PROJECT REPORT

LANDSLIDE PREDICTION USING MACHINE LEARNING

TEAM 2

Aayushi Priya (20CE10002)
Abhiraj Rananajay Singh (20CE30035)
Ankush Agarwal (20CE10013)
Dheemen Pattanayak (20CE10091)
Diya A. P. Chodnekar (20CE10022)
Ikshita Agarwal (20CE30012)
Sisodiya Jitendrasingh (20CE10069)
Pradeep Kumar Kumawat (20CE10046)

INDEX

ACKNOWLEDGEMENT

We would like to express our heartfelt gratitude to Professor Baidurya Bhattacharya for providing us with the excellent opportunity to work on the project 'Landslide Prediction Using Machine Learning,' as well as for mentoring and motivating us at every step of the way. It has been an honour to be a part of such a fantastic and socially useful endeavour. We, as a team, would want to express our gratitude to everyone who assisted us with our initiative. It's been a fantastic experience for practically all of us because it was our first time working together as a on a project.

PROBLEM STATEMENT

Landslides are one of the most widespread natural disasters and are wreaking havoc on the world's economy and resources. Geospatial data is complex and undergoes drastic changes over time making it difficult to determine the landslides. Landslides can be detected with real-time data but sometimes there is not enough time to react and respond to prevent infrastructure damage and loss of life. Therefore to prevent this, an early forecasting system is necessary as if the collapse could be predicted then residents could flee to a safer place earlier and appropriate preventive measures could be taken by the authorities.

ABSTRACT

Landslides are one of the most dangerous natural disasters, so early warning system can save a lot of lives. In this report, four prediction modelling approaches were used, namely Logistic Regression, Random Forest, AdaBoost and Support Vector Machine. These approaches utilised pseudo-random generated data which included various parameters such as rainfall, cohesion, soil depth etc. A term factor of safety was introduced to determine whether a landslide occurs or not for a set of given values of the features. The final prediction is made by taking the weighted mean of all ML Models.

INTRODUCTION

Landslides are a type of "mass-wasting", which are defined as the movement of a mass of rock, debris, or earth down a slope. It may include a wide range of ground movements, such as rockfalls, deep-seated slope failures, mudflows, and debris flows.

The impact of a landslide is extensive and may lead to loss of life, destruction of infrastructure, damage to land and loss of natural resources.
They also affect the environment in a detrimental way by leading to changes in the topography of the earth's surface, character and quality of water bodies, forests and natural habitat.

ENVIRONMENTAL IMPACTS

Landslides can overwhelm, and even pollute water bodies with excess sediment which can dam the water bodies, impacting marine life, increasing the risk of floods and deterioration of water quality.

One of the major causes of landslides is rainfall. Since mountain slopes are governed by the laws of gravity, water along with debris, rocks and the materials that compose the slope move downwards, causing immense loss to both life and property. It may even cause flash floods in valleys.

Landslides are often triggered by earthquakes. If these earthquakes happen underwater, they may induce submarine landslides. Submarine landslides cause the movement of large quantities of rock and sediment and hence can threaten underwater infrastructures like pipelines and seabed cables. They even have the potential to generate tsunamis which can cause immense damage to coastal areas. [1][2]

IMPACT ON LIFE AND INFRASTRUCTURE

Landslides are annual recurring phenomena and cause substantial loss of human lives, livelihoods and seriously impacts the physical, socio-economic and cultural landscape of the country. India is the most affected country by human-triggered landslides. 2011 estimate suggested that India suffers Rs. 150-200 crore of monetary loss every year from landslides[3]

In India, 4.2 lakh km2 of the hilly region in 16 states and two Union Territories, are vulnerable to landslide hazards. Out of this,

- 0.18 million sq. km falls in North East Himalaya, including Darjeeling and Sikkim Himalaya
- 0.14 million sq. km falls in North West Himalaya (Uttarakhand, Himachal Pradesh and Jammu & Kashmir)
- 0.09 million sq. km in the Western Ghats and Konkan hills (Tamil Nadu, Kerala, Karnataka, Goa and Maharashtra)
- 0.01 million sq. km in Eastern Ghats of Aruku area in Andhra Pradesh.

[4]

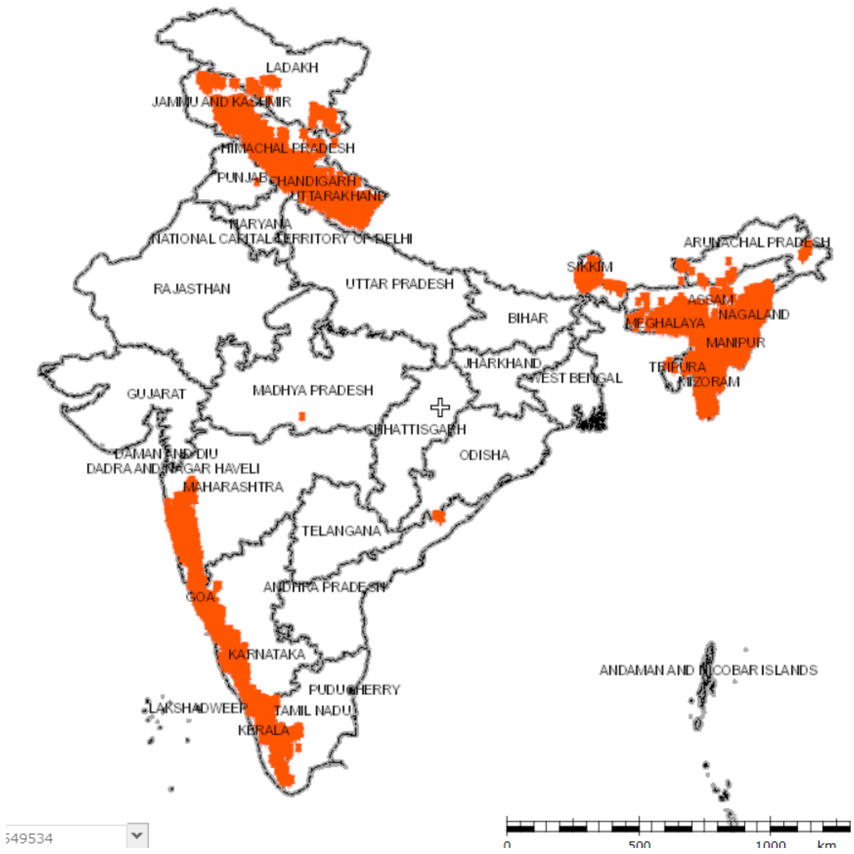

Fig 1. Bhukosh - Geographical survey of India [5]

Some of the past incidents reported:

- On July 30, 2014, a landslide occurred in the main village in the Ambegaon tehsil of Pune district. Around 150 people were feared being trapped under the debris. Five people have died as a result, according to TV reports. [6]
- A passenger train from Karnataka's Mangaluru to Mumbai was hit by a landslide in Goa on July 24, 2021. The train was derailed on the Dudhsagar-Sonaulim section in Goa with the engine and the first general running off the tracks, The Hindu newspaper reported. No passengers were injured. [7]
- In September 2021, Landslide debris blocked Badrinath Highway, damaged Vehicles and due to the blockage of the highway and the link road, the Rudraprayag and Chamoli districts are not able to receive even essential commodities. [8]

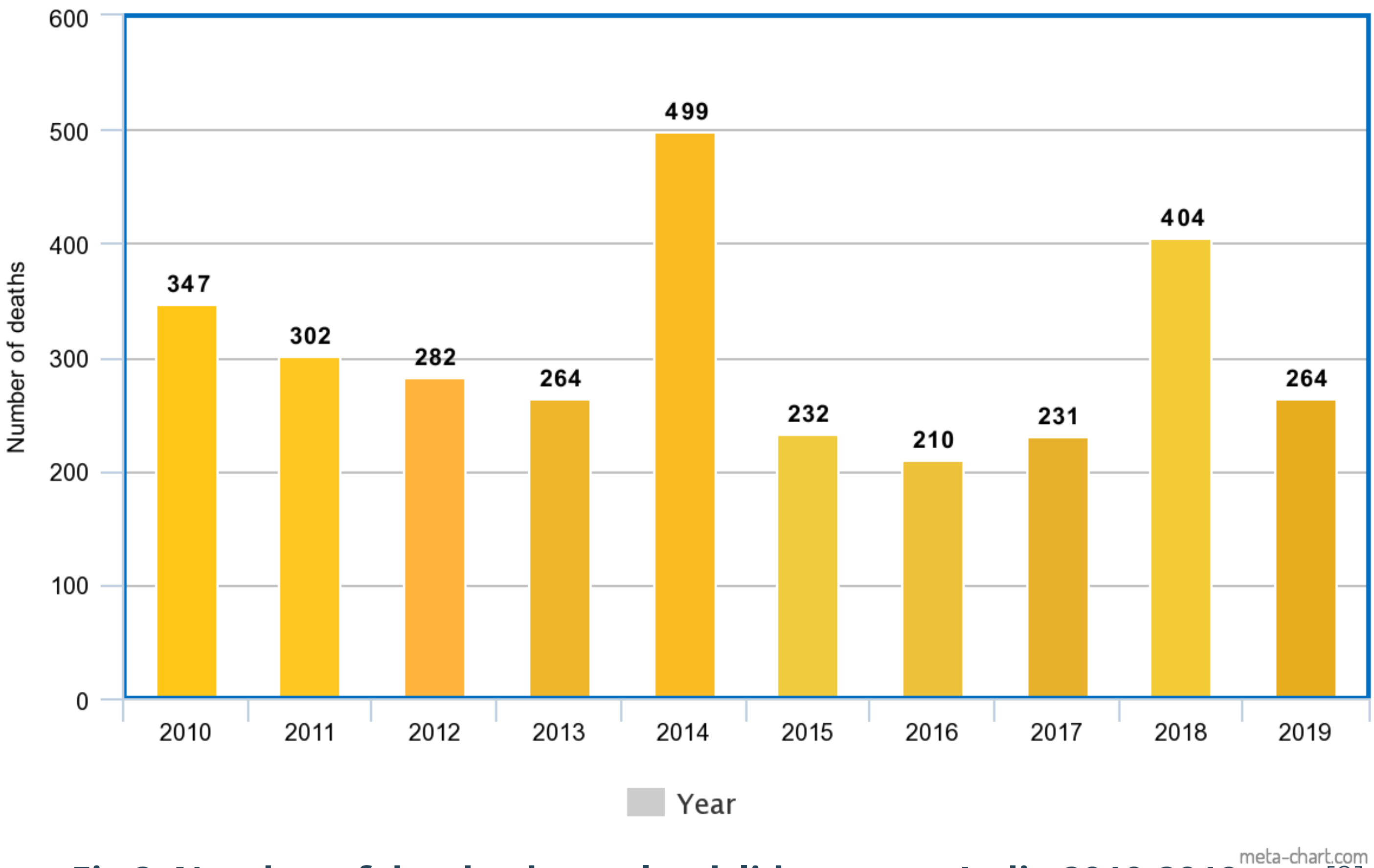

Fig 2. Number of deaths due to landslides across India 2010-2019 [9]

Therefore, it becomes necessary to have effective early warning systems so that people can take necessary measures.

CURRENT STATUS

LIn collaboration with the Geological Survey of India (GSI), the National Disaster Management Authority (NDMA) has developed a National Landslide Risk Management Strategy (NLRMS) that addresses all aspects of landslide disaster risk reduction and management, including

- hazard mapping
- monitoring and early warning systems
- awareness programmes.
- preparation of mountain zone legislation and policies
- capacity building and training
- The development of an SPV (special purpose vehicle) for landslide management, including stabilisation and mitigation.[10]

GSI has recently completed research on the creation of more advanced tools and procedures for landslide risk assessment, including

- the establishment of a regional landslide early warning system and debris flow modelling.
- utilising satellite data to measure ground deformation
- investigating landslides with cutting-edge ground-based instruments, etc.

By 2025, GSI plans to use regional landslide forecasting in an operational mode in other vulnerable districts of the country, and it has a plan to build a state-of-the-art landslide warning centre at GSI's GHRM (Geohazard Research and Management) Centre in Kolkata, similar to the Tsunami warning centre of the Indian National Centre for Ocean Information Services (INCOIS), Ministry of Earth Sciences (MoES) in Hyderabad.

At various scales, landslide susceptibility maps are being created in order to extract information on hill landslide vulnerability that may be used for regional and specific land use planning. Some of its maps, such as the meso-scale landslide susceptibility map of Nainital, Uttarakhand, and the macro-scale landslide susceptibility map of Nilgiri District, Tamil Nadu, are used in the respective landslide disaster management plans. The Tamil Nadu Disaster Management Authority uses the NLSM data product in TN-SMART, a mobile app that distributes alerts and is linked to an alarm system to all registered members.

On September 7, 2018, GSI installed a people-centric Landslide Early Warning System on an experimental basis at Giddapahar Village in Kurseong Block, Darjeeling District, West Bengal for the first time in India. It was a collaborative endeavour between GSI and the West Bengal government. The project's goal was to create an early warning system in which the community plays a major part in landslide management.

PROBLEMS WITH CURRENT TECHNOLOGIES

The National Landslide Susceptibility Map(NLSM) of GSI divides areas into three zones-high, moderate and low- based on how likely they are to experience landslides

CHALLENGES:

- NLSM cannot predict the time of occurrence of a landslide.
- It also cannot predict location. Only displays landslides prone areas
- The susceptibility of a particular geographic area may also change over a period of time depending on factors like land-use patterns, rainfall patterns and deforestation. Real-time prediction is difficult.

EARLY-WARNING SYSTEMS

- SITE-SPECIFIC: For identification of an exact slope we gather data related to that slope but Capital and human resource intensive. Cannot be extrapolated to wider areas.

- REGION-SPECIFIC: Conducted over a wide geographical area. Formulated based on the correlation between rainfall and landslides. A prototype was developed in 2020. Model is being used in at least 26 places worldwide. successful implementation after 10-12 years of experimenting.

- CHALLENGES: Lack of data with regards to the time of landslides and rainfall in hilly and remote areas. No time for intensive experimentation. Prediction of weather events that trigger a landslide

CHALLENGES WITH CURRENT MACHINE LEARNING MODELS

- Each area has a different type of dataset, which therefore might not give an accurate result.
- Interference in learning between two or more classes in the datasets, which results in a low true positive rate.
- The problem of determining when to re-train the models.

DATA ACQUIRED

The data utilized in this project is of Akpa (Kinnaur), Himachal Pradesh. The features considered were the depth of the soil layer (m), density of soil (kg/m^3), density of soil water (kg/m^3), internal angle of friction of the soil (°), slope gradient (°), rainfall (mm), cohesion of soil (N/sq. m) and water table depth (m).

The minimum and maximum values of these features were obtained from research papers and climate API. Further using these values, pseudo-random numbers forming a normal distribution were generated.[11]

Table 1: Minimum and maximum values of the factors/features considered

FEATURE	MINIMUM	MAXIMUM
Soil depth (m)	0.7568	8.4166
Soil density (kg/m^3)	1629.354	1640.216
Angle of friction (°)	34.89037	43.61002
Rainfall (mm)	0.01747	7.9475
Cohesion (N/m^2)	24971.2	28595.96
Gradient (°)	42.74673	51.61566
Soil water density (kg/m^3)	993.8053	1003.477
Water table depth (m)	7.4112	16.5495

CASE STUDY

LANDSLIDE COMMUNITY MODEL(DARJEELING)

Since 2016, the area has been investigated using modern InSAR technology, including installing five corner reflectors, rainfall threshold analysis for landslide initiation within the Kurseong area, landslide awareness programmes, and developing a People-centric Landslide Early Warning System. It is backed up by GSI's long-term study of rainfall thresholds, landslide vulnerability mapping, and surface deformation using a radar-based approach. In 2016 and 2018, GSI worked with the Darjeeling District Administration to offer community members risk information through a series of community-based landslide awareness programmes.

A tipping bucket rain gauge has been installed in Giddapahar village. A village resource person, a Gram Panchayat staff member, has been assigned the task of monitoring daily rainfall at predetermined intervals and reporting the results to the local BDMO (Block Disaster Management Officer) /DDMC (District Disaster Management Centre) for documentation. GSI has instructed the in-charge on estimating rainfall thresholds, and he has been given a threshold chart on rainfall requirements for "ALERT" and "ALARM" warnings. When there is an IMD alert for heavy rainfall and the frequency of measuring the data increases than on normal days, he compares 3-day preceding rainfall with the measured event day rainfall to decide on threshold crossover for the warning. Two-way communication is required for dissemination and communication: one between the in-charge and local authorities (BDMO), and the other between the in-charge and the community at risk (three continuous whistles with pause). The BDMO can be reached by phone, text, or WhatsApp group, with the message text "ALERT landslide warning issued." If the threshold is surpassed, he activates the "alert" (continuous whistling and texts). Villagers are evacuated to safe shelters at strategic locations when the "alert" signal is activated. [12]

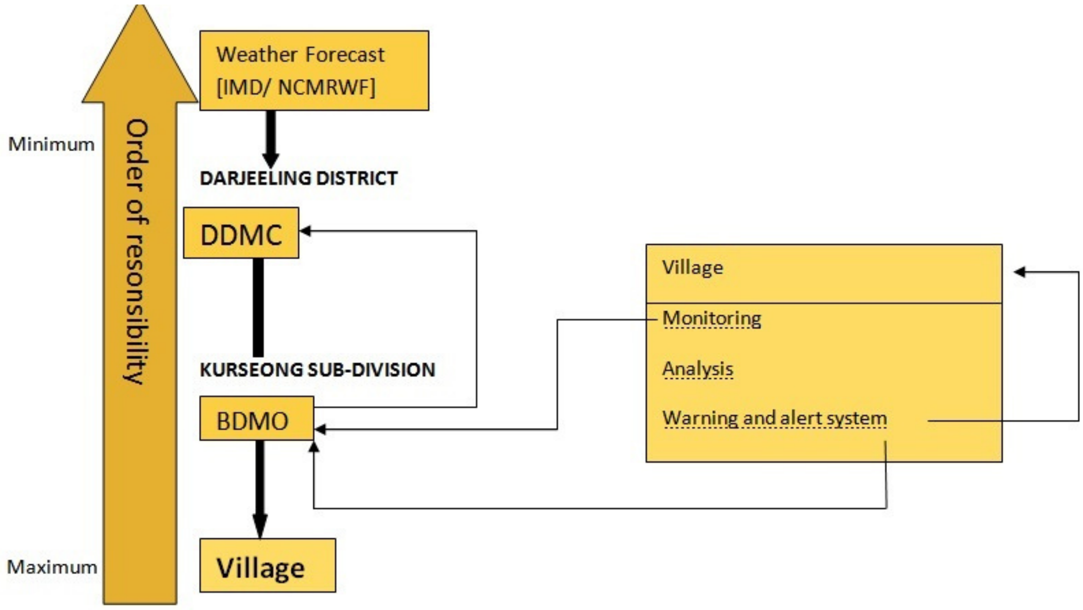

Fig 3. Working of People-centric Landslide Early Warning System

Every year, rainfall-induced landslides that were shallow and fast moving claimed the lives of several people. The "top-down" method to making L-EWS successful in such instances is not always practicable, thus GSI proposed a "bottom-up" approach to early warning, including active participation of local people.

MAPS

To better understand the objectives of this project, the topography and landslide causing factors of Darjeeling were analyzed.

The necessary data for this analysis were collected from various sources. These include collecting relevant landslide point data from Bhukosh- Geological survey of India, DEM data from Bhuvan- Indian geo-platform of ISRO and district shape file and Rainfall data from India Meteorological Department website.
Using this data and ARCGIS PRO, the following raster maps of various parameters effecting landslides like aspect, slope, curvature, elevation etc. were prepared:

<u>Elevation</u>: Height above or below a fixed reference point(Earth's sea level)
<u>Aspect</u>: The compass direction or azimuth that a terrain surface faces
<u>Slope</u>: One of the most important features, it determines the steepness of the land
<u>Curvature</u>: Displays the shape of slope, a concave or convex portion of a surface can exist.
<u>Profile curvature</u>: It affects the acceleration and deceleration of flow
<u>Plan curvature</u>: Perpendicular to the slope, it affects convergence and divergence of flow across the surface

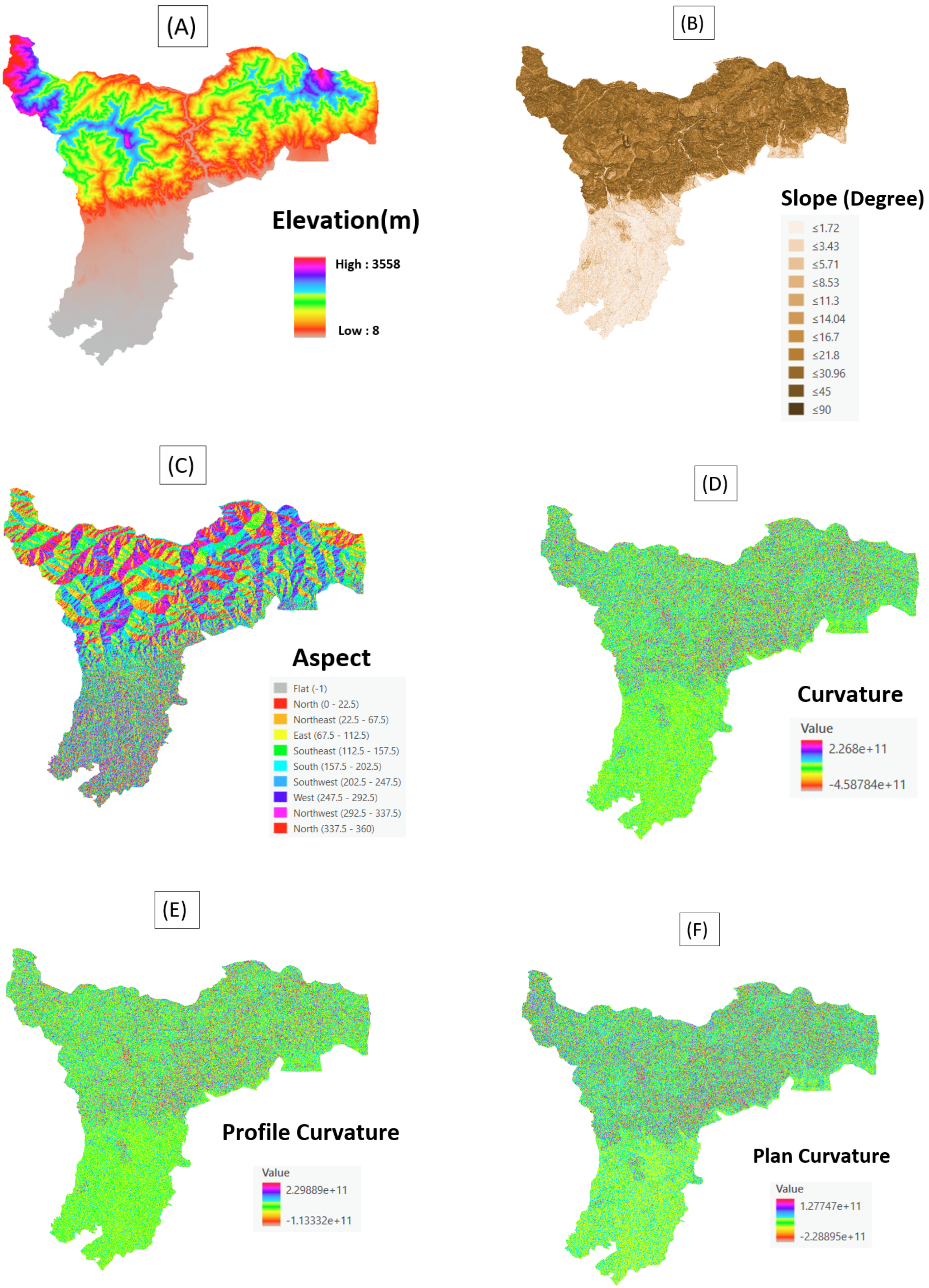

Fig. 4: Raster Maps : - (A): DEM , (B): Slope , (C): Aspect , (D): Curvature , (E): Profile Curvature (F): Plan Curvature

LANDSLIDES

FACTORS AFFECTING SLOPE STABILITY [13]

- **Material involved:**
  - Material properties (cohesion and the internal friction)
  - Fracture density and quality (Weathering of the material)

Following heavy rains, recent volcaniclastic material, for example, may become exceedingly unstable and collapse into debris flows and lahars. A hard and compact rock, such as complete gneiss, on the other hand, is typically quite stable.

- **The geometry of material and slope angle:**

Layers of rocks dipping toward the slope are particularly unstable. The slope angle is another important variable.

- **The distribution of weight along the slope:**

Loading the top of a slope can have a significant impact on its stability. Similarly, reducing the slope at its base reduces the buttressing of the lower layers underneath it and encourages sliding.

- **Water Content:**

One of the most major instability variables is water. In soils, it reduces cohesiveness, while in granular media, it increases weight and pore water pressure. It's also possible that the velocity at which water seeps through the slope is crucial. Some slopes can become unstable if small amounts of water penetrate quickly, while others are more sensitive to the amount of water that falls over a lengthy period of time. Rapid flows, which occur in many parts of the world where rock is covered by a thick layer of soil, are much more striking. Several landslides may occur at the same time after heavy rain, resulting in a typical desolate landscape.

- **Vegetation:**

Vegetation may influence stability through mechanical cohesion and removal of water via evapotranspiration.

- **External Impulsive Forces:**

External impulsive forces such as earthquakes, waves, and volcanic eruptions.

FACTOR OF SAFETY (FOS)

The factor of safety of a slope is a ratio of resisting forces to driving forces that describes the slope's stability. A slope that has a factor of safety greater than one is said to be stable. The slope angle, friction, cohesion, and water content are all taken into account when calculating FOS. The factor of safety falls as the water content and slope angle increase. Increasing friction and cohesion enhances strength and, as a result, the safety factor. In homogeneous soils with no favoured weak layer for failure, the sliding plane computation of the factor of safety cannot be used. In homogeneous soils, the failure surface is sub-spherical, resulting in a rotating slide.[15]

The equation of factor of safety (FOS) can be given by:

$$FOS = \frac{c + \cos^2\theta[\rho_r g(D - D_w) + (\rho_r - \rho_w)gD_w]\tan\phi}{D\rho_r g \sin\theta\cos\theta} \quad (1)$$

where,

c : cohesion of soil (N/sq. m)

ρ_r : density of soil (kg/m^3)

ρ_w : density of soil water (kg/m^3)

D : depth of the soil layer (m)

D_w : water table depth (m)

θ : slope gradient (°)

ϕ : internal angle of friction of the soil (°)

g : acceleration due to gravity (m/s^2)

OUR SOLUTION

GENERATION OF PSEUDO-RANDOM DATA

One of the challenges faced was acquisition of data. Hence to overcome this, pseudo-random data based on real data was generated to train and validate the ML models. Using the minimum and maximum values of each feature, computer-generated random numbers fitting a normal distribution were generated. The procedure is as follows:

- For a given feature compute a factor,
  f = (Maximum value - Minimum value) / (No. of samples)
- Using this factor, generate (N+1) samples between the minimum and maximum values having a linear nature.
- Compute the mean and standard deviation of these (N+1) samples.
- Finally, to draw the pseudo-random data from a normal distribution, use the function *numpy.random.normal(loc, scale, size)* from the NumPy library which take three primary parameters as input:
  loc : mean of the distribution
  scale : Standard deviation of the distribution
  size : size and shape of the output array

Therefore the calculated mean and standard deviation are given as input to the NumPy function to get the final data to be used.

FACTOR OF SAFETY (FOS)

To generate the target variable, i.e., whether a landslide occurs or not for a set of given values of the features, factor of safety(FOS) was used.
If the value of FOS is greater than one, the value of the target label is one, implying a landslide would occur. But if FOS is lesser than one, the value of the target label is zero, implying a landslide would not occur. [15]

$$\text{Target label} = \begin{cases} 1 \quad ; \quad \text{FOS} > 1 \\ 0 \quad ; \quad \text{FOS} < 1 \end{cases} \tag{2}$$

ML ALGORITHMS

LOGISTIC REGRESSION

Logistic Regression is a statistical model similar to Linear Regression where it uses independent factors to predict the dependent variable, but the dependent variable must be categorical. The dependent variable will always be categorical, regardless of whether the independent factors are numeric or categorical. Logistic regression can be used to classify objects into binary or multi-class categories. The conditional probability is modelled using the Logistic function (also known as logit).

For example, in binary regression, we calculate the conditional probability of the dependent variable Y, given independent variable X. It can be written as P(Y=1/X) or P(Y=0/X) which is read as the conditional probability of Y=1, given X or conditional probability of Y=0, given X. [18]

RANDOM FOREST

Random forest is a tree-based strategy that combines a large number of decision trees to generate a single consensus forecast, resulting in dependable and good prediction performance. The main feature of random forest is that a majority of the available features are not allowed to be considered at each tree split (Friedman et al., 2001). For instance, a popular method for determining the number of splits is to:

$$n' = sq.\ root\ (n) \tag{3}$$

where n is the number of predictors at each split and n is the number of all predictors [16]

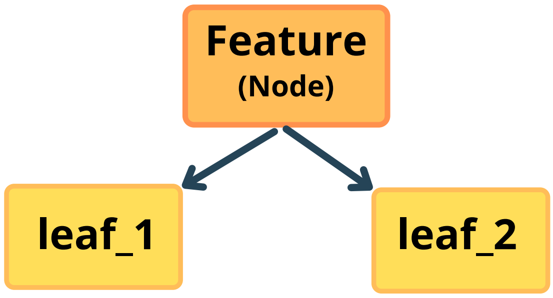

Fig 5. Decision tree stump

ADABOOST

AdaBoost algorithm, short for Adaptive Boosting, is a Boosting technique in machine learning wherein the weights are re-assigned to each example, with higher weights assigned to incorrectly classified example.

Given a dataset with N examples, each example is assigned a sample weight equal to (1/N). The algorithm creates the same number of decision tree stumps (learners) as the number of features, where a stump is one node with two leaves (Fig.5). Out of these learners the algorithm selects only one by computing their Gini and Entropy indices and choosing the one with least value. Then, the performance of the stump (as shown in (4)) is calculated using total error which is defined as the sum of all the errors in the classified example for sample weights. Next, the weights of the wrongly classified examples are increased and that of correctly classified are decreased (as shown in (5)) and then are passed to the next learner.

Finally, after passing through all the learners, the output from the learners having the majority of votes is chosen as the final output. [17]

$$\text{Performance of stump} = 0.5*\ln([1-TE] / TE) \quad ; \text{where } TE : \text{Total error} \tag{4}$$

$$\text{New Sample Weight} = \begin{cases} \text{Sample Weight} * e\hat{\ }(\text{Performance}) & ; \text{if wrongly classified} \\ \text{Sample Weight} * e\hat{\ }(-\text{Performance}) & ; \text{if correctly classified} \end{cases} \tag{5}$$

SUPPORT VECTOR MACHINE

SVMs (short for Support Vector Machines) are machine learning algorithms that are used for classification and regression. SVMs are a type of machine learning method that may be used for classification, regression, and outlier detection. SVM classifiers create a model that allocates new data points to one of the predetermined categories. As a result, it may be considered a binary linear non-probabilistic classifier.

SVMs may be used to do linear classification. SVMs can efficiently conduct non-linear classification utilising the kernel method in addition to linear classification. It allows us to translate the inputs into high-dimensional feature spaces implicitly.

LANDSLIDE PREDICTION

$$W = \frac{\sum_i^N (accuracy_i * label_i)}{Number\ of\ models}$$

; where, $accuracy_i$: accuracy of i-th model
$label_i$: predicted label vector of i-th model
N : Number of models

$$\text{Final prediction} = \begin{cases} 1 & ; W >= 0.5 \\ 0 & ; W < 0.5 \end{cases}$$

The above mentioned ML algorithm has been trained and validated by the pseudo-random data. The models can be deployed through a website/app for a wider application.

OVERVIEW OF OUR SOLUTION

GENERATION OF DATA

Using the minimum and maximum values for features of real data, pseudo-random numbers from normal distribution were generated.

VISUALIZATION OF DATA

The data was visualised using boxplot and histogram.

SPLITTING OF DATA INTO TRAINING AND TESTING DATA

The data was split 80% training and 20% testing data.
The training data was used in training the model whereas, the testing data was used to validate the model.

APPLYING MACHINE LEARNING MODELS

In this report, landslide identification is a binary classification problem, each sample is given a prediction of either 1 or 0 using the trained models. Four types of ML algorithms are chosen to evaluate the feasibility of ML in landslide identification.

LOGISTIC REGRESSION

Accuracy = 97.98%

RANDOM FOREST

Accuracy = 86.87 %

ADABOOST

Accuracy = 93.59 %

SVM

Accuracy = 97.98 %

LANDSLIDE PREDICTION

By considering a weighted mean of the above ML models, the following accuracy was achieved :

Accuracy = 97.98%

[Link to Our Dataset] [Link to Our Machine Learning Model]

REFERENCES

1. Geertsema, Marten & Highland, Lynn & Vaugeouis, Laura. (2009). Environmental Impact of Landslides. 10.1007/978-3-540-69970-5_31.
2. https://www.usgs.gov/faqs/what-a-landslide-and-what-causes-one?qt-news_science_products=0#qt-news_science_products
3. 12% Indian land prone to landslides as climate change increases the risks | Business Standard News (business-standard.com)
4. Landslide Hazard (gsi.gov.in)
5. Bhukosh (gsi.gov.in)
6. https://www.firstpost.com/india/pune-landslide-in-ambegaon-village-five-dead-150-feared-trapped-1641109.html
7. https://www.ndtv.com/india-news/goa-train-hit-by-landslide-after-massive-rain-goes-off-tracks-2493511
8. https://www.ndtv.com/india-news/uttarakhand-landslide-debris-blocks-badrinath-highway-damages-vehicles-2537564#:~:text=Landslide%20debris%20has%20blocked%20Uttarakhand's,dozens%20of%20vehicles%20in%20Sirobagad.&text=The%20landslides%20were%20triggered%20due,lives%20of%20the%20local%20residents.
9. https://www.meta-chart.com/share/number-of-deaths-due-to-landslides-across-india-from-2010-2019-2
10. https://employee.gsi.gov.in/cs/groups/public/documents/document/b3zp/odi0/~edisp/dcport1gsigovi824758.pdf
11. Vipin Kumar et al. Inferring potential landslide damming using slope stability, geomorphic constraints, and run-out analysis: a case study from the NW Himalaya. Earth Surf. Dynam., 9, 351–377, 2021
12. https://employee.gsi.gov.in/cs/groups/public/documents/document/b3zp/mziw/~edisp/dcport1gsigovi320028.pdf
13. International Journal of Science and Engineering Investigations, Volume 4, Issue 46, November 2015
14. https://nhess.copernicus.org/articles/15/1835/2015/nhess-15-1835-2015.pdf
15. https://dx.doi.org/10.1016/j.gsf.2020.02.012
16. https://www.mygreatlearning.com/blog/adaboost-algorithm/

18. https://towardsdatascience.com/quick-and-easy-explanation-of-logistics-regression-709df5cc3f1e