

1. Data analysis

Unfortunately I have not been able to load any of the provided masks so I could not do a proper data analysis. I have added the jupyter notebook file however it only contains the code from the blog post linked in the instruction.

2. Proposed model

Since this is a semantic segmentation task we could use some of the publicly available state of the art models like U-net. The biggest challenge with this task however is probably going to be the small overall area of eroded soil compared to the whole image. Looking at the polygon geometries we can see that the eroded samples are pretty small and even though there are 935 of them it is still going to be too few. Therefore we could e.g. oversample the eroded polygons or use a special cost function that would heavily prioritize this class. Alternatively we could use data augmentation to create new samples.

3. Literature review

Modelling and mapping of soil erosion susceptibility using machine learning in a tropical hot sub-humid environment

<https://doi.org/10.1016/j.jclepro.2022.132428>

In this study four models were used: Support Vector Machine, Random Forest, Boosted Regression Trees, Classification and Regression Trees. To select the most important variables for classifying the eroded soil recursive feature elimination was used. To assess the precision of soil erosion susceptibility maps ROC assessment was used. Random Forest achieved the best ROC of 0.97. The study also showed that the soil elevation, the drainage density and the NDVI index contributed the most to soil erosion.

A New Approach For Smart Soil Erosion Modelling: Integration of Empirical And Machine Learning Models

<https://doi.org/10.21203/rs.3.rs-809330/v1>

In the study researchers used four different kinds of factors like topographical factors, lithology and soil factors, hydrologic factors and anthropogenic factors. By analyzing the trained models it was determined that slope angle, land use/land cover, elevation, and rainfall erosivity are the most important soil erosion conditioning factors.