Data mining and statistical approaches in debris-flow susceptibility modelling using airborne LiDAR data

ABSTRACT

Cameron Highland is a popular tourist hub in the mountainous area of Peninsular Malaysia. Most communities in this area suffer frequent incidence of debris flow, especially during monsoon seasons. Despite the loss of lives and properties recorded annually from debris flow, most studies in the region concentrate on landslides and flood susceptibilities. In this study, debris-flow susceptibility prediction was carried out using two data mining techniques; Multivariate Adaptive Regression Splines (MARS) and Support Vector Regression (SVR) models. The existing inventory of debris-flow events (640 points) were selected for training 70% (448) and validation 30% (192). Twelve conditioning factors namely; elevation, plan-curvature, slope angle, total curvature, slope aspect, Stream Transport Index (STI), profile curvature, roughness index, Stream Catchment Area (SCA), Stream Power Index (SPI), Topographic Wetness Index (TWI) and Topographic Position Index (TPI) were selected from Light Detection and Ranging (LiDAR)-derived Digital Elevation Model (DEM) data. Multi-collinearity was checked using Information Factor, Cramer's V, and Gini Index to identify the relative importance of conditioning factors. The susceptibility models were produced and categorized into five classes; not-susceptible, low, moderate, high and very-high classes. Models performances were evaluated using success and prediction rates where the area under the curve (AUC) showed a higher performance of MARS (93% and 83%) over SVR (76% and 72%). The result of this study will be important in contingency hazards and risks management plans to reduce the loss of lives and properties in the area.

Keyword: Debris flows; Susceptibility; Machine learning; MARS; SVR; LiDAR; GIS; Remote sensing