Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method

ABSTRACT

Flood is one of the most commonly occurred natural hazards worldwide. Severe flood occurrences in Kelantan, Malaysia cause damage to both life and property every year. Due to the huge losses in this area, development of appropriate flood modeling is required for the government. Remote sensing and geographic information system techniques can support overall flood management as they can produce rapid data collection and analysis for hydrological studies. The existing models for flood mapping have some weak points that may improve through more sophisticated and ensemble methods. The current research aimed to propose a novel ensemble method by integrating support vector machine (SVM) and frequency ratio (FR) to produce spatial modeling in flood susceptibility assessment. In the literature, mostly statistical and machine learning methods are used individually; however, their integration can enhance the final output. The FR model can perform bivariate statistical analysis and evaluate the correlation between the flooding and classes of each conditioning factors. The weights achieved by FR can be assigned to each conditioning factor and the resulted factors can be used in SVM analysis. In order to examine the efficiency of the proposed ensemble method and to show the proficiency of SVM, another machine learning algorithm such as decision tree (DT) was applied and the results were compared. To perform the methods, the upper catchment of the Kelantan basin in Malaysia was chosen. First, a flood inventory map with a total of 155 flood locations were extracted from various sources over the study area. The flood inventory map was randomly divided into two dataset; 70 % (115 flood locations) for the purpose of training and the remaining 30 % (40 flood locations) was used for validation. The spatial database included digital elevation model, curvature, geology, river, stream power index, rainfall, land use/cover, soil type, topographic wetness index and slope. For model validation, area under curve method was used and both success and prediction rate curves were calculated. The validation results for ensemble method showed 88.71 and 85.21 % for success rate and prediction rate respectively. The DT model showed 87.00 and 82.00 % for the success rate and prediction rate respectively. It is evident that the accuracies were increased using the ensemble method. The acquired results proved the efficiency of the proposed ensemble method as rapid, accurate and reasonable in flood susceptibility assessment.

Keyword: Flood susceptibility; Support vector machine (SVM); Decision tree (DT); Remote sensing; GIS; Malaysia