Spatial prediction of flood susceptible areas using rule based decision tree (DT) and ensemble bivariate and multivariate statistical models in GIS

ABSTRACT

Flood is one of the natural hazards which occur all over the world and it is critical to be controlled through proper management. Severe flood events in Kelantan, Malaysia cause damage to both life and property every year, and therefore the development of flood model to recognize the susceptible areas in watersheds is important for decision makers. Remote sensing (RS) and geographic information system (GIS) techniques could be useful in hydrological studies while they are able to fulfill all the requirements for comprehensive, rapid and accurate analysis. The aim of the current research is to compare the prediction performances of two different approaches such as rule-based decision tree (DT) and combination of frequency ratio (FR) and logistic regression (LR) statistical methods for flood susceptibility mapping at Kelantan, Malaysia. DT is based on the rules which are created precisely and strongly by considering all the characteristics of the variables which can enhance the performance of the flood susceptibility mapping. On the other hand, LR as multivariate statistical analysis (MSA) has some weak points. For that reason, FR was used to analyze the impact of classes of each variable on flood occurrence and overcome the weakness of LR. At first, flood inventory map with a total of 155 flood locations was extracted from various sources over the part of the Kelantan. Then the flood inventory data was randomly divided into a testing dataset 70% (115 flood locations) for training the models and the remaining 30% (40 flood locations) was used for validation purpose. The spatial database includes digital elevation model (DEM), curvature, geology, river, stream power index (SPI), rainfall, land use/cover (LULC), soil type, topographic wetness index (TWI) and slope. For validation both success and prediction rate curves were performed. The validation results showed that, area under the curve for the results of DT and integrated method of FR and LR was 87% and 90% for success rate and 82% and 83% for prediction rate respectively.

Keyword: Flood susceptibility, Decision tree (DT), Ensemble, GIS, Remote sensing, Malaysia