

EGU21-12046

<https://doi.org/10.5194/egusphere-egu21-12046>

EGU General Assembly 2021

© Author(s) 2021. This work is distributed under the Creative Commons Attribution 4.0 License.



Automatic landslide detection using the Random Forest classification - the importance of the train-test split ratio

Kamila Pawluszek-Filipiak and Andrzej Borkowski

Wrocław University of Environmental and Life Sciences, Institute of Geodesy and Geoinformatics, The Faculty of Environmental Engineering and Geodesy, Wrocław, Poland (kamila.pawluszek-filipiak@upwr.edu.pl)

Landslide identification is the fundamental step to reduce the potential damaging effects of landslide activities. A variety of techniques and approaches has been developed to detect landslides. Conventional landslide identification is a complex and laborious task due to a large amount of the field work and materials that have to be investigated. Additionally, the conventional geomorphological mapping mainly provides a subjective representation of landscape complexities at different scales. Sometimes, in certain conditions, such as densely-vegetated terrain, conventional landslide mapping is ineffective or even impossible.

Therefore, innovative methods that allow for the reduction of subjectivism, time, and effort have increasingly become the subject of interest in landslide research. These methods mainly focus on semi-automated or automatic landslide mapping and include analysis of remote sensing data, such as optical images, Digital Elevation Models (DEMs) derived by Light Detection and Ranging etc. Among them, the pixel-based approach (PBA) and the object-based image analysis (OBIA) methods can be distinguished, for which supervised classification methods are usually utilized.

The accuracy of supervised classification methods strongly corresponds to the training samples - its quality and amount. Supervised classification methods require the collection of training as well as testing data to generate and assess the accuracy of the classification results. It is a challenging task, especially in forested areas, to capture ground truths of the good quality to train the classifier and to identify landslides. Considering this, we decided to investigate the following research question: What is the appropriate training-testing dataset split ratio in supervised classification to detect landslides in a testing area based on DEMs? Since PBA and OBIA approaches are nowadays widely utilized, we investigated this issue for both methods. The Random Forest classifier was implemented for both methods. The experiments were performed in Poland in the Outer Carpathians.

Accuracy measures calculated for the region growing validation indicated that the training area should be similarly large to the testing area in DEM-based automatic landslide detection. Additionally, we found that the OBIA approach performs slightly better than PBA when the quantity of training samples is lower. Besides this, we also attempted to increase the detection performance and to generate final landslide inventory. For this purpose, the intersection of the OBIA and PBA results together with median filtering and the removal of small elongated objects

were carried out. We achieved the Overall Accuracy of 80% and F1 Score of 0.50.