EGU23-4851, updated on 24 Oct 2023
https://doi.org/10.5194/egusphere-egu23-4851
EGU General Assembly 2023

Resolution of data, type of inventory and data splitting in machine learning-based landslide susceptibility mapping

Neelima Satyam[1], **Minu Treesa Abraham**[2,4], and Kunal Gupta[3]
[1]Indian Institute of Technology Indore, Civil Engineering, Indore, India (neelima.satyam@gmail.com)
[2]Indian Institute of Technology Indore, Civil Engineering, Indore, India (phd1901204011@iiti.ac.in)
[3]Indian Institute of Technology Indore, Civil Engineering, Indore, India (phd2101104003@iiti.ac.in)
[4]Methods for Model-based Development in Computational Engineering, RWTH Aachen, Germany, (abraham@mbd.rwth-aachen.de)

The use of machine learning (ML) approaches for developing landslide susceptibility maps (LSM) has gained wide popularity in the recent past. The choice of ML algorithms, spatial resolution, the ratio of train-to-test data, and the landslide conditioning factors are some of the crucial factors that decide the performance of the developed LSM. However, there are no formal guidelines on the selection of any of these factors, as the choice highly depends upon the study area. In most cases, site-specific comparative analysis are required to find the best-suited combination. Two case studies were conducted for parts of the Western Ghats in India to develop pixel-based LSM for Idukki and Wayanad districts. Five different ML algorithms, two different spatial resolutions, multiple train-to-test ratios and two different types of landslide inventory data were used for developing the best-suited LSM. After detailed analysis, it was observed that the random forest (RF) algorithm has resulted in the best-performing LSM for both regions. The effects of spatial resolution and data splitting were found to be different for different algorithms, and among all the factors considered, data splitting is found to be the least influencing factor.