

Landslide susceptibility mapping: machine and ensemble learning based on remote sensing big data

ABSTRACT

Predicting landslide occurrences can be difficult. However, failure to do so can be catastrophic, causing unwanted tragedies such as property damage, community displacement, and human casualties. Research into landslide susceptibility mapping (LSM) attempts to alleviate such catastrophes through the identification of landslide prone areas. Computational modelling techniques have been successful in related disaster scenarios, which motivate this work to explore such modelling for LSM. In this research, the potential of supervised machine learning and ensemble learning is investigated. Firstly, the Flexible Discriminant Analysis (FDA) supervised learning algorithm is trained for LSM and compared against other algorithms that have been widely used for the same purpose, namely Generalized Logistic Models (GLM), Boosted Regression Trees (BRT or GBM), and Random Forest (RF). Next, an ensemble model consisting of all four algorithms is implemented to examine possible performance improvements. The dataset used to train and test all the algorithms consists of a landslide inventory map of 227 landslide locations. From these sources, 13 conditioning factors are extracted to be used in the models. Experimental evaluations are made based on True Skill Statistic (TSS), the Receiver Operation characteristic (ROC) curve and kappa index. The results show that the best TSS (0.6986), ROC (0.904) and kappa (0.6915) were obtained by the ensemble model. FDA on its own seems effective at modelling landslide susceptibility from multiple data sources, with performance comparable to GLM. However, it slightly underperforms when compared to GBM (BRT) and RF. RF seems most capable compared to GBM, GLM, and FDA, when dealing with all conditioning factors.

Keyword: Big data; Landslide susceptibility; Flexible discriminant analysis; Random forest; Ensemble model