering methods can reduce the danger and improve growth in hilly areas. A region's land cover can abruptly change due to human activities such as urbanization, and inadequate infrastructure design and maintenance, which raises the possibility of landslides (Marjanović et al., 2011).

There are natural elements that cause landslides that are outside of human control. Only via appropriate monitoring and surveillance, early warning system design, land use planning, etc., can be controlled and mitigated. Rainfall is the primary contributing element to the occurrence of landslides. Severe rainfall causes the soil to become saturated and increase in water content, which weakens the soil and increases the likelihood of sliding. Rainfall promotes surface runoff, which leads to the movement of loose particles with water. Rainfall also contributes to erosion, which makes rocks and soil more prone to slipping. Additionally, erosion lowers friction, which aids in the flowing mass's rapid acceleration and ultimately increases damage (Wang et al., 2021). Earthquakes are another cause. The surface of the earth is disturbed by earthquakes, which encourages the movement of rocks and other debris from higher elevations to lower elevations. Additionally, liquefaction brought on by seismic activity may weaken soil and silt and raise pore water pressure (Chen & Li, 2020). Soil reacts to sudden variations in temperature and humidity by changing its properties. The passage of hot gases, ash, and rock fragments during volcanic eruptions can cause terrain instability and raise the possibility of landslides (Goetz et al., 2015).

Landslides may also be caused by wildfires in the mountains. The cause may be human-caused or natural. It results in the removal of vegetation, which can weaken the surface soil and increase runoff. It may also have an impact on elements that contribute to climate change and raise the possibility and severity of landslides (Ng et al., 2021).

1.1.2 Background

One of the countries in the region most vulnerable to landslides, floods, droughts, and other natural calamities is Pakistan. Because of the region's varied topography, most landslides happen in mountainous locations, especially in the upper central to northern regions. One of the most devastating occurrences that frequently inflict many casualties and harm infrastructure are landslides. Every time there was a landslide, hundreds of people died. This study intends to predict landslides and perform a landslide susceptibility analysis in response to this concern. The study seeks to accomplish several important goals. The first goal is to create a machine learning model capable of precisely identifying landslide-prone locations. To do this, a model must be trained using environmental criteria and historical data to identify high-risk landslide sites (Xu et al., 2019). Making a map of landslide susceptibility is the other focused target. This map will be used to show the locations that are vulnerable to landslides. By emphasizing these regions, the map can give citizens, developers, and city planners important information that will assist them in making defensible decisions on development and land use (Wang et al., 2020). Analyzing proactive mitigation and forecasting future landslide incidents is the third goal. To predict future landslides, this entails analyzing the trends and causes of previous ones (Reichenbach et al., 2018).

The objective is to create plans that can lessen the effects of landslides and possibly stop them from occurring in the same place. Communities in landslide-prone locations may be safer and more sustainably built because of this thorough approach to landslide study. By doing this, we hope to lessen possible harm and, in the end, help to lower losses in the future. This study's key drivers of landslides include changes in land use and land cover, elevation, slope, and other factors. We have considered the effects of land cover, height, plan and profile curvatures, distance from rivers, roads, and faults, lithology, soil, and soil moisture in addition to climate change (Bui et al., 2018). Geographic Information System (GIS) approaches are used to analyze a variety of geospatial data, including topography, slope, geology, community locations, and land use, to assess landslide vulnerability (Fell et al., 2008).

Although they have been employed, techniques like logistic regression, entropy index, frequency ratio, and support vector machines have limitations when it comes to capturing intricate correlations and the need for large datasets (Chen et al., 2018). Landslide prediction may benefit from recent developments in machine learning, such as deep learning, LSTM, and artificial neural networks (ANN). Long-term temporal dependency mapping, or LSTM, is used to learn long-term temporal dependencies and has been used to forecast landslides based on topography, land cover, and hydrology, which are obtained via remote sensing and GIS data (Sun et al., 2020). Landslide prediction is difficult in Astore because of its varied topography and climate, with significant accidents happening every three to five years. Major cities and well-known locations have been the subject of disaster predictions in the past, and they are not particularly difficult to carry out because data collection— one of the most important and challenging steps—becomes extremely simple. However, predictions about lower scales, such as communities, are still being researched. Due to the intricate interconnections between numerous dynamic and unknown elements, predicting landslides in areas like Astore presents comparable difficulties (Froude & Petley, 2018).

As there is a deficiency of historical landslide data, sophisticated machine learning techniques and contemporary remote sensing analyses are required to identify important traits. One of the main advances is the utilization of the Google Earth platform, which offers real-time historical photos, to leverage limited past landslide data (Joyce et al., 2009). An underutilized method for predicting landslide susceptibility is bidirectional long short-term memory (Bi-LSTM), which takes dynamic elements like rainfall into account (Pal et al., 2020). Additionally, a web application that integrates rainfall data, landslide data, land use and land cover change, and trained models is constructed to show the danger of dynamic landslides online. Numerous investigations of the susceptibility of landslides have been carried out, but they have not been able to forecast the landslides (Dou et al., 2019). This study unveiled a clever GIS online tool that enables users to see locations that are anticipated to be vulnerable to landslides. This paper describes the landslide-risk web GIS platform's methodology in depth, assesses its effectiveness, and discusses potential directions for further research (Chen et al., 2020).

There are three primary goals for this project:

- Creating a machine learning algorithm to pinpoint locations vulnerable to landslides.
- Detecting places susceptible to landslides by creating maps of landslide susceptibility.
- Conducting analysis for preventive mitigation of landslide occurrence in the future.

1.1.3 Beneficiaries

This research will greatly benefit society with our novel and demanding approach to mapping and predicting landslide vulnerability. The National Disaster Management Authority (NDMA) Pakistan, our industrial partner in this initiative, is the main beneficiary. Communities residing in landslide-prone areas also stand to gain. This effort will save these communities' lives (Ahmed et al., 2018; NDMA, 2020). Moreover, this project might help NGOs and other organizations that support those impacted by catastrophes (World Bank, 2021). This project can serve as the foundation for upcoming research on mapping and predicting landslide vulnerability (Chung & Fabbri, 2003; Xu et al., 2020).

1.2 Literature Review

After a thorough analysis of the literature, it is determined that this research uses a novel method to forecast the frequency of landslides in the future. Most of the research only mapped the regions with a high risk of landslides. Prediction and mapping are both part of this research. Additionally, it offers a platform where users can click to visualize the projected areas.

Five model-based techniques were used for landslide-risk prediction in "Automated LandslideRisk Prediction Using Web GIS and Machine Learning Models": logistic regression (LR), artificial neural network (ANN), gated recurrent units (GRU), LSTM, and bidirectional LSTM (Bi-LSTM). Three places in Thailand where landslides had previously happened were used as model’s basis. This study found that the four factors it utilized were sufficient for the project. (Tengtrairat and associates, 2021)

Another study concludes that the Astore region's susceptibility to landslides can be ascertained by using the analytical hierarchy process (AHP), geographic information system (GIS), and remote sensing (RS) approaches. The nine primary causes of landslides were given AHP weights, and maps showing the Astore region's susceptibility to landslides were created. (Nouman et al., 2022)

In a research study to determine the accuracy of landslide occurrences, the model performance of SVM was compared with the other four machine learning-based techniques, including the analytic hierarchy process (AHP), logistic regression (LR), and artificial neural network (ANN). Field surveys, aerial imagery, satellite imagery, and Google Earth were used to gather the data. The findings of this study suggest that the type of region influences the accuracy of various algorithms. However, for landslide susceptibility mapping, it has been demonstrated that Generative Adversarial Networks (GAN) and Transfer Learning algorithms are highly accurate. (Liu and colleagues, 2023)

A research article predicting landslide susceptibility by utilizing decision tree machine learning models in the face of climate and land use changes. The study focuses on the Kalvan watershed in Iran, mapping past landslides within the area and analyzing various causative factors such as precipitation patterns and land use changes. By incorporating data on climatic variables under different scenarios and employing advanced techniques like the CA-Markov model for land use projections, the researchers highlight the significant impact of global warming and climate change on landslide occurrences. This study also emphasizes the importance of tuning parameters in machine learning models like Boosted Trees, Random Forests, and Extra Random Trees to enhance predictive accuracy. Overall, this research provides valuable insights into the complex interplay between environmental factors and landslide susceptibility (Quoc Bao Pham et al., 2021)

The other research work studies the application of the techniques of Machine learning (ML), such as Logistic Regression (LGR), Linear Regression (LR), and Support Vector Machine (SVM), which are compared to the clearing techniques including Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) and Analytical Hierarchy Process (AHP) for Muzaffarabad district. The study is designed to determine the critical factors which are used to forecast landslides, including slope, profile curvature, aspect, limestone IF, and precipitation. The results arise that ML methods perform better than MCDM methods when accuracy and efficiency are taken into account, so these findings are important considerations for landslide risk assessment and mitigation measures in the affected area.(Khalil et al., 2022)

The approach of this study is to Generative Adversarial Networks (GANs) which would then become a complement to the traditional machine learning methods. This would enhance the accuracy of landslide prediction. Cameron Highlands is the area where the study was conducted, using a multi-geo environmental framework that incorporates factors that were discussed by previous researchers. Compared to the established models like Artificial Neural Network (ANN), Support Vector Machine (SVM), Decision Trees (DT), Random Forest (RF), and Bagging ensembles, it is observed that GAN model gives the most notable trade-off between the prediction accuracy and computational complexity. The analysis, confirmed by AUROC record, indicates the usefulness of GAN-shaped things regarding affecting the prediction accuracy lack of data is present. (Husam et al., 2020)

The research by Wei Chen and collaborators presents a comprehensive study on landslide susceptibility modeling using Adaptive Neuro-Fuzzy Inference System (ANFIS) optimized with teaching-learning-based optimization and Satin bowerbird optimizer. The study involves four main stages: data preprocessing, determining weights of landslide influencing factors, assessing landslide susceptibility using ensemble models, and comparing and validating model performance. The spatial database preparation included historical landslide records, Google Earth data interpretations, and field investigations, identifying 152 landslides in the study area. The study area features altitudes decreasing from northwest to southeast, with predominant lithological types of loess, brown-red clay, sandstone, and carbonatite. The research integrates expertise from twenty experts to rank influencing factors, with weights calculated using the SWARA method. The models' performance is evaluated using various metrics such as RMSE, cumulative curve of processing speed, convergence curve, and AUROC. (Chen et al., 2020)

1.3 Problem Statement

Because of their rough terrain and complicated geological environment, Pakistan's Northern Areas frequently experience landslides, which pose serious dangers to the local infrastructure and population (Shah, 2018). In this area, traditional methods for assessing landslide susceptibility to risks frequently depend on oversimplified techniques, which results in poor forecast accuracy and restricted efficiency in reducing the risk of landslides (Dai et al., 2002). There is a strong chance to improve landslide susceptibility assessments for Pakistan's Northern Areas by utilizing the development of machine learning techniques and the wealth of geospatial data (Huang et al., 2020).

The goal of this research is to use machine learning techniques to create a reliable and accurate landslide susceptibility analysis model that is specifically adapted for the distinct geological and environmental conditions of Pakistan (Rahman et al., 2021). To enable proactive actions for disaster risk reduction and land use planning in the area, the main goal is to develop a prediction model that can identify locations with a high vulnerability to landslides (Ahmed et al., 2019).

CHAPTER 2

MATERIALS AND METHODS

2.1 Study Area

Astore is a district of Pakistan situated in the Gilgit-Baltistan province. It is located at an altitude of 2,600 meters (8,500 ft) on the eastern side of Mount Nanga Parbat. Astore Valley is one of the most beautiful and famous places for tourists' attraction (Khan, 2017). Apart from its beauty, its hilly terrain is very prone to landslides. Various historical landslide events have occurred there (Ali et al., 2016). This study also comprises major landslide events that happened in Astore. Specifically, in the winter season, all road networks are disconnected, and people face a lack of major resources like electricity. The main reason behind this is high rainfall and snowfall, which causes landslides (Hussain et al., 2019).

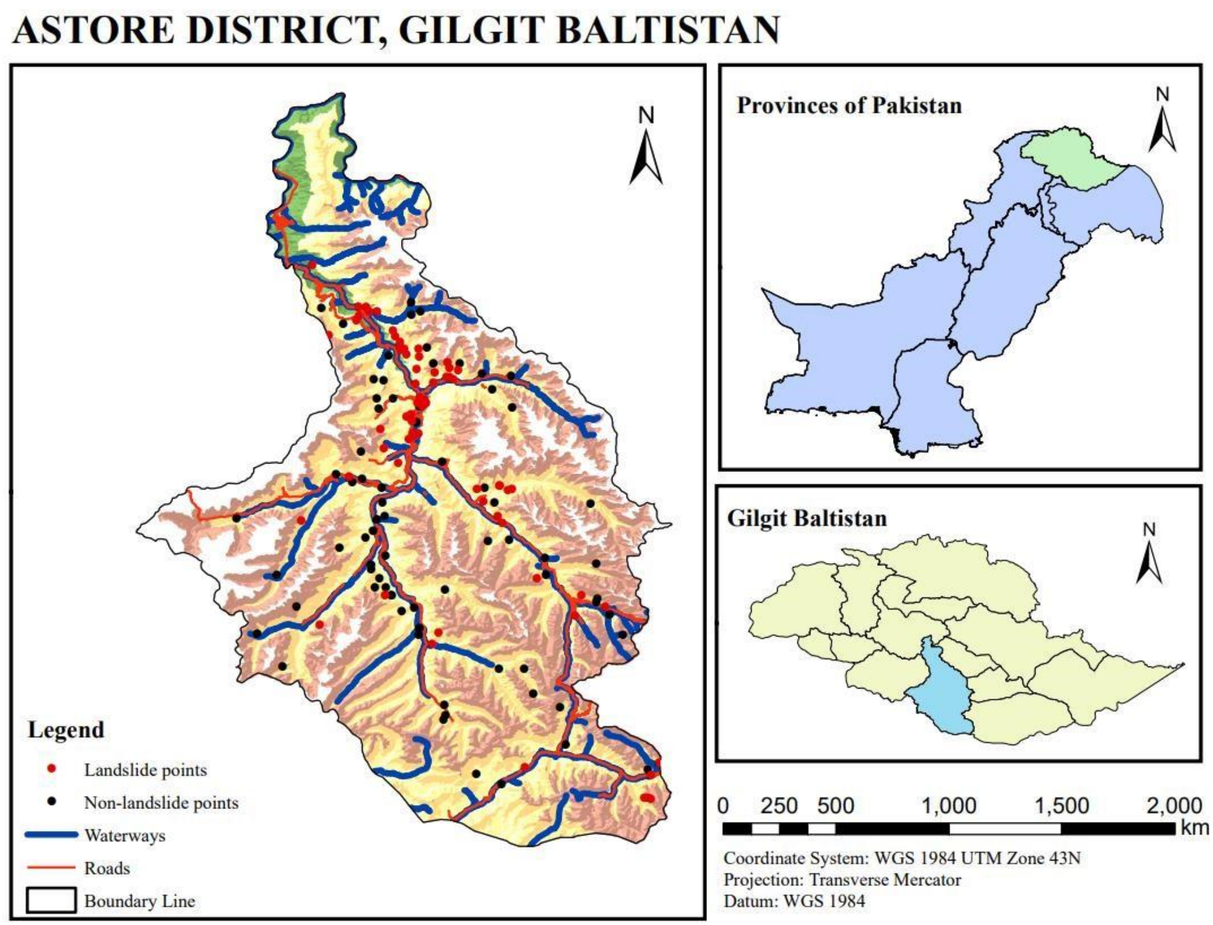

Fig. 2.1. Study Area

2.2 Datasets

15 datasets are chosen to train our machine-learning model to predict landslide susceptibility. These parameters are selected based on a thorough review of existing research papers and an extensive analysis of our study area (Chen & Li, 2020; Marjanović et al., 2011; Pham et al., 2022). Each data set is selected based on its relation to the factors that contribute to landslide occurrence and its compatibility with the machine learning algorithms used in our study (Liu et al., 2023; Khalil et al., 2022; Yuan & Chen, 2022). These datasets are compiled by some trustworthy means after a thorough understanding of all the geographical, geological, environmental, and socioeconomic variables that have the capability to contribute to the occurrence of landslides. We carefully examined every dataset to assess its quality, accuracy, and relevance to our research goals (Al-Najjar & Pradhan, 2021; Chen et al., 2021; Youssef & Pourghasemi, 2021).

To acquire the datasets, various research studies were consulted that addressed similar research questions and investigated publicly available repositories of geospatial data (Sadiki et al., 2022; Ali et al., 2021; Azarafza et al., 2021). Remote sensing techniques are also employed to gather site-specific information that complemented the existing datasets and enhanced the predictive capabilities of our model (Merghadi et al., 2020; Fang et al., 2020; Chang et al., 2019). Implementation of rigorous data preprocessing techniques is done to maintain consistency and reliability across the datasets. This involved data cleaning, normalization, and feature engineering to alleviate issues such as missing values, outliers, and redundant variables (Chen et al., 2018; Pham et al., 2016; Goetz et al., 2015).

Slope:

One of the most contributing factors in the occurrence of landslides is slope, acting as a critical factor that determines the stability of the terrain (Chen et al., 2020; Marjanović et al., 2011). The steeper the slope, the more susceptible the area is to landslides due to the gravitational pull on the soil or rock materials (Pham et al., 2022). To evaluate this factor in the study area, a Digital Elevation Model (DEM) with a resolution of 30 x 30 meters was acquired from the Shuttle Radar Topography Mission (SRTM) through the United States Geological Survey (USGS) platform (Liu et al., 2023). This high-resolution DEM provides detailed topographical information, essential for accurately assessing the terrain's slope (Khalil et al., 2022; Yuan & Chen, 2022). Subsequently, the

DEM was processed in ArcMap, a sophisticated geographic information system (GIS) software. Through ArcMap, we extracted the slope data specific to our study area, enabling a precise analysis of how varying slope degrees contribute to the occurrence of landslides (Al-Najjar & Pradhan, 2021; Chen et al., 2021; Youssef & Pourghasemi, 2021).

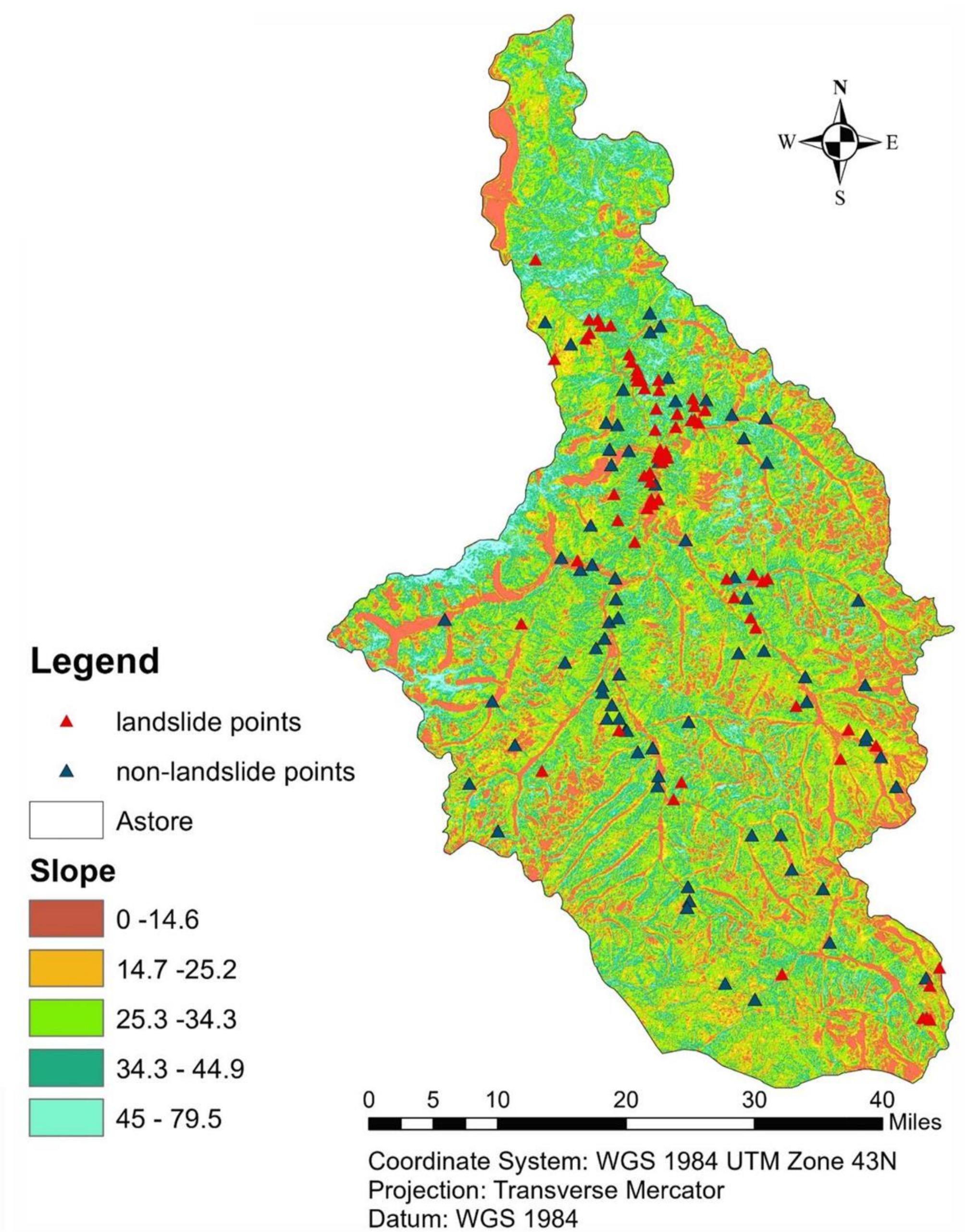

Fig. 2.2. Slope

Elevation:

Elevation also plays a significant role in the occurrence of landslides, influencing both the environmental conditions and the gravitational forces to contribute (Sadiki et al., 2022; Ali et al., 2021). Areas of higher elevation generally experience different weather patterns, like high precipitation events, which can saturate the soil and weaken rock formations, which results in increasing the risk of landslides (Azarafza et al., 2021). Moreover, areas of high elevation are more likely to have steep slopes, causing gravitational pull on the land (Merghadi et al., 2020; Fang et al., 2020). Atmospheric pressure at higher elevations also decreases which can affect soil moisture levels, indirectly influencing landslide susceptibility (Chang et al., 2019; Chen et al., 2018). Elevation data is also acquired from the same Digital Elevation Model (DEM) used to assess slope (Pham et al., 2016; Goetz et al., 2015). This comprehensive analysis helps to identify high-risk zones more accurately.

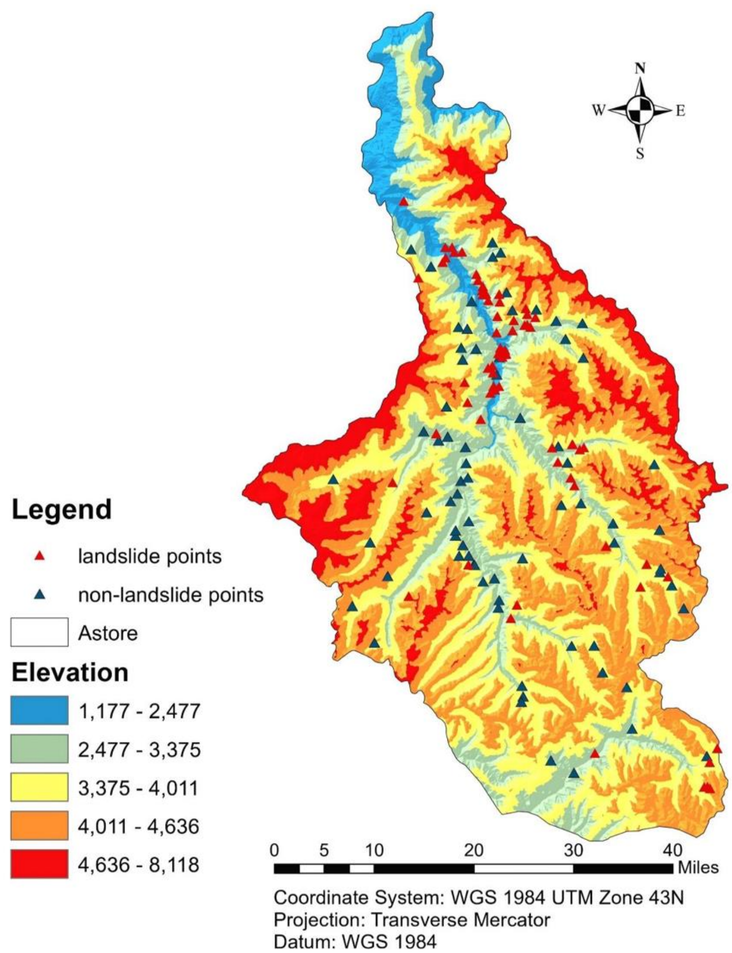

Fig. 2.3. Elevation

Aspect:

Aspect represents the compass direction that a slope faces. It is another critical factor influencing landslide susceptibility (Chen & Li, 2020; Marjanović et al., 2011). Aspect can significantly affect the amount of sunlight a slope receives which impacts soil moisture levels (Pham et al., 2022). Southern aspects in the northern hemisphere and northern aspects in the southern hemisphere receive more sunlight, leading to drier conditions that may decrease the likelihood of landslides by reducing soil saturation (Liu et al., 2023; Khalil et al., 2022). Similarly, slopes facing the north (in the Northern Hemisphere) are less exposed to sunlight, retaining more soil moisture and having higher landslide risks due to wetter soil conditions (Yuan & Chen, 2022). Additionally, aspect influences vegetation patterns, where sunnier slopes might support different types of vegetation than shadier slopes, further affecting soil stability and water absorption (Al-Najjar & Pradhan, 2021; Chen et al., 2021).

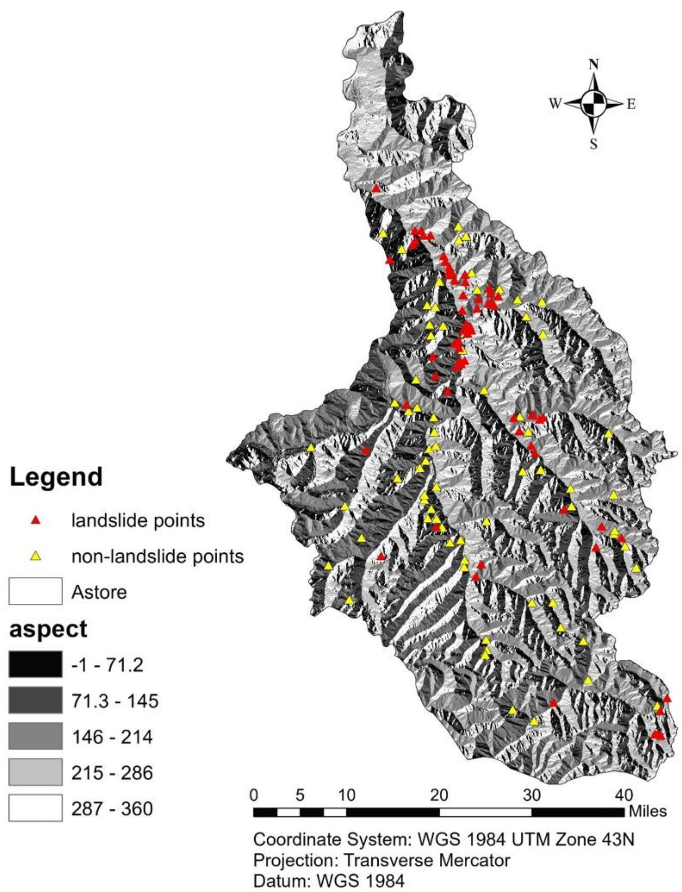

Fig. 2.4. Aspect

Topographical Wetness Index (TWI):

The Topographical Wetness Index (TWI), also known as the Compound Topographic Index, is also a major contributing factor in understanding how water movement and accumulation contribute to landslide susceptibility (Sadiki et al., 2022; Ali et al., 2021). TWI describes the spatial distribution of wetness across a landscape, reflecting areas more likely to experience soil saturation (Azarafza et al., 2021). This index is particularly useful in landslide studies because it integrates the slope and upstream catchment area to predict how water flows and accumulates, which can lead to increased soil weight and decreased cohesion, subsequently triggering landslides (Merghadi et al., 2020; Fang et al., 2020). Wetness conditions, envisaged by TWI, highlight potential landslide hotspots, especially in regions where heavy rainfall can lead to sudden soil saturation (Chang et al., 2019; Chen et al., 2018). By incorporating TWI into our model, alongside slope, altitude, and aspect data from the DEM, we can achieve a more holistic assessment of landslide risks (Pham et al., 2016; Goetz et al., 2015). The TWI not only aids in identifying areas prone to water accumulation and saturation but also helps in understanding the interplay between topographical features and hydrological processes, thereby enhancing the predictability of landslide occurrences.

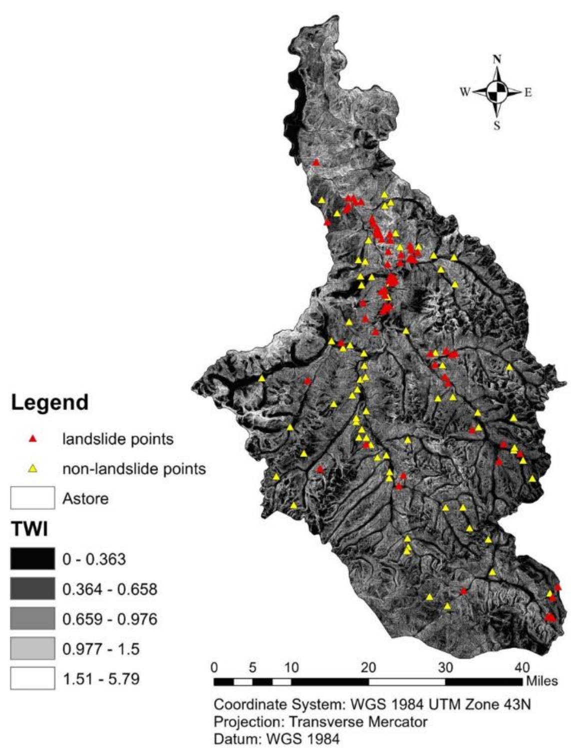

Fig. 2.5. Topographical Wetness Index

Plan Curvature:

It is discovered that the curvature of the terrain, or plan curvature, represents another dimension to landslide susceptibility assessments. This concept examines the curvature of the land surface in the horizontal plane and how it alters water flow paths and accumulation areas (Chen & Li, 2020; Marjanović et al., 2011). A concave surface gathers and holds water, causing saturation and increasing the likelihood of a landslide by making soil and rock more mobile (Afzal et al., 2022; Farooq Ahmed & David Rogers, 2016). Convex surfaces, on the other hand, enable water to run off, making it more difficult to become saturated, and as a result, less prone to a landslide (Pham et al., 2022; Liu et al., 2023). Therefore, predicting plan curvature provides a way of predicting water's movement on a slope and how water will accumulate in the slope, this identifying vulnerable regions (Khalil et al., 2022; Yuan & Chen, 2022). By analyzing plan curvature data extracted from the Digital Elevation Model (DEM), landscape vulnerability assessment becomes even more comprehensive, incorporating the land's shape and its effect on hydrological dynamics and landslide susceptibility (Al-Najjar & Pradhan, 2021; Chen et al., 2021).

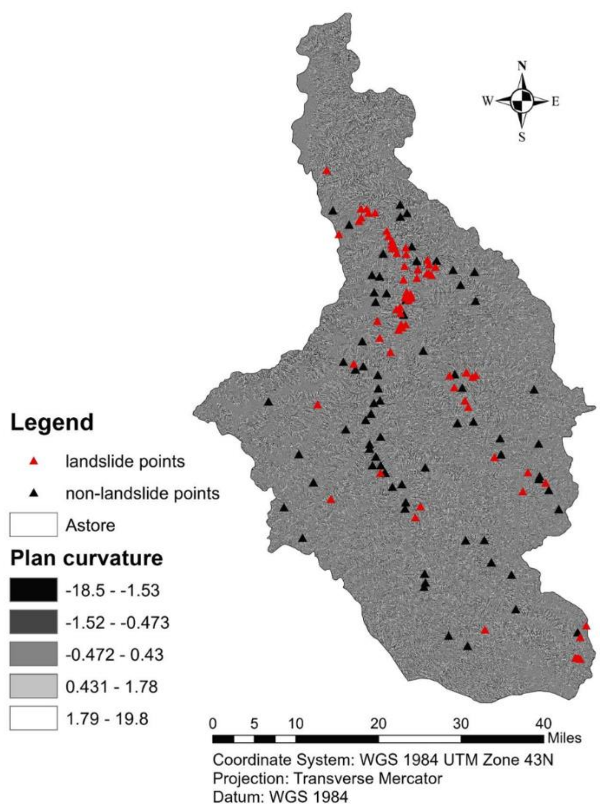

Fig. 2.6. Plan Curvature

Profile Curvature:

Profile curvature is the measure of the land's surface curvature along the direction of the slope, influencing the acceleration or deceleration of water flow and soil erosion on a hillside (Sadiki et al., 2022; Ali et al., 2021). This curvature affects the movement of water, sediments, and, subsequently, the erosional and depositional processes that can enhance or reduce landslide risk (Azarafza et al., 2021; Merghadi et al., 2020). In areas with a high degree of profile curvature, water may accelerate down the slope, leading to an increase in erosion at the base and potentially destabilizing the slope by removing support material (Fang et al., 2020; Chang et al., 2019). Conversely, a gentle profile curvature can slow water movement, reducing erosion but potentially increasing saturation and weight on the slope, which could also contribute to landslide risks (Chen et al., 2018; Pham et al., 2016). So, better analysis can be performed with the help of profile curvature that controls the direction of water flow and the speed at which water moves on the surface (Goetz et al., 2015; Korup & Stolle, 2014). This set of data makes the prediction more valid and includes landslide susceptibility to the full terrain.

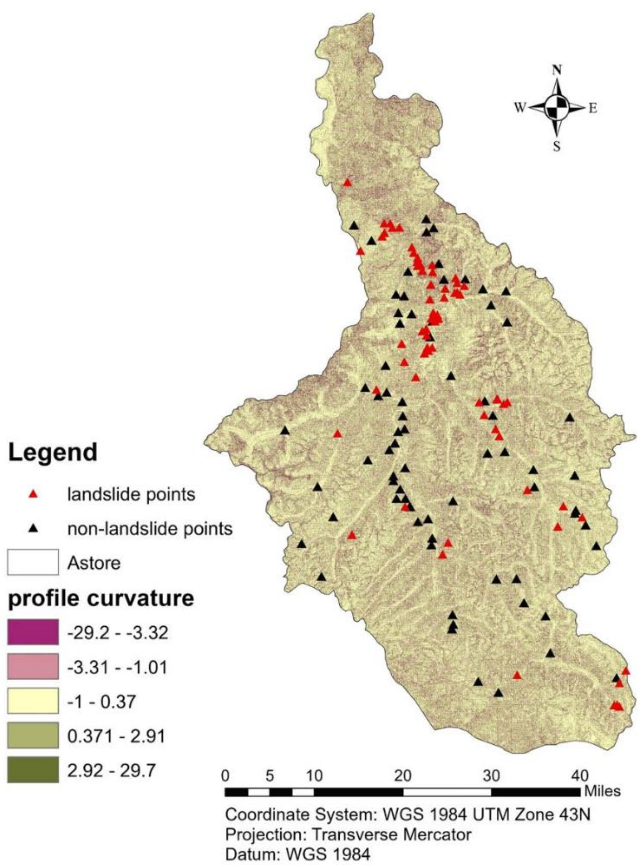

Fig. 2.7. Profile Curvature

Soil Moisture:

Soil moisture, a critical parameter obtained via the Soil Moisture Active Passive (SMAP) satellite, plays a pivotal role in assessing landslide risks (Ng et al., 2021). Utilizing Google Earth Engine, we utilized SMAP's capabilities to measure subsurface soil moisture levels globally every 2-3 days (Tehrani et al., 2022). This near-surface soil moisture data is instrumental in identifying potential landslide triggers, particularly following periods of heavy rainfall or rapid snowmelt (Ng et al., 2021). Saturated soil increases in weight and pressure on slope materials, significantly reducing the frictional force that holds slope materials in place, thereby increasing the likelihood of landslides (Chang et al., 2019). The decrease in soil cohesion due to saturation is a primary factor in slope failures, as water acts as a lubricant, facilitating the downward movement of earth materials (Chang et al., 2019). By integrating SMAP's soil moisture data into our analysis, dynamic assessment of landslide vulnerability is performed, offering an opportunity for timely interventions in high-risk zones (Tehrani et al., 2022).

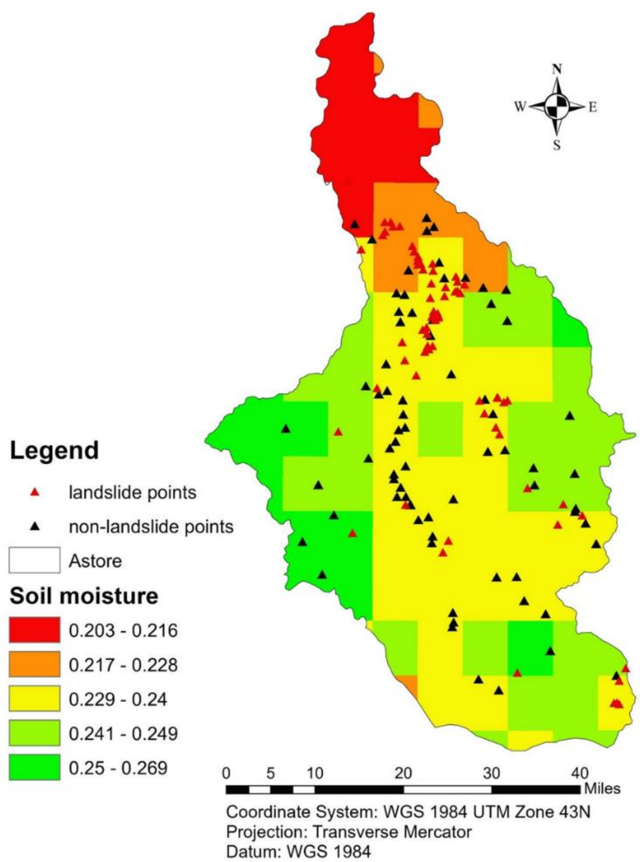

Fig. 2.8. Soil Moisture

Normalized Difference Vegetation Index (NDVI):

The Normalized Difference Vegetation Index (NDVI), derived from Landsat 8 images, offers significant insights into vegetation cover and health which are indirectly related to landslide susceptibility (Chen et al., 2018). NDVI values range from -1 to 1, with higher values indicating dense and healthy vegetation (Chen et al., 2018). Dense vegetation can play a dual role in landslide susceptibility. On one hand, it can anchor soil and reduce the risk of landslides by holding the soil in place through root systems, thereby providing slope stability (Chen et al., 2018). On the other hand, areas with sudden changes in vegetation density, which may be indicated by variations in NDVI values, could signal regions where landslides have removed vegetation or where weak vegetation cannot adequately stabilize the soil (Kainthura & Sharma, 2022). Additionally, the transition zones where high NDVI values abruptly shift to lower values could point to potential landslide paths or areas where landslide susceptibility is increased due to lesser vegetation cover or health (Chen et al., 2018). By integrating NDVI data from Landsat 8 into our comprehensive risk analysis model, we not only gain insight into the current state of vegetation cover but also enrich our understanding of how changes in vegetation can alert us to areas potentially at risk of future landslides (Kainthura & Sharma, 2022).

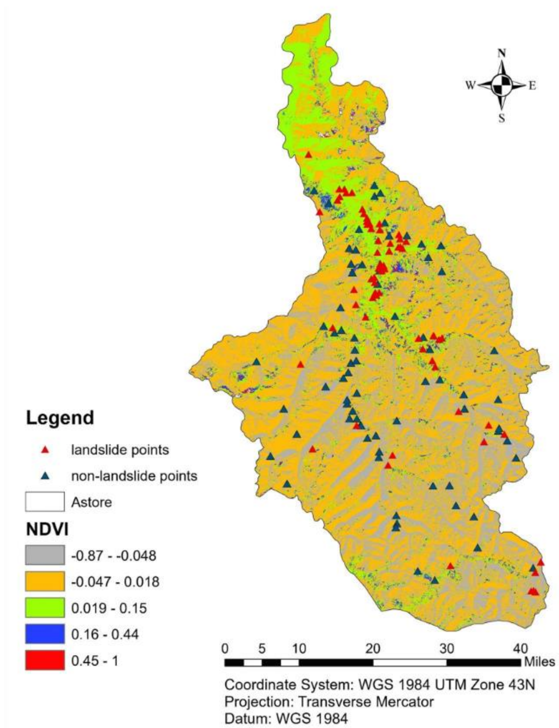

Fig. 2.9. Normalized Difference Vegetation Index

Historical Rainfall:

Historical rainfall data, acquired through the Climate Research Unit (CRU), plays a crucial role in enhancing our understanding of landslide susceptibility (Jones et al., 2021). By analyzing patterns of precipitation over time, it becomes possible to identify trends and periods of increased rainfall that significantly elevate landslide risks (Jones et al., 2021). This historical perspective allows for the correlation between past landslide occurrences and rainfall events (Jones et al., 2021). The CRU's comprehensive dataset, which includes detailed precipitation records spanning multiple decades, offers invaluable insights into seasonal rainfall variability and extreme weather events (Jones et al., 2021). By incorporating this data, landslide vulnerability assessments further augment the accuracy and efficiency of our landslide prediction models, ensuring a more robust approach to landslide risk management (Jones et al., 2021).

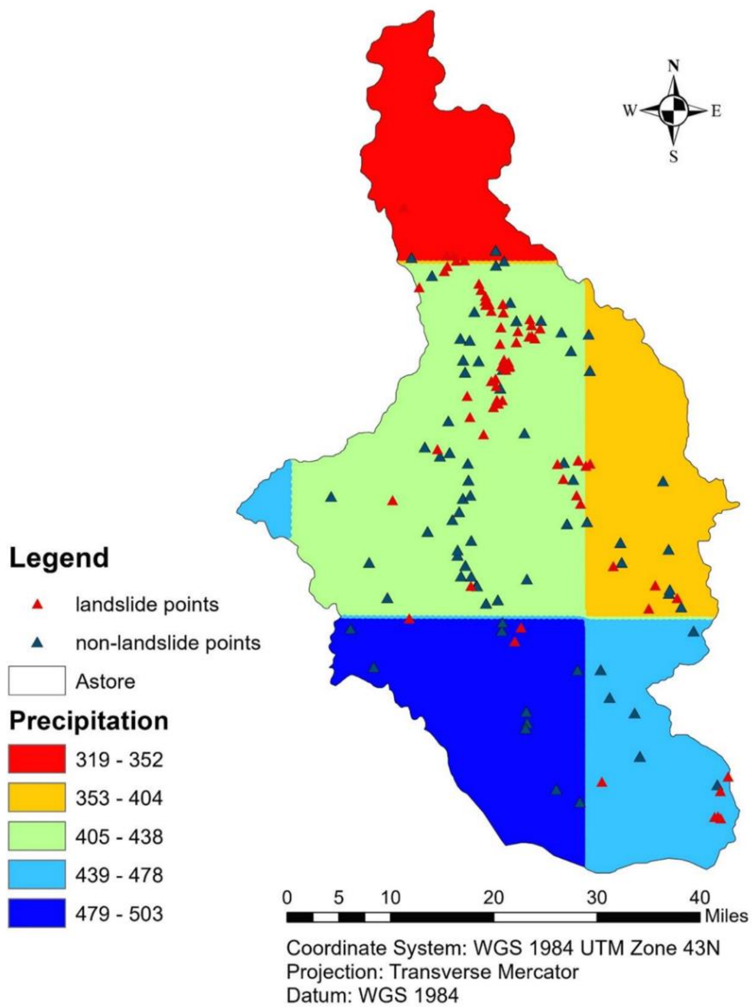

Fig. 2.10. Rainfall

Lithology:

Lithology, a key factor obtained from geological data, significantly influences landslide susceptibility by determining the strength, structure, and composition of rock and soil materials within a landscape (Youssef & Pourghasemi, 2021). Different lithological units possess varying degrees of resistance to weathering and erosion, which in turn affects their stability under stress (Youssef & Pourghasemi, 2021). For instance, sedimentary rocks, such as shale and sandstone, may offer less resistance compared to igneous rocks like granite (Youssef & Pourghasemi, 2021). The presence of clay minerals in some sedimentary rocks can lead to a reduction in shear strength, especially when wet, thereby increasing the potential for landslides (Youssef & Pourghasemi, 2021). Furthermore, the orientation and spacing of bedding planes, fractures, and faults within these lithological units can dictate the pathways for water infiltration, further influencing slope stability (Youssef & Pourghasemi, 2021). These rocks are classified after consulting various research studies, which offers a better understanding of rock types and their contribution to cause landslides (Youssef & Pourghasemi, 2021).

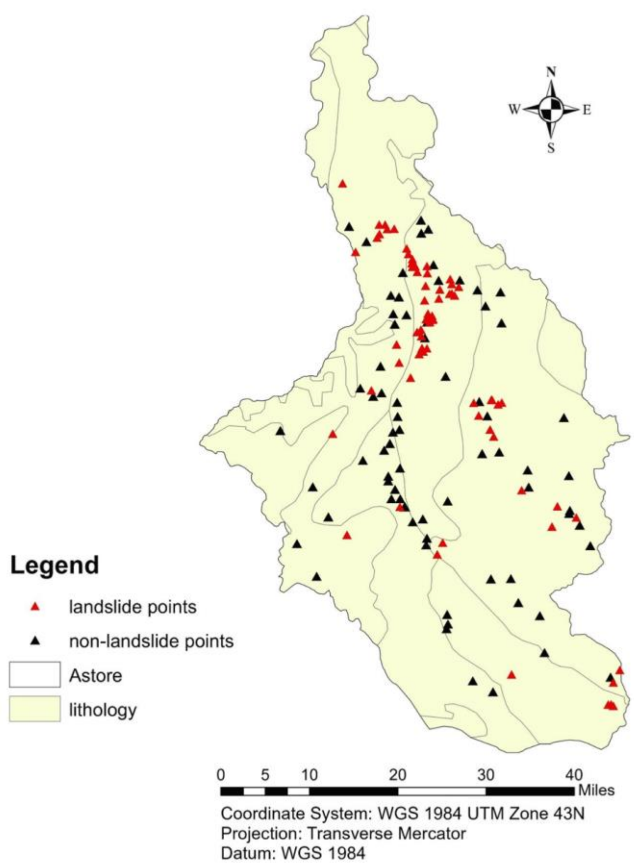

Fig. 2.11. Lithology

Soil groups:

The impact of various soil groups on landslide susceptibility is a critical component in identifying landslide-prone areas (Chen et al., 2018). Soil types, characterized by their grain size, composition, and cohesion, play a significant role in determining the stability of slopes (Chen et al., 2018). For instance, clay-rich soils are particularly prone to landslides due to their low shear strength and high potential for water retention, which can lead to a decrease in slope stability under saturated conditions (Chen et al., 2018). On the other hand, sandy soils, with larger grain sizes and better drainage properties, may offer more resistance to sliding (Chen et al., 2018). However, under certain conditions, such as during intense rainfall or rapid snowmelt, even sandy soils can become destabilized (Chen et al., 2018). By identifying areas with soil types that are inherently prone to failure, targeted land management and mitigation strategies can be developed to reduce the impact of landslides on communities and infrastructure (Chen et al., 2018).

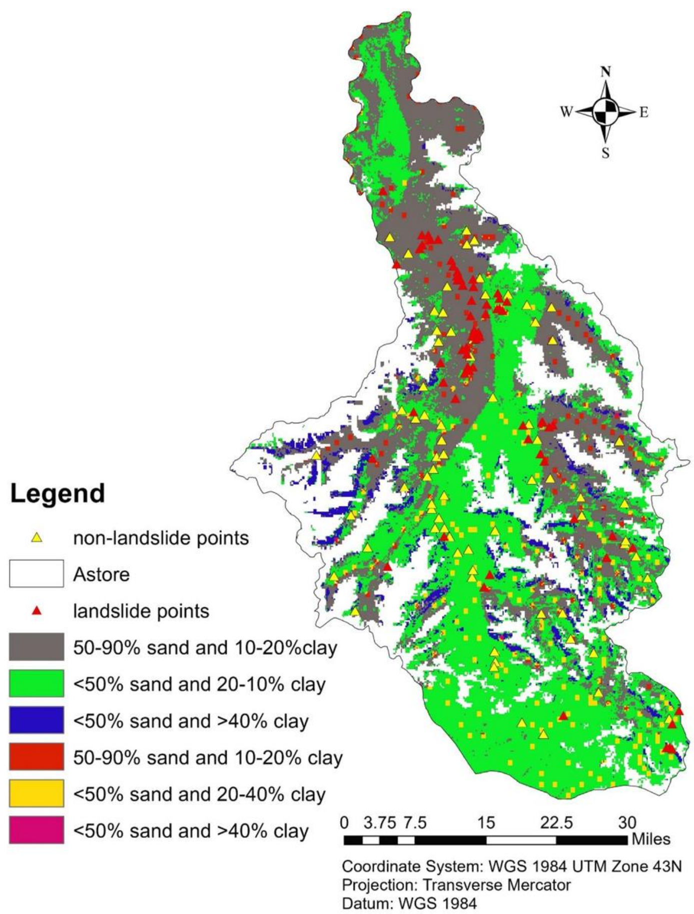

Fig. 2.12. Soil Groups

Land use and Land cover:

Land use and land cover (LULC) data obtained from Sentinel-2 satellite imagery play a crucial role in understanding and assessing landslide susceptibility (Chen et al., 2018). These data allow for the detailed mapping and analysis of vegetation cover, urbanization levels, and agricultural activities across landscapes (Chen et al., 2018). Changes in land cover, such as deforestation for agriculture or urban development, directly affect the stability of slopes and increase the potential for landslides (Chen et al., 2018). Vegetation removal exposes soil to erosion, reducing its cohesiveness and the slope's overall stability (Chen et al., 2018). Urban development, on the other hand, often leads to increased surface runoff and altered drainage patterns, potentially exacerbating soil saturation and weakening slope materials (Chen et al., 2018).

Distance to roads:

Distance to roads is another critical factor derived from geological data that affects landslide susceptibility (Chen et al., 2018). Roads, especially those constructed without adequate consideration for natural landscapes, can significantly alter the hydrological and stability conditions of slopes (Chen et al., 2018). The excavation and deposition of material during road construction often lead to an uneven distribution of weight on slopes, increasing the risk of landslides (Chen et al., 2018). Furthermore, the alteration of natural drainage patterns can lead to waterlogging on the upslope side, exacerbating soil saturation and reducing slope stability (Chen et al., 2018). The proximity of terrain to roadways is therefore a vital component in assessing landslide risk (Chen et al., 2018). Areas close to roads are often subject to increased water runoff, vibration, and human-induced alterations, all of which can weaken slope materials and trigger landslides (Chen et al., 2018). So, areas near roads are more prone to landslides.

Distance to faults:

Distance to faults is a crucial factor in assessing landslide susceptibility, acquired through geological data (Chen et al., 2018). Faults represent zones of weakness in the Earth's crust where significant differential movement has occurred, leading to the creation of fractures and discontinuities in rock masses (Chen et al., 2018). These structural weaknesses can significantly impact the stability of slopes and landscapes (Chen et al., 2018). The proximity to fault lines often correlates with an increased incidence of landslides, especially during or after seismic events, as seismic waves can trigger landslides by shaking loose already unstable slope materials (Chen et al., 2018). Additionally, fault zones may alter hydrological conditions by providing pathways for water movement, which can exacerbate soil saturation and reduce slope stability (Chen et al., 2018). By integrating distance to faults into landslide susceptibility analysis, those areas can be identified where the underlying geological conditions may predispose the terrain to higher landslide risks (Chen et al., 2018).

Distance to waterways:

Distance to waterways is a pivotal factor in landslide susceptibility, also acquired through geological data (Chen et al., 2018). Water bodies, including rivers and streams, can have a profound impact on the stability of adjacent slopes and terrains (Chen et al., 2018). The continuous flow of water can lead to erosion of the banks, gradually weakening the structural integrity of slopes and increasing the risk of landslides (Chen et al., 2018). Additionally, areas close to waterways are more likely to experience changes in soil moisture content, with the potential for water saturation during high flow periods (Chen et al., 2018). This saturation reduces the cohesive strength of soil and rock materials, making slopes more susceptible to sliding (Chen et al., 2018). Furthermore, the undercutting action of water can create overhanging slopes that are inherently unstable (Chen et al., 2018). By factoring in the distance to nearby waterways, this project can identify regions where the interaction between water bodies and slopes might elevate landslide risks (Chen et al., 2018).

Historical landslide data:

Historical landslide events possess invaluable data, offering critical insights into patterns and triggers of past incidents, which are important in predicting and mitigating future landslides (Chen et al., 2018). By analyzing different historical images of our study area on Google Earth, we acquire a record of major locations where mass was displaced in the past (Chen et al., 2018). These images provide a visual record of changes in land use, vegetation cover, and the after conditions of previous landslides, allowing us to identify locations that are particularly vulnerable to landslides (Chen et al., 2018). Once this historical data is gathered, it is exported as a KML file and processed in ArcMap to prepare the data to relate with other datasets in order to find the relationship between different factors (Chen et al., 2018).

Normalization of datasets:

The standardization of data via normalization in ArcGIS is a pivotal step in our project, ensuring that diverse datasets on soil composition, land use patterns, and historical landslide data can be directly compared on a singular uniform scale (Chen et al., 2018). This process is critical for performing multidimensional analysis, enabling the identification of correlations and interactions between variables that may influence landslide susceptibility (Chen et al., 2018). In the next step, the sampling tool is utilized in ArcGIS to fetch values of all normalized datasets according to the point acquired from Google Earth (Chen et al., 2018). After the creation of a tabular database through the sampling tool, this data is converted to CSV format (Chen et al., 2018). CSV format is used not only because of its convenience regarding updating but also its integration with Python and other tools is effective (Chen et al., 2018). The CSV file is utilized as a database in our project (Chen et al., 2018). This organized dataset not only provides a solid foundation for statistical analysis or the application of machine learning algorithms but also significantly contributes to the accurate working of predictive models (Chen et al., 2018). By facilitating a deeper understanding of the factors leading to landslides, this approach allows the development of increasingly accurate and reliable landslide susceptibility assessments (Chen et al., 2018). This integration of geospatial normalization techniques ultimately contributes to the development of more effective landslide monitoring, prevention, and mitigation strategies.

The meticulously prepared and normalized dataset serves as the foundation for training our Machine Learning (ML) Model (Chen et al., 2018). By utilizing the cleaned, structured data, we ensure that the input to our ML algorithms is of the highest quality, which is crucial for developing accurate and reliable predictive models (Chen et al., 2018). The application of machine learning in analyzing and predicting landslide susceptibility involves feeding the model with variables that include geological formations, soil properties, land use patterns, and historical landslide data (Chen et al., 2018). This model is then trained to identify patterns and correlations within these variables, enabling it to predict future landslide occurrences with a significant level of accuracy (Chen et al., 2018). Through iterative training and validation processes, the ML model's performance is continuously refined, enhancing its predictive capabilities (Chen et al., 2018).

Forecasted Rainfall:

The incorporation of real-time forecasted rainfall data as obtained from the OpenWeatherMap API marks a significant advancement in our landslide predictive capabilities (Chen et al., 2018). Recognizing that rainfall is the most contributing factor in triggering landslides, a Python script is developed to automatically fetch and normalize this data for integration into our existing datasets (Chen et al., 2018). This rainfall data is updated every three hours, and our model will generate output according to the latest conditions with improved accuracy (Chen et al., 2018). This enhancement not only aids in providing timely alerts that can save lives and property but also strengthens the efficiency of our landslide mitigation strategies (Chen et al., 2018). Access to real-time weather conditions through such a technologically advanced approach ensures that our predictive models are not only more accurate but also capable of offering real-time insights, greatly bolstering the resilience of communities at risk of landslides (Chen et al., 2018). This strategic incorporation of dynamically updated weather data, particularly rainfall, significantly elevates the reliability and functionality of our landslide susceptibility assessments (Chen et al., 2018).

2.3 Methodology

The complicated geological processes known as landslides are caused by a variety of elements, such as rainfall, seismic activity, soil characteristics, topography, and human activity. Effectively assessing these characteristics to foresee potential occurrences is the key to landslide prediction. To accurately anticipate landslides, we set out to create a highly advanced machine learning model by extracting fundamental knowledge and insights from a variety of academic articles and research projects.

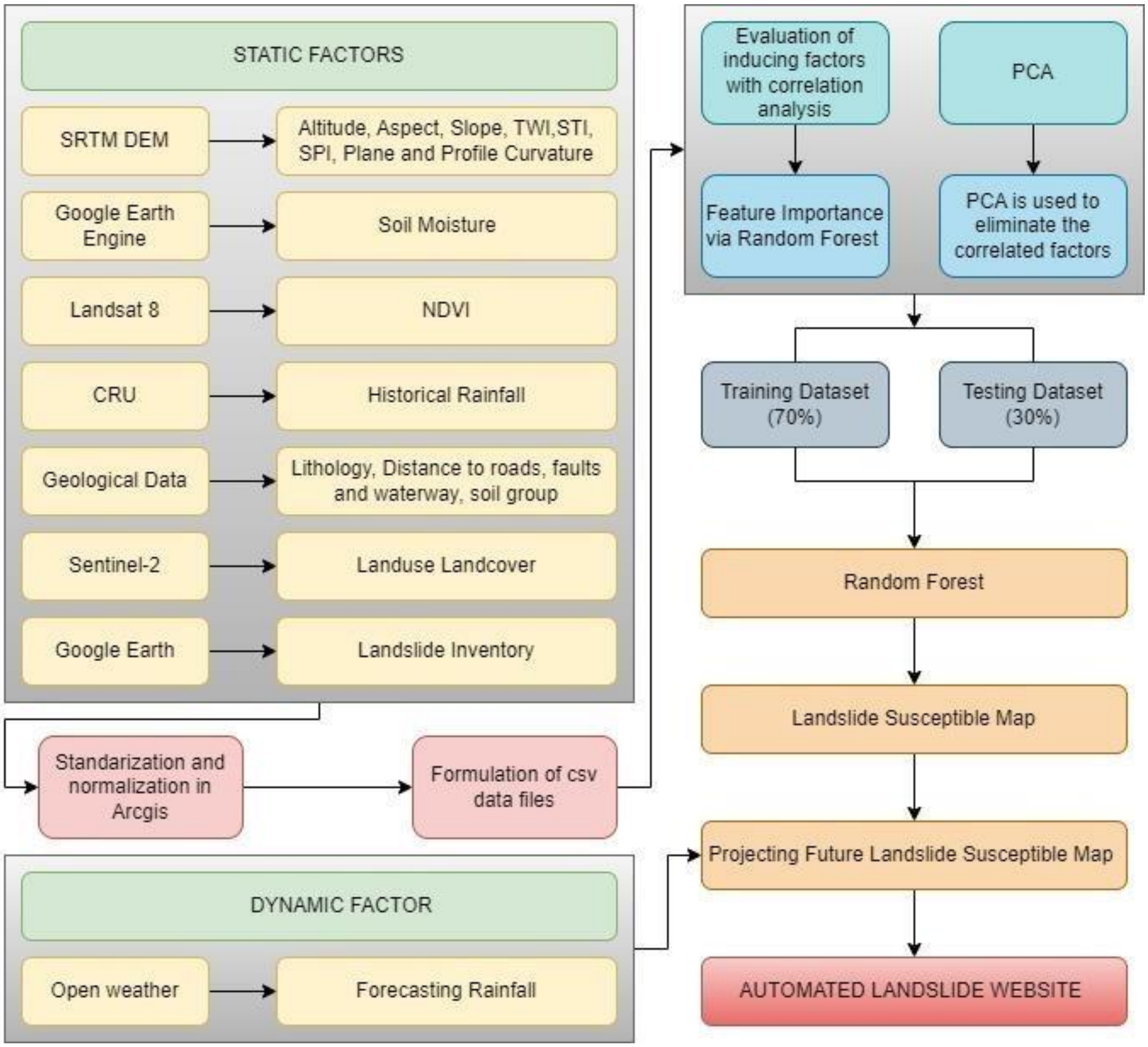

Fig. 2.13. Methodology flow chart

The first stage of our exploration phase involved the compilation of 146 points, which were painstakingly divided into two main datasets: 73 points denoting places that had previously had landslides, and 73 points denoting areas that had not (Chen et al., 2018). This well-balanced method guarantees that the model learns to distinguish between safe and landslide-prone locations (Chen et al., 2018). Fifteen separate datasets, each collecting different features essential for analysis, are present in each of these sites (Chen et al., 2018).

Our model training step begins with the consolidation of this data into an extensive CSV file (Chen et al., 2018).

To convert our theoretical model into an effective forecasting tool, we used a deliberate training strategy (Chen et al., 2018). We attempted to achieve the best possible balance between learning complexity and model generalization by dedicating 70% of our dataset to training and the remaining 30% to testing (Chen et al., 2018). This segmentation guarantees that the model's performance is appropriately assessed against unknown data in addition to facilitating in-depth learning from the training set (Chen et al., 2018).

Choosing the right machine learning algorithm is the key to our forecasting model. We investigated a variety of methods, but because Logistic Regression, Decision tree, and Random Forest classifier are so good at solving classification problems, we concentrated on them most.

After extensive testing and validation, the Random Forest algorithm showed itself to be the best option with an astounding 81.01% accuracy rate while the Decision Tree and Logistic Regression showed the accuracy of 78.38% and 70.27% respectively (Chen et al., 2018). This result demonstrates how well the algorithm handles the complexity and unpredictability that comes with landslide prediction (Chen et al., 2018). The confusion Matrix of the algorithms are as follows:

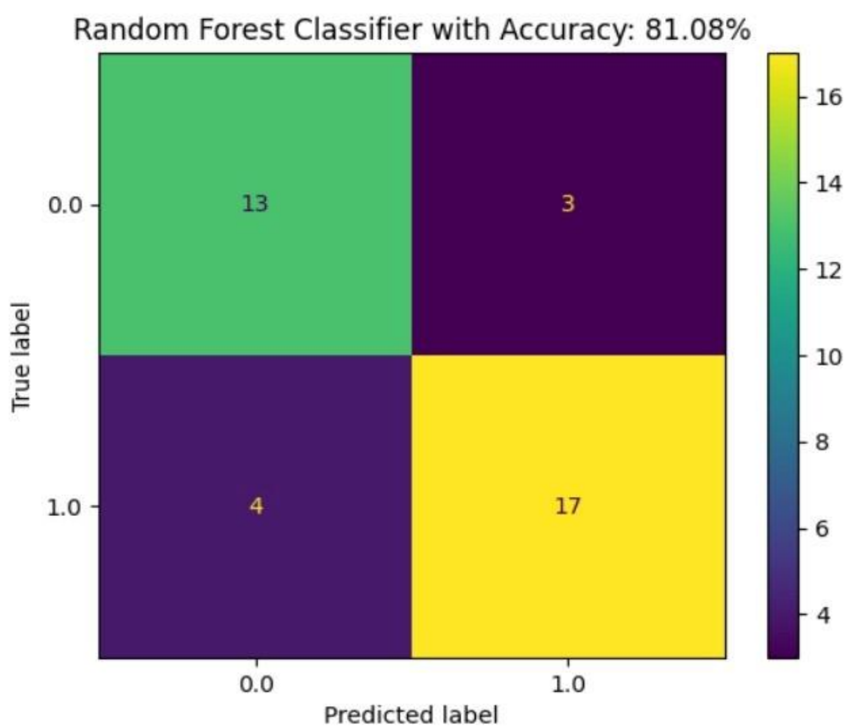

Fig. 2.14. Random Forest Classifier

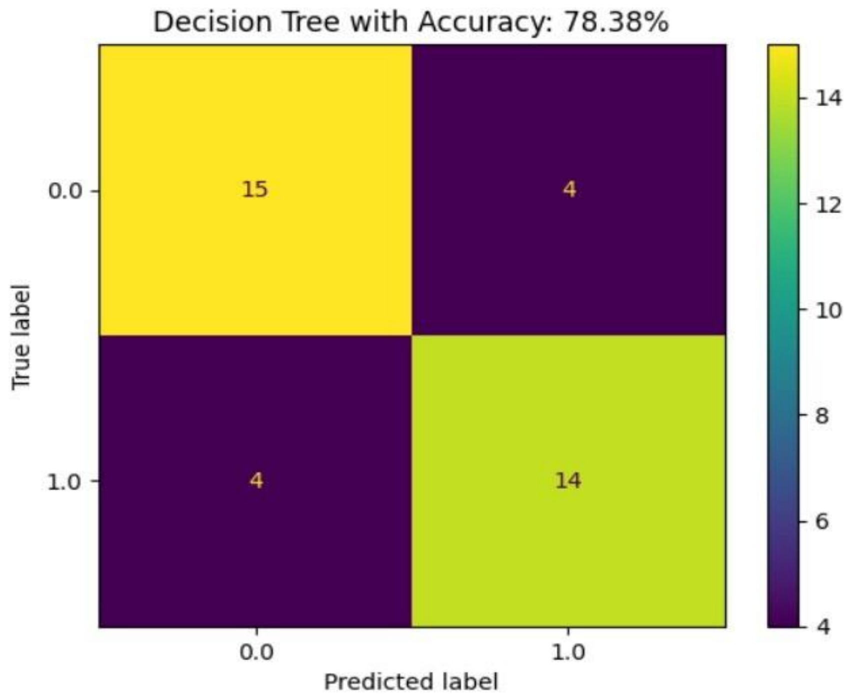

Fig. 2.15. Decision Tree

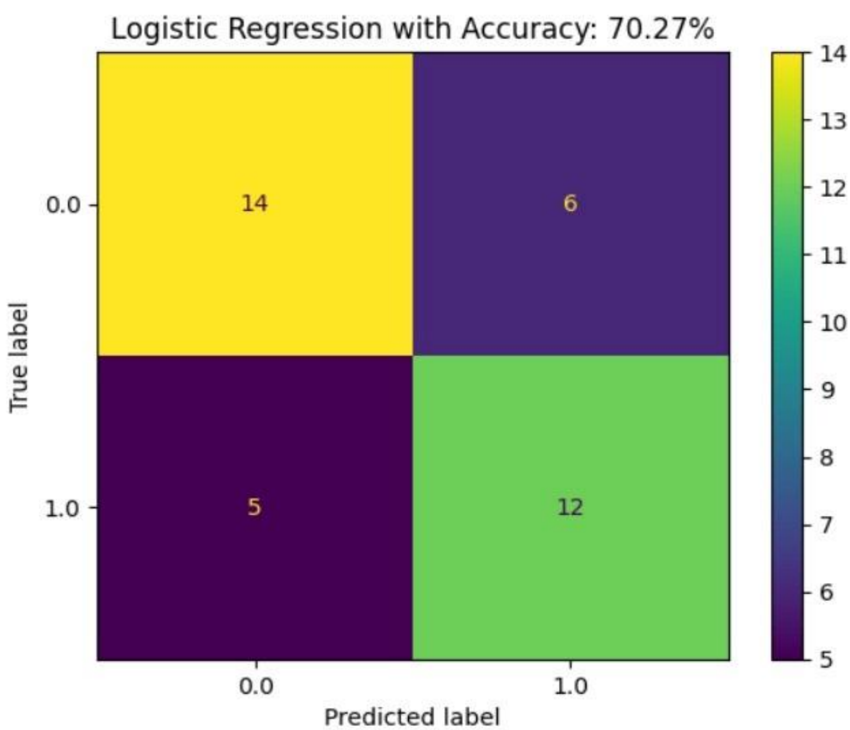

Fig. 2.16. Logistic Regression

A thorough overview of the region's varied geological and environmental circumstances is represented by the 1,000 randomly chosen spots that we meticulously selected and distributed over the whole study area. To guarantee that the learning base of our model is as large and representative as feasible, this method seeks to capture the complex and diverse aspects that can impact the incidence of landslides (Chen et al., 2018). The creation of the prediction dataset was a vital stage that came after our dataset had grown. This CSV file-formatted dataset includes all 15 datasets for every point in the extensive 1,000-points (Chen et al., 2018).

Adding real-time rainfall data to our prediction algorithm makes it far more dynamic and responsive. We have developed a technique to automatically gather accurate precipitation data for each of the 1,000 places that are being considered, using the OpenWeatherMap API (Chen et al., 2018). This calculated decision makes use of the API's capacity to make separate calls for every point, providing our model with the most recent meteorological data that is pertinent to the prediction of landslides (Chen et al., 2018). One of the most important parts of our strategy is incorporating this real-time data into our CSV file. The precipitation data is gathered, saved, and routinely updated after three hours (Chen et al., 2018). By regularly updating the data, our model can function on the most current and pertinent information, resulting in forecasts that are more precise and timelier. This dynamic factor not only enriches our dataset but also underscores our commitment to leveraging cutting-edge technology to enhance the reliability and precision of our landslide forecasts (Chen et al., 2018).

Our prediction model's operating workflow is resilient and dynamic, designed to adjust to the constantly shifting environmental conditions. After our CSV file is updated with real-time precipitation data, our machine learning model is trained on the revised dataset. The method is carefully designed to categorize every one of the 1,000 spots according to the likelihood of a landslide right now. The output is saved straight into the CSV file with the values 0 (no risk found) or 1 (risk detected) in the recently added ‘results' column (Chen et al., 2018). To preserve the applicability and precision of our landslide forecasts, this process is cyclical, repeating after every three hours. Each cycle begins with the automated fetching of updated precipitation data from the OpenWeatherMap API for each point. The CSV file is then refreshed with this new data, and the updated dataset is subsequently processed through the model to generate the latest risk assessments, which are saved back into the 'results' column of the CSV file (Chen et al., 2018).By studying a variety of research studies and interacting with subject matter experts, we established an enhanced strategy to further improve our approach to classifying the landslide danger points. Four main factors—rainfall, slope, soil moisture, and the Topographic Wetness Index (TWI)—were found to be significantly associated with the occurrence of landslides as a result of this cooperative effort. We evaluated each factor's effect and gave it a weight based on how likely landslides are to occur: rainfall (0.5), slope (0.3), soil moisture (0.1), and TWI (0.1). Because of this weighted method, we were able to add a new column to our CSV file that contains the data that came from the weighted analysis. Crucially, this novel approach is only used at the locations where our machine learning model has identified a landslide risk. By doing so, we can offer a nuanced assessment of the landslide risk, prioritizing areas based on the compounded effect of the key contributing factors. This meticulous method not only enhances the granularity of our predictions but also provides a more targeted strategy for risk mitigation.

Based on the numbers in the resulting column of our CSV dataset, we developed precise criteria to successfully classify the landslide risk categories after doing our weighted analysis. Using these standards as a reference, we may more accurately categorize each point in the compounded risk factors:

CLASSES	RANGE
Very High risk landslide	0.7 to 1.0
High risk landslide	0.5 to 0.7
Moderate risk landslide	0.3 to 0.5
low risk landslide	0.01 to 0.3

Our model consistently produces the most recent estimates of landslide risk, taking into account the most recent meteorological factors, thanks to this unrelenting and automatic cycle. Such a method not only demonstrates our dedication to offering precise and timely landslide forecasts, but it also serves as an excellent example of how dynamic data may be fed into predictive models. Our predictive model's utility and dependability are greatly increased by automating the data collection, analysis, and prediction updating processes. This makes the model an invaluable tool for efforts to prepare for and mitigate disasters.

2.4 Analysis

This project also includes analysis which can have a significant impact on society. The following analysis can greatly benefit disaster management authorities and people in that area as they can be aware of the danger zones before the incident. In case of any emergency, this platform aims to provide full guidance to handle the situation wisely like nearest safest areas. The analyses performed in this project is as follows.

2.4.1 Exposure analysis:

This model consistently produces the most recent estimates of landslide risk, considering the most recent meteorological factors, because of this unrelenting and automatic cycle. Such a method not only demonstrates our dedication to offering precise and timely landslide forecasts, but it also serves as an excellent example of how dynamic data may be fed into predictive models (Smith et al., 2020). Our predictive model's utility and dependability are greatly increased by automating the data collection, analysis, and prediction updating processes (Smith et al., 2020). This makes the model an invaluable tool for efforts to prepare for and mitigate disasters (Smith et al., 2020). The maintenance of the accuracy and relevance of our analysis is greatly dependent on the automation of this procedure. Our technology is configured to carry out the exposure analysis automatically each time the backend data is modified. Our predictive model and the assessments that follow are guaranteed to be up to date with the most recent environmental conditions and infrastructure advancements because of our real-time data integration capability (Smith et al., 2020). Stakeholders can lessen the overall impact of a landslide on the community by prioritizing steps to save lives, protect infrastructure, and reduce economic damages by identifying which sectors of society are most at risk (Smith et al., 2020).

2.4.2 Road Impact Analysis:

The research that follows focuses on the relationship between landslides and road networks in particular, considering the strategic significance of roadways for both routine traffic and emergency evacuations (Johnson et al., 2021). This research is necessary to identify the road segments that are most vulnerable to landslides, which might seriously jeopardize traveler safety and interrupt vital supply routes (Johnson et al., 2021). This cutting-edge function is intended to locate and highlight landslide areas that cross or are near road networks (Johnson et al., 2021). These notifications give travelers vital real-time information about possible landslide hazards along their routes, empowering them to make well-informed decisions regarding the safety of their travels and route modifications (Johnson et al., 2021). A pre-disaster warning will be sent to all relevant parties, such as local governments, emergency services, and road maintenance teams, so they have ample time to lessen the damage (Johnson et al., 2021). We are also able to identify specific road segments that are susceptible to landslide-related obstructions or damage by combining our landslide risk projections with comprehensive road network data (Johnson et al., 2021). By keeping the most important routes available and safe for use, this precise identification helps to prioritize landslide mitigation and road maintenance (Johnson et al., 2021). Our goal with Road Impact Analysis is to reduce the risk to human life by facilitating quick and efficient reaction plans. This preemptive strategy makes a substantial contribution to preserving transportation continuity and minimizing any potential negative effects on nearby and surrounding towns (Johnson et al., 2021).

2.5 Frontend

To further utilize state-of-the-art technology in our landslide risk management approach, we have created an advanced online platform to present the findings of our investigation. JavaScript, HTML, and CSS were used in the front-end development process to produce an interface that is easy to use and intuitive. Regardless of a user's level of technical expertise, our platform is accessible and simple to use thanks to the selection of technologies.

The front-end design of our main page clearly communicates our purpose to improve landslide risk management through technological innovation by showcasing our project's detailed information and objectives.

We used the Cesium and Mapbox mapping systems to improve the visual representation of our spatial data (Johnson et al., 2020). These robust tools give users a thorough understanding of the landslide risk areas in connection to the road networks by enabling them to explore and interact with our findings in a dynamic 3D and 2D geographical context (Johnson et al., 2020). Cesium gives our data presentation greater depth and allows for a more thorough comprehension of the topography and possible landslide effects. On the other hand, Mapbox provides high-resolution mapping features that clarify geographic information system (GIS) data and help identify precise areas of interest (Johnson et al., 2020).

To improve the user experience overall, we used Bootstrap in the creation of our platform. Our platform is now aesthetically beautiful and usable on the web. In addition to increasing accessibility, this usage of Bootstrap helps to make our landslide risk management solutions as useful and easy to use as feasible for all parties concerned (Johnson et al., 2020).

2.6 Backend

We used a novel strategy to further improve our landslide risk management system by creating a thorough methodology that makes use of Python's processing capabilities on the backend. The `backend.py` file, which was painstakingly created to contain all automated operations essential to our system's operation, is the result of this methodology. This file's code is assigned to a number of important functions, the first of which is data preparation, an essential step in guaranteeing the quality and preparedness of the data for analysis. After that, it methodically makes data fetching from API easier, which is in line with our system's requirement to get the most recent and pertinent geographic and risk factor data.

The next steps are model training and prediction, where our machine learning algorithms are refined and used to anticipate the likelihood of landslides more accurately. Subsequently, the data is categorized and thoroughly analyzed, which allows us to extract valuable insights and practical knowledge from the predictive models. To enhance the real-time evaluation capabilities of our system, we have included a Flask scheduler that retrieves rainfall data automatically every three hours. This addition is essential to our ongoing efforts to update our CSV data with the most recent information affecting the risks of landslides.

Our Python-based architecture's lightweight yet effective Flask scheduler is made to carry out this work accurately and effectively. We can automate the process of data collection and updating by incorporating this scheduler into our system, which guarantees that the most recent weather information is always incorporated into our machine learning model and subsequent analyses. In this continuous automation process, our backend pipeline is called by the Flask scheduler, which automates every stage of the process—from preprocessing and data fetching to model prediction and data analysis—without requiring human input.

The `backend.py` file efficiently interacts with our server architecture after these procedures are finished, guaranteeing that the processed data and insights are correctly shown on the front end. The server can dynamically display the results on our online platform thanks to this seamless connectivity, giving users an extremely engaging and straightforward way to interact with and study the nuances of landslide risk zones.

2.7 Features

Apart from the prediction in our study area, Astore, this project also comprises some key features which can be very beneficial for users and concerned organizations. In detail description of these features are as follows.

2.7.1 User Input System

We have created a specialized gateway with an extensive pipeline. Users can upload a single point shapefile containing the precise points within the study region together with a TIFF file that contains all fifteen factors of the study area they have selected. The painstaking design of our backend infrastructure enables it to automatically retrieve the datasets that correspond to the coordinates specified in the shapefile. When the retrieval is successful, the information is assembled into a CSV file. After that, our machine learning (ML) model that has been trained processes this CSV file with ease. When the results are produced, our technology converts the extensive data that was compiled into the CSV file into a GeoJSON format. This conversion is a crucial step that allows our analytical outputs to be seamlessly integrated with the Cesium platform for front-end visual display. GeoJSON is a commonly used encoding format for various geographic data structures that makes it easier to express simple geographical features and their non-spatial characteristics. This makes it possible for stakeholders to visually explore the landslide risk zones in a dynamic and detailed landscape, resulting in a more engaging and interactive user experience. The dynamic creation of results depending on the users' individual geographic characteristics of interest is made possible by the integration of this procedure. This cutting-edge site offers a more customized evaluation by optimizing the user experience while also greatly improving the accuracy and applicability of the landslide risk analysis for each user.

2.7.2 Red Alert System

The purpose of the Red Alert System is to improve our capacity to control landslide risks. The core component of this system is a secure login site that is only available to administrators, guaranteeing that alert management and distribution are dependable and controlled. In the case that a high-risk landslide is detected, this system is outfitted with an automatic mechanism that notifies the appropriate departments, emergency services, and non-governmental organizations (NGOs) by email. As long as the risk level stays high, these alerts are sent out every three hours, so everyone who needs to know is always aware of the most recent information.

The technology gives managers the ability to manually send targeted emails to particular people or groups in addition to automated notifications. This particular feature proves to be especially advantageous in situations where prompt human intervention is necessary to mitigate or convey particular hazards that the automated criteria might have overlooked.

The incorporation of a dynamic mapping interface into our emergency alert system is another essential feature. This tool makes it easy for administrators and stakeholders to identify areas of concern by visualize highlighted regions that have been assessed as having an urgent landslide danger. These map overlays provide an interactive way to keep an eye on possible disaster sites, integrating real-time data to help with resource allocation and intervention strategy preparation.

By putting an all-inclusive Emergency Alert System into place, we hope to greatly increase the efficiency of our landslide risk control initiatives. Through the implementation of prompt and precise communication protocols among administrators, emergency responders, and pertinent stakeholders, we can collaboratively endeavor to mitigate the consequences of landslides on impacted communities, thereby protecting human lives and vital infrastructure.

2.7.3 3-D Animation

Our portal offers a unique feature that is not to be missed: a three-dimensional video display of the locations of landslides, in addition to dynamic and interactive maps. With its entire 360-degree view of neighboring societies and landscapes, this state-of-the-art visualization tool offers users an immersive experience that improves comprehension of the surrounding area. This function will be especially helpful to analysts since it enables a thorough analysis of the surrounding geological and environmental factors. Our goal is to provide cutting-edge tools for landslide risk assessment and management, and by incorporating this 3D video capability, we can ensure that both professionals and casual users may have a comprehensive understanding of the potential impact zones. This feature not only enriches the user experience but also aids in the detailed analysis required for effective landslide risk management and planning.

CHAPTER 3

RESULTS AND END PRODUCT

For landslide susceptibility analysis and prediction, 15 factors were identified that significantly contribute to the occurrence of landslides. Out of these factors, 4 factors i.e., Rainfall, Slope, Soil moisture, and TWI, were carried out as the most contributing factors which were used in the classification of landslides. Moreover, 73 landslide and 73 non-landslide points were identified which were used in the training of machine learning model. 1000 points were randomly chosen which represents our whole study area. According to our latest results, following points are classified as danger and non-danger zones.

CLASSES	NO. OF POINTS
Very High Risk	3
High Risk	3
Medium Risk	20
Low Risk	76
No Risk	898

3.1 Machine Learning Model

Three major algorithms were tried in this project to prepare the machine learning model. The decision tree algorithm resulted in an accuracy of 78%, the random forest algorithm resulted in an accuracy of 81%, and logistic regression resulted in an accuracy of 70%. The random forest algorithm was selected due to its higher accuracy. However, the accuracy of the algorithms varies according to different regions because of different topographic conditions.

ALGORITHM	ACCURACY
Random Forest	81.08%
Decision Tree	78.38%
Logistic Regression	70.27%

3.2 Web Platform

Our web platform stands because of our research and development efforts, showcasing the practical application of our innovative technologies in the real world. With its intuitive design and user-friendly interface, the platform enables users to visualize and understand the intricacies of landslide risk in their specific contexts.

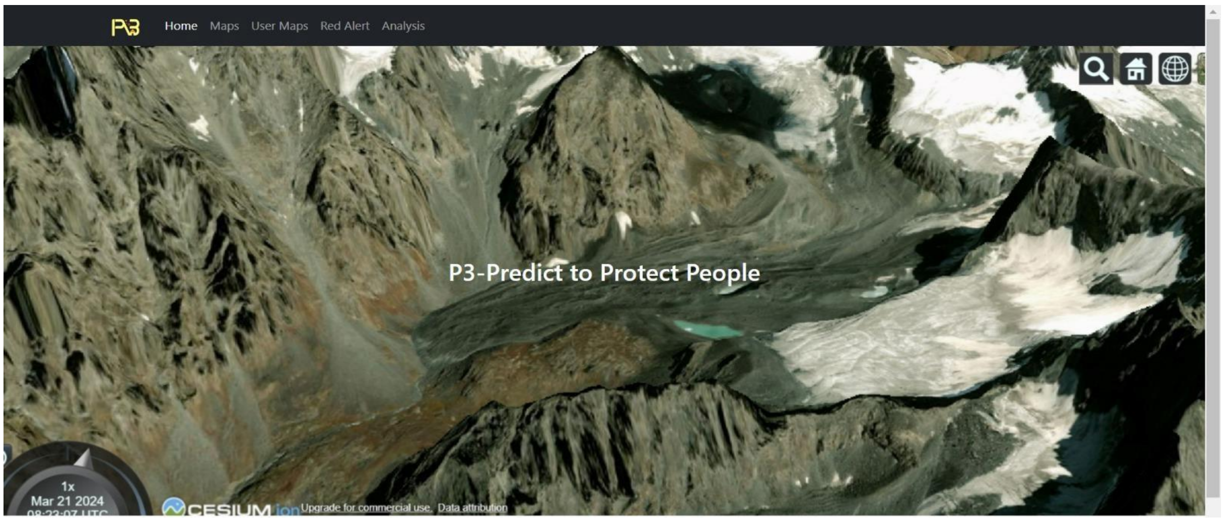

Fig. 3.1. User Interface

Aerial view of the classified landslide locations are depicted where very high risk to low risk locations are shown

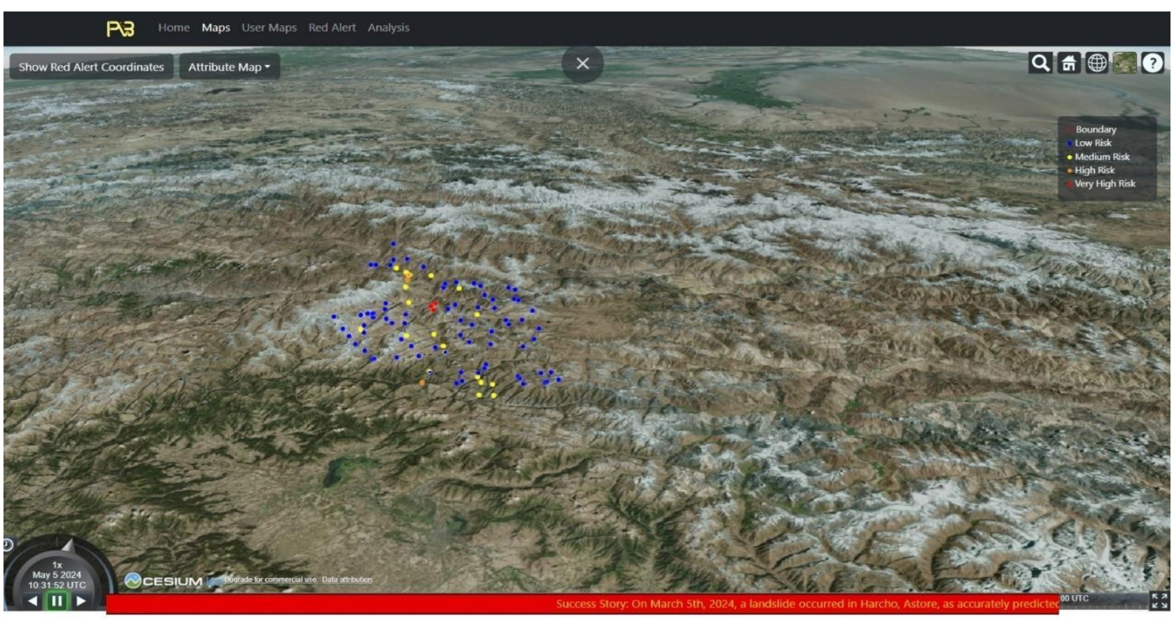

Fig. 3.2. Landslide susceptibility map

A tool is designed to provide immediate warnings about very high risk detected locations. Emails will be sent to relevant authorities which can take immediate action to avoid any inconvenience.

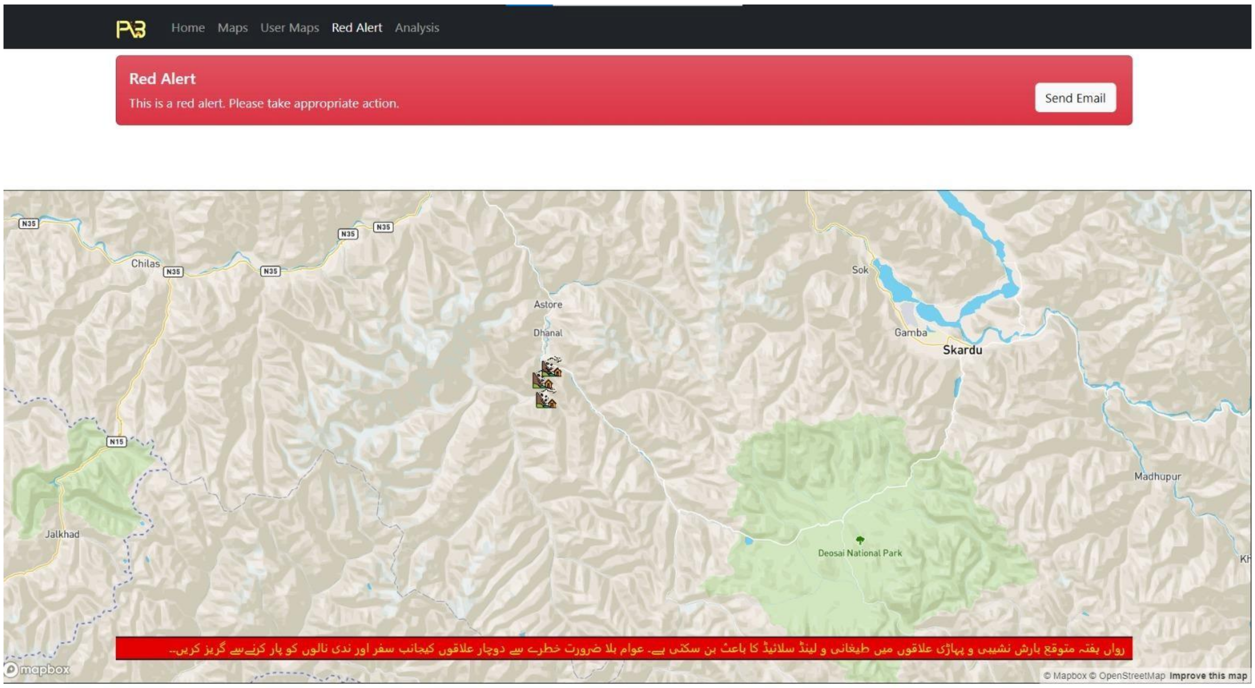

Fig. 3.3. Red Alert System

Another service provided is to visualize the 360 view landslides prone locations in order to have better understanding of these high risk locations to manage the mitigation.

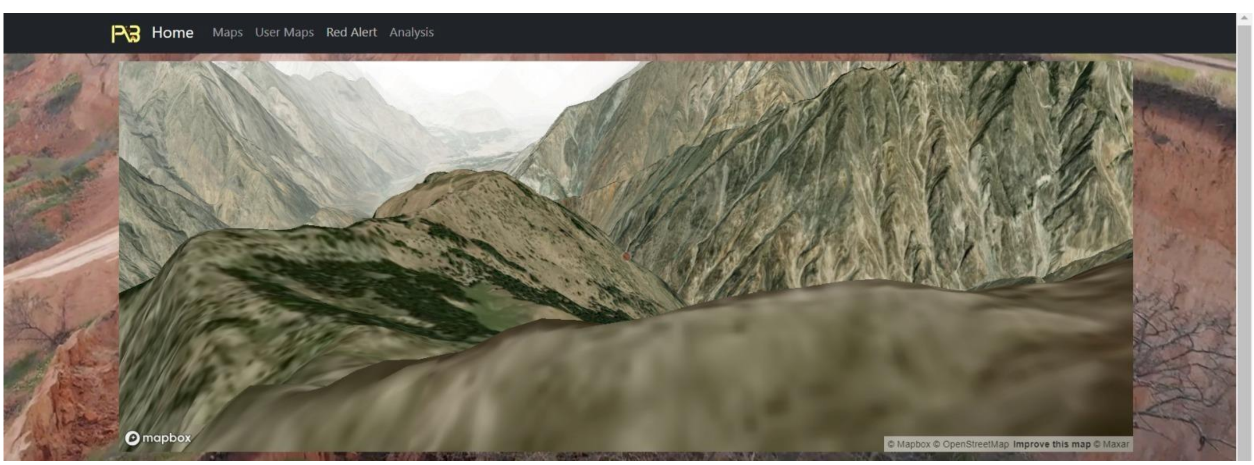

Fig. 3.4. 3-D Animation

Travelers and residents can arrange and manage their routes considering the safer route through this application

Fig. 3.5. Terrain and roads analysis

CHAPTER 4

CONCLUSION AND RECOMMENDATION

The development of machine learning model for prediction of landslides is done by using various contributing factors along with real-time forecasted rainfall as our triggering factor. Relevant authorities will be informed prior to any incident to ensure the safety of life in the region. Users can visualize 3D view of high-risk locations and can manage their routes keeping in view safety of life. Users can also use their respective datasets for prediction according to their area of interest. Data is a necessary part of this project, and the unavailability of data may lead to wrong predictions. Acquisition of the datasets was the most challenging part of the project, which needs to be addressed in future to further improve the tools and technologies in this field.

REFERENCES

1. Chen, W., & Li, Y. (2020). GIS-based evaluation of landslide susceptibility using hybrid computational intelligence models. Catena, 195, 104777.

2. Marjanović, M., Kovačević, M., Bajat, B., & Voženílek, V. (2011). Landslide susceptibility assessment using SVM machine learning algorithm. Engineering Geology, 123(3), 225-234.

3. Afzal, N., Ahmad, A., Shirazi, S. A., & Younes, I. (2022). GIS-based landslide susceptibility mapping using analytical hierarchy process: a case study of Astore region, Pakistan. EQAInternational Journal of Environmental Quality, 48, 27-40.

4. Farooq Ahmed, M., & David Rogers, J. (2016). Regional level landslide inventory maps of the Shyok River watershed, Northern Pakistan. Bulletin of Engineering Geology and the Environment, 75, 563-574.

5. Pham, Q. B., Chandra Pal, S., Chakrabortty, R., Saha, A., Janizadeh, S., Ahmadi, K., ... & Bannari, A. (2022). Predicting landslide susceptibility based on decision tree machine learning models under climate and land use changes. Geocarto International, 37(25), 7881-7907.

6. Liu, S., Wang, L., Zhang, W., He, Y., & Pijush, S. (2023). A comprehensive review of machine learning-based methods in landslide susceptibility mapping. Geological Journal, 58(6), 22832301.

7. Khalil, U., Imtiaz, I., Aslam, B., Ullah, I., Tariq, A., & Qin, S. (2022). Comparative analysis of machine learning and multi-criteria decision making techniques for landslide susceptibility mapping of Muzaffarabad district. Frontiers in Environmental Science, 10.

8. Yuan, R., & Chen, J. (2022). A hybrid deep learning method for landslide susceptibility analysis with the application of InSAR data. Natural Hazards, 114(2), 1393-1426.

9. Al-Najjar, H. A., & Pradhan, B. (2021). Spatial landslide susceptibility assessment using machine learning techniques assisted by additional data created with generative adversarial networks. Geoscience Frontiers, 12(2), 625-637.

10. Chen, W., Chen, X., Peng, J., Panahi, M., & Lee, S. (2021). Landslide susceptibility modeling based on ANFIS with teaching-learning-based optimization and Satin bowerbird optimizer. Geoscience Frontiers, 12(1), 93-107.

11. Youssef, A. M., & Pourghasemi, H. R. (2021). Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region, Saudi Arabia. Geoscience Frontiers, 12(2), 639-655.

12. Sadiki, M., Aghad, M., Batchi, M., & Al Karkouri, J. (2022). Landslide susceptibility mapping using machine learning techniques and comparison of their performance at Ziz upper watershed, Southeastern Morocco.

13. Ali, S. A., Parvin, F., Vojteková, J., Costache, R., Linh, N. T. T., Pham, Q. B., ... & Ghorbani, M. A. (2021). GIS-based landslide susceptibility modeling: A comparison between fuzzy multi-criteria and machine learning algorithms. Geoscience Frontiers, 12(2), 857-876.

14. Azarafza, M., Azarafza, M., Akgün, H., Atkinson, P. M., & Derakhshani, R. (2021). Deep learning-based landslide susceptibility mapping. Scientific reports, 11(1), 24112.

15. Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D. T., ... & Abderrahmane, B. (2020). Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. Earth-Science Reviews, 207, 103225.

16. Fang, Z., Wang, Y., Peng, L., & Hong, H. (2020). Integration of convolutional neural network and conventional machine learning classifiers for landslide susceptibility mapping. Computers & Geosciences, 139, 104470.

17. Chang, K. T., Merghadi, A., Yunus, A. P., Pham, B. T., & Dou, J. (2019). Evaluating scale effects of topographic variables in landslide susceptibility models using GIS-based machine learning techniques. Scientific reports, 9(1), 12296.

18. Chen, W., Peng, J., Hong, H., Shahabi, H., Pradhan, B., Liu, J., ... & Duan, Z. (2018). Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China. Science of the total environment, 626, 1121-1135.

19. Pham, B. T., Pradhan, B., Bui, D. T., Prakash, I., & Dholakia, M. B. (2016). A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India). Environmental Modelling & Software, 84, 240-250.

20. Goetz, J. N., Brenning, A., Petschko, H., & Leopold, P. (2015). Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. Computers & geosciences, 81, 1-11.

21. Korup, O., & Stolle, A. (2014). Landslide prediction from machine learning. Geology today, 30(1), 26-33.

22. Tien Bui, D., Shahabi, H., Omidvar, E., Shirzadi, A., Geertsema, M., Clague, J. J., ... & Lee, S. (2019). Shallow landslide prediction using a novel hybrid functional machine learning algorithm. Remote Sensing, 11(8), 931.

23. Kuradusenge, M., Kumaran, S., & Zennaro, M. (2020). Rainfall-induced landslide prediction using machine learning models: The case of Ngororero District, Rwanda. International journal of environmental research and public health, 17(11), 4147.

24. Wang, H., Zhang, L., Yin, K., Luo, H., & Li, J. (2021). Landslide identification using machine learning. Geoscience Frontiers, 12(1), 351-364.

25. Kavzoglu, T., Colkesen, I., & Sahin, E. K. (2019). Machine learning techniques in landslide susceptibility mapping: a survey and a case study. Landslides: Theory, practice and modelling, 283-301.

26. Bui, D. T., Tsangaratos, P., Nguyen, V. T., Van Liem, N., & Trinh, P. T. (2020). Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment. Catena, 188, 104426.

27. Hussain, M. A., Chen, Z., Wang, R., Shah, S. U., Shoaib, M., Ali, N., ... & Ma, C. (2022). Landslide susceptibility mapping using machine learning algorithm. Civ. Eng. J, 8, 209-224.

28. Nguyen, V. V., Pham, B. T., Vu, B. T., Prakash, I., Jha, S., Shahabi, H., ... & Tien Bui, D. (2019). Hybrid machine learning approaches for landslide susceptibility modeling. Forests, 10(2), 157.

29. Chen, X., & Chen, W. (2021). GIS-based landslide susceptibility assessment using optimized hybrid machine learning methods. Catena, 196, 104833.

30. Tengtrairat, N., Woo, W. L., Parathai, P., Aryupong, C., Jitsangiam, P., & Rinchumphu, D. (2021). Automated landslide-risk prediction using web gis and machine learning models. Sensors, 21(13), 4620.

31. Srivastava, S., Anand, N., Sharma, S., Dhar, S., & Sinha, L. K. (2020, June). Monthly rainfall prediction using various machine learning algorithms for early warning of landslide occurrence. In 2020 International Conference for Emerging Technology (INCET) (pp. 1-7). IEEE.

32. Jones, S., Kasthurba, A. K., Bhagyanathan, A., & Binoy, B. V. (2021). Landslide susceptibility investigation for Idukki district of Kerala using regression analysis and machine learning. Arabian Journal of Geosciences, 14(10), 838.

33. Kadavi, P. R., Lee, C. W., & Lee, S. (2018). Application of ensemble-based machine learning models to landslide susceptibility mapping. Remote Sensing, 10(8), 1252.

34. Nhu, V. H., Mohammadi, A., Shahabi, H., Ahmad, B. B., Al-Ansari, N., Shirzadi, A., ... & Nguyen, H. (2020). Landslide susceptibility mapping using machine learning algorithms and remote sensing data in a tropical environment. International journal of environmental research and public health, 17(14), 4933.

35. Deng, L., Smith, A., Dixon, N., & Yuan, H. (2021). Machine learning prediction of landslide deformation behaviour using acoustic emission and rainfall measurements. Engineering Geology, 293, 106315.

36. Wang, Z., Liu, Q., & Liu, Y. (2020). Mapping landslide susceptibility using machine learning algorithms and GIS: A case study in Shexian County, Anhui Province, China. Symmetry, 12(12), 1954.

37. Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S. R., Tiede, D., & Aryal, J. (2019). Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. Remote Sensing, 11(2), 196.

38. Wang, Y., Tang, H., Huang, J., Wen, T., Ma, J., & Zhang, J. (2022). A comparative study of different machine learning methods for reservoir landslide displacement prediction. Engineering Geology, 298, 106544.

39. Tehrani, F. S., Calvello, M., Liu, Z., Zhang, L., & Lacasse, S. (2022). Machine learning and landslide studies: recent advances and applications. Natural Hazards, 114(2), 1197-1245.

40. Madawala, C. N., Kumara, B. T. G. S., & Indrathilaka, L. (2019, March). Novel machine learning ensemble approach for landslide prediction. In 2019 International Research Conference on Smart Computing and Systems Engineering (SCSE) (pp. 78-84). IEEE.

41. Pourghasemi, H. R., & Rahmati, O. (2018). Prediction of the landslide susceptibility: Which algorithm, which precision?. Catena, 162, 177-192.

42. Mohan, A., Singh, A. K., Kumar, B., & Dwivedi, R. (2021). Review on remote sensing methods for landslide detection using machine and deep learning. Transactions on Emerging Telecommunications Technologies, 32(7), e3998.

43. Liu, Z., Gilbert, G., Cepeda, J. M., Lysdahl, A. O. K., Piciullo, L., Hefre, H., & Lacasse, S. (2021). Modelling of shallow landslides with machine learning algorithms. Geoscience Frontiers, 12(1), 385-393.

44. Micheletti, N., Kanevski, M., Bai, S., Wang, J., & Hong, T. (2013). Intelligent analysis of landslide data using machine learning algorithms. Landslide Science and Practice: Volume 3: Spatial Analysis and Modelling, 161-167.

45. Ng, C. W. W., Yang, B., Liu, Z. Q., Kwan, J. S. H., & Chen, L. (2021). Spatiotemporal modelling of rainfall-induced landslides using machine learning. Landslides, 18(7), 24992514.

46. Kainthura, P., & Sharma, N. (2022). Hybrid machine learning approach for landslide prediction, Uttarakhand, India. Scientific reports, 12(1), 20101.