HTT07-03 Room:422 Time:April 28 09:30-09:45

Implementation of Random Forest in landslide susceptibility study, a case study of the Tokamachi area, Niigata, Japan

PAUDEL, Uttam[1]* ; OGUCHI, Takashi[2]

[1]Graduate School of Frontier Sciences, The University of Tokyo, [2]Center for Spatial Information Science, The University of Tokyo

Random Forest (RF), a bagged trees ensemble, is widely appreciated for its superiority amongst classification algorithms and is popular in various fields of data mining. However, the application of RF in susceptibility analysis of landslide hazard remains very limited. This study highlights the results of such an attempt. The study area was selected on the basis of landslide density distribution. A density map of landslide distribution in Japan was prepared from the landslide inventory provided by the National Research Institute for Earth Science and Disaster Prevention (NIED). The Tokamachi area in Niigata Prefecture has a very high density of events and was thence selected for this study. Seven topographic factors (aspect, curvature, drainage density, elevation, plan curvature, profile curvature, and slope) derived from the 10 m DEM obtained by the Geospatial Information Authority of Japan (GSI) were used for the analysis. The classification data concern 9747 landslide events and 20685 randomly generated instances from the areas with no landslides. Unlike the values of a centroid used in many other studies, each landslide event in the classification data was represented by a mean of values of the respective factors in each landslide feature. Information gain for each factor was also evaluated and it was found that the profile curvature is the most effective factor in classifying landslides in the area, whereas elevation is the least effective. A 10-fold cross validation of the RF model with 200 trees resulted in an 'out of bag error' of 0.1443, an accuracy of 85.87%, and an ROC area of 0.926. These results suggest the suitability of RF in susceptibility analysis, the stability of which can be further strengthened with an increase of factors and the number of trees.

Keywords: Landslide susceptibility, GIS, Machine learning, Random Forest