Critical Infrastructure Vulnerability due to Landslides

Pritesh Maheshwari, Anindita Ghosh, Himanshu Dey, Nick Meyer, Sarvesh Patidar, Vasundhara Acharya
mahesp@rpi.edu, ghosha7@rpi.edu, deyh@rpi.edu, meyern4@rpi.edu, patids@rpi.edu, acharv2@rpi.edu.
Rensselaer Polytechnic Institute, Tetherless World Constellation, Troy, NY, United States,

Abstract and Problem Area

With the increase in population and consequent climate change, the number of hazards has risen gradually. Among the natural hazards, landslides are among the most dangerous natural disasters, resulting in significant economic damage and human deaths worldwide. Comprehensive disaster management strategies must be devised and implemented to deal with natural risks.

The proposed work aims to assess the risk of landslides in the northeastern part of the United States and provide sufficient information to assist in critical infrastructure planning (focusing on power plants) development, and upgrade strategies.

To achieve the objective, data related to topography, precipitation, and other geographically related factors will be collected. The geographically related data will then be merged to form a complete training set. The data preprocessing (data cleaning, replacing missing values) will be performed. It will then be used to train a model that will predict the probability of a landslide occurring at a particular geographic location. The labels of the final dataset will be derived from NASA's landslide dataset.

Questions/Hypothesis:

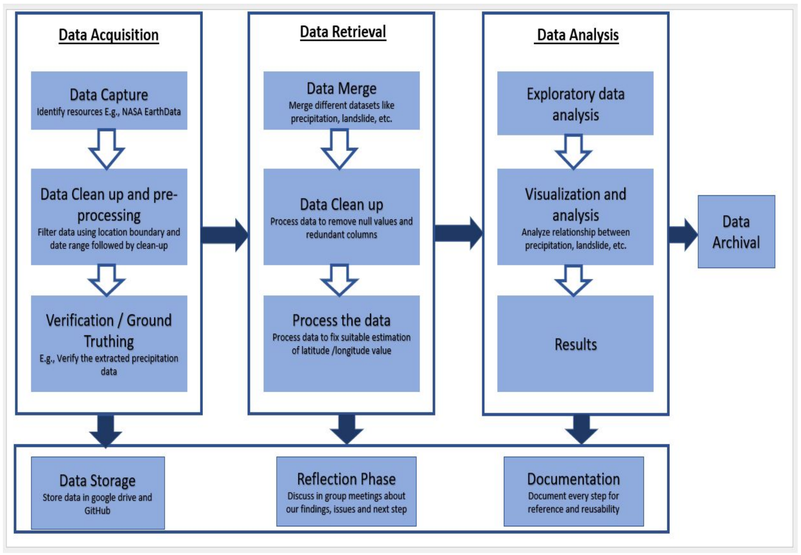
Assess the vulnerability of critical infrastructures across the northeastern part of the US caused due to landslides. Compute a risk index for each of the critical infrastructure.

•What is the relationship between factors such as the precipitation, energy, latitude, longitude, elevation, slope with the landslide probability.

The Data

- **Monthly Precipitation Data:** Goddard Earth Sciences Data and Information Services Center (GES DISC)
- **Elevation and Slope Data:** Google Earth Pro, GPS Visualizer
- **Earthquake Data:** The USGS Earthquake Hazards Program
- **Power Plants Data:** Homeland Infrastructure Foundation-Level Data (HIFLD)

Workflow

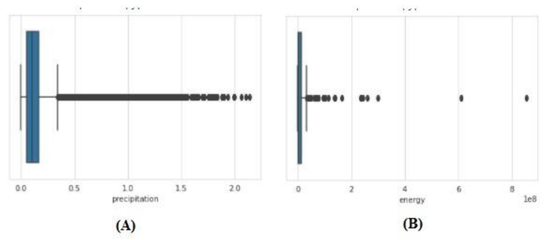


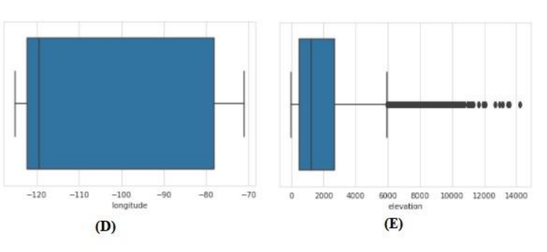
Our workflow can be broadly divided into data acquisition, data retrieval, data analysis, and data archival. Furthermore, each step has data storage, reflection phase, and documentation step in common. The project report, data, and python scripts will be archived in GitHub with all necessary documentation to ensure data preservation.

Sponsors:

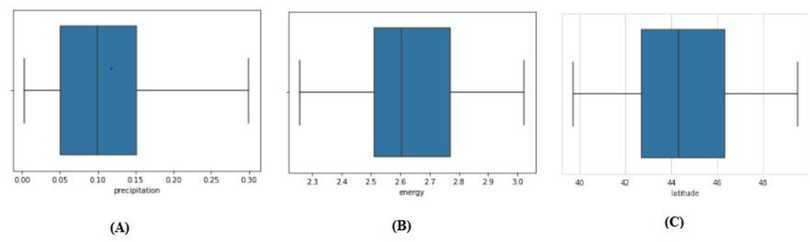
Visualization

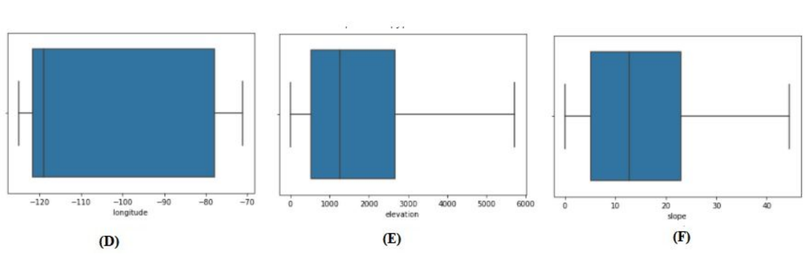
1. With Outliers





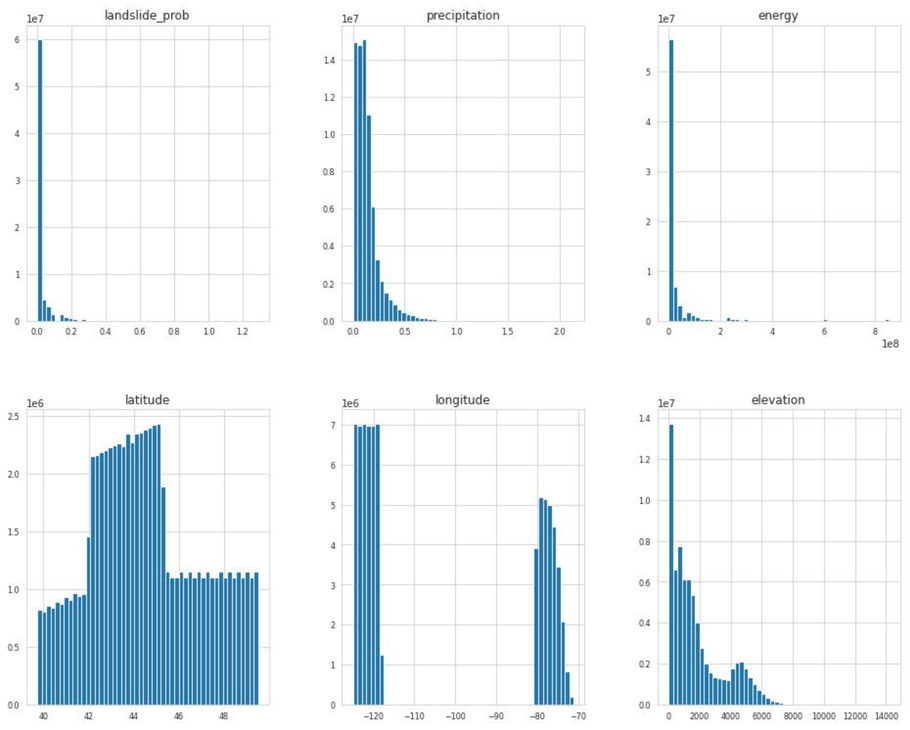
2. After elimination of Outliers





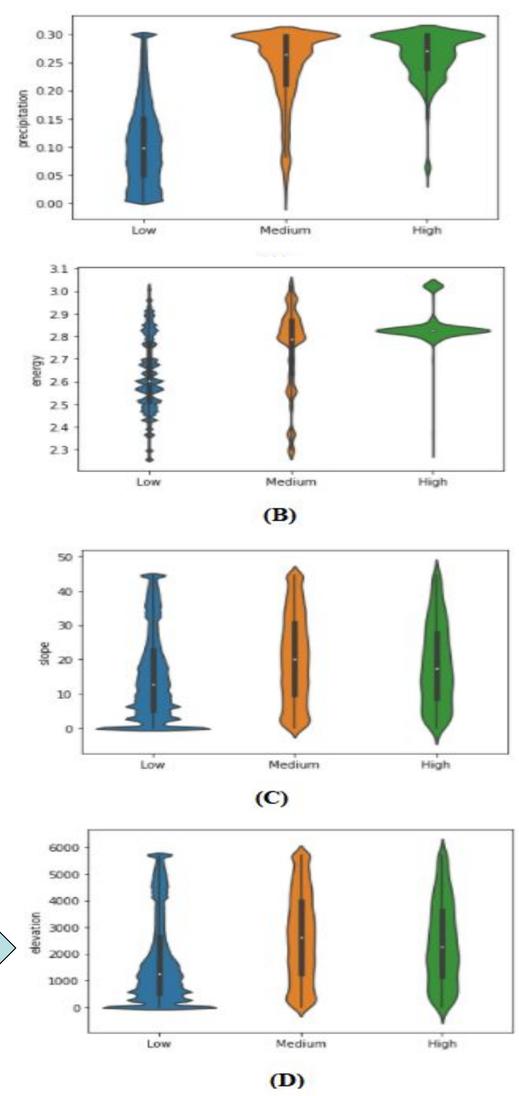
(A). Box plot of precipitation. (B). Box plot of energy. (C). Box plot of latitude. (D). Box plot of longitude. (E). Box plot of elevation. (F). Box plot of slope

3. Distribution data



Elevation and slope almost share similar distribution here. The figure shows the plot of the distribution of the numerical features.

4. Violin plot



.(A).Precipitation versus Landslide probability. (B). Energy versus Landslide probability. (c). Slope versus Landslide probability. (D). Elevation versus Landslide Probability

A violin plot of the highly correlated features is shown

5. Correlation Matrix

There is a strong positive correlation between the precipitation and landslide probability and between the elevation and landslide probability. This means the landslide probability increases with increased precipitation and increased elevation.

	landslide_prob	precipitation	energy	latitude	longitude	elevation	run	slope
landslide_prob	1.000000	0.239832	0.063006	0.069174	-0.131648	0.102372	0.129903	0.087323
precipitation	0.239832	1.000000	0.081829	-0.121491	0.272923	-0.266793	-0.274953	-0.233792
energy	0.063006	0.081829	1.000000	0.285470	-0.345650	-0.190577	0.342838	-0.239789
latitude	0.069174	-0.121491	0.285470	1.000000	-0.590008	0.129006	0.617797	0.040858
longitude	-0.131648	0.272923	-0.345650	-0.590008	1.000000	-0.420460	-0.999235	-0.290438
elevation	0.102372	-0.266793	-0.190577	0.129006	-0.420460	1.000000	0.420512	0.984160
run	0.129903	-0.274953	0.342838	0.617797	-0.999235	0.420512	1.000000	0.289989
slope	0.087323	-0.233792	-0.239789	0.040858	-0.290438	0.984160	0.289989	1.000000

Conclusion/Outcome

- To further validate our hypothesis, we collected the landslide data from the northwestern region to expand the range of the training data.
- We are able to predict the probabilities for critical vulnerabilities with respect to 7 parameters.
- Three different bins were created as follows in the following intervals: [(-0.00129, 0.43] $<$ (0.43, 0.86] $<$ (0.86, 1.29].
- The values that fell within the first bin were labeled as the "Low Risk".
- The values that fell within the second bin were labeled as the "Moderate Risk".
- The values that fell within the last bin were labeled as the "High Risk".

Resources:
R Chambers, P Kokic, P Smith, and M Cruddas. Winsorization for identifying and treating outliers in business surveys. In Proceedings of the Second International Conference on Establishment Surveys, pages 717–726. American Statistical Association Alexandria, Virginia,2000.
Pandas: Pandas.cut() method in python, Jul 2021. Pandas series.isnull(), Feb 2019.
GPS Visualizer: https://www.gpsvisualizer.com/elevation.
Power plant Data: https://hifld-geoplatform.opendata.arcgis.com/datasets/power-plants/.
NASA monthly precipitation Data: https://disc.gsfc.nasa.gov/datasets/GPM_3IMERGM_06/summary.
Earthquake Data: https://earthquake.usgs.gov/earthquakes/search/