Science of the Total Environment 651 (2019) 2087–2096

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines

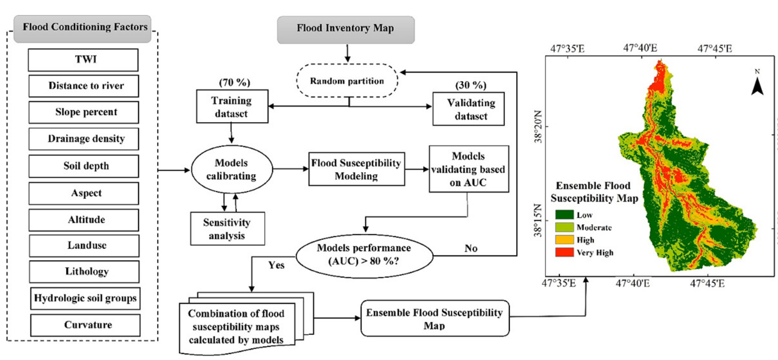
Bahram Choubin [a,b,*], Ehsan Moradi [b], Mohammad Golshan [a], Jan Adamowski [c], Farzaneh Sajedi-Hosseini [b], Amir Mosavi [d,e]

[a] *Department of Watershed Management, Sari Agricultural Sciences and Natural Resources University, Sari, Iran*
[b] *Department of Reclamation of Arid and Mountainous Regions, University of Tehran, Karaj, Iran*
[c] *Department of Bioresource Engineering, Faculty of Agricultural and Environmental Sciences, Macdonald Campus, McGill University, Canada*
[d] *Institute of Automation, Kalman Kando Faculty of Electrical Engineering, Obuda University, Budapest, Hungary*
[e] *Institute of Advanced Studies Koszeg, IASK, Koszeg, Hungary*

HIGHLIGHTS

- Ensemble machine learning (ML) predicting flood susceptibility
- Contribution of models with accuracy values above 80% in ensembling process
- Area under curve (AUC) for the models ranges from 0.83 to 0.89.
- ML allows quick priority of prone areas for the remediation of floods.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 3 July 2018
Received in revised form 1 October 2018
Accepted 5 October 2018
Available online 06 October 2018

Editor: Ashantha Goonetilleke

Keywords:
Ensemble approach
Flood susceptibility
Support vector regression
Multivariate discriminant analysis
Classification and regression trees

ABSTRACT

Floods, as a catastrophic phenomenon, have a profound impact on ecosystems and human life. Modeling flood susceptibility in watersheds and reducing the damages caused by flooding is an important component of environmental and water management. The current study employs two new algorithms for the first time in flood susceptibility analysis, namely multivariate discriminant analysis (MDA), and classification and regression trees (CART), incorporated with a widely used algorithm, the support vector machine (SVM), to create a flood susceptibility map using an ensemble modeling approach. A flood susceptibility map was developed using these models along with a flood inventory map and flood conditioning factors (including altitude, slope, aspect, curvature, distance from river, topographic wetness index, drainage density, soil depth, soil hydrological groups, land use, and lithology). The case study area was the Khiyav-Chai watershed in Iran. To ensure a more accurate ensemble model, this study proposed a framework for flood susceptibility assessment where only those models with an accuracy of >80% were permissible for use in ensemble modeling. The relative importance of factors was determined using the Jackknife test. Results indicated that the MDA model had the highest predictive accuracy (89%), followed by the SVM (88%) and CART (0.83%) models. Sensitivity analysis showed that slope percent, drainage density, and distance from river were the most important factors in flood susceptibility mapping. The

* Corresponding author at: Department of Watershed Management, Sari Agricultural Sciences and Natural Resources University, Sari, Iran.
E-mail address: bahram.choubin@ut.ac.ir (B. Choubin).

https://doi.org/10.1016/j.scitotenv.2018.10.064

ensemble modeling approach indicated that residential areas at the outlet of the watershed were very susceptible to flooding, and that these areas should, therefore, be prioritized for the prevention and remediation of floods.

1. Introduction

Climate change is contributing to an increase in abnormal weather events such as flooding, which is one of the most devastating natural disasters (Hirabayashi et al., 2013; CRED and UNISDR, 2015). Flooding can have a profound impact on ecosystems and human life (Alexander et al., 2011) and can result in environmental damage and economic loss to residential areas, agriculture and water resources (Lee et al., 2015; Klaus et al., 2016; Wu et al., 2017). Population increase and the rapid growth of urban development and riverside construction have increased the risk of flooding in many areas (Ahmadisharaf et al., 2016). Floods generally begin suddenly and increase rapidly over several hours (Borga et al., 2008). They occur most commonly in the spring due to severe rainfall and/or snowmelt (Furl et al., 2018). In addition, morphological changes to river channels, caused by human or natural interference, may sometimes result in changes to river flow and cause flooding (Yousefi et al., 2018). Iran has faced many flood events in different regions due to large climatic and rainfall variations (Khosravi et al., 2016a, 2016b; Norouzi and Taslimi, 2012). The locations most vulnerable to flooding are areas with high population density and agricultural lands with marginal drainage, as well as river networks where runoff from rainfall events has concentrated (Dankers et al., 2014). Modeling flood susceptibility in watersheds and reducing the damage caused by flooding is an important component of environmental and water management (Siahkamari et al., 2017).

Various methods have been used to identify and evaluate flood susceptible areas. For instance, some studies used multi-criteria decision analysis (MCDA) methods, such as the analytic hierarchy process (AHP), to assess flood susceptibility (Yahaya et al., 2010; Chen et al., 2011; Luu et al., 2018; Tang et al., 2018). These methods are based on expert knowledge, which may be skewed by ambiguous judgments and uncertainty (Miles and Snow, 1984). Other studies used statistical methods such as frequency ratio (FR) and regression logistics for mapping flood susceptibility and hazardous areas (Rahmati et al., 2016; Siahkamari et al., 2017; Samanta et al., 2018). These methods depend on predicted variables based on relationships with various explanatory parameters and depend greatly on the size of datasets (McLay et al., 2001). Physically based models such as HEC-RAS and MIKE11 are also used for studying floods (e.g., Gharbi et al., 2016) but require a large amount of data, substantial processing power, and often use parameters that demand extensive calculation time (Shrestha et al., 2013; Hong et al., 2018b). Newer methods include machine learning techniques such as random forest (RF), artificial neural networks (ANN) and support vector machines (SVM), which can identify areas prone to flooding and produce flood susceptibility maps (Lee et al., 2017; Zhao et al., 2018; Termeh et al., 2018). Each of these approaches exhibits certain weaknesses in terms of flood susceptibility mapping that can be improved through an ensemble modeling approach. Ensemble modeling is a process of combining the predictions of single models into an integrated model to increase prediction accuracy (Rokach, 2010). The main objective of this study was to use and apply an ensemble approach to generate a flood susceptibility map. To achieve this, SVM, a universal machine learning method, was applied in conjunction with two other robust machine learning methods (i.e., MDA and CART) that have not been used to date for the spatial modeling of floods. The main contribution of SVM is excellent generalization, but it is difficult to capture important modeling variables using this method. Most important and critical modeling variables can be applied with MDA (Elith et al., 2008; Sajedi-Hosseini et al., 2018b). The CART model has several capabilities, such as the use of many trees for process description and modeling, insensitivity to data distribution and the existence of data outliers, and the ability to easily account for categorical and numerical variables in the modeling process, such as land use and slope, respectively (Sutton, 2004; Choubin et al., 2018). Ensemble modeling, a synthesis of the individual model predictions, was employed in the current study to address the limitations of the SVM, MDA, and CART models in the context of flood susceptibility mapping (Lee et al., 2012b; Sajedi-Hosseini et al., 2018b).

Accordingly, the objectives of the current research were to: i) explore the capability of the CART, MDA and SVM methods to predict flood susceptibility, ii) compare the results of the new algorithms used in this study (i.e., MDA and CART) with a commonly used algorithm (i.e., SVM) to predict flood susceptibility areas, iii) evaluate the importance of flood conditioning factors in producing flood susceptibility maps, and iv) use and apply an ensemble modeling approach to generate a flood susceptibility map. The current paper first provides a description of the study area, the Khiyav-Chai watershed in Iran. Next, in the context of flood susceptibility prediction and assessment, the Methodology section demonstrates the preparation of a flood inventory map, the selection of the appropriate flood conditioning factors, and introduces tools and techniques. Subsequently, the prediction results from the different applied models are presented and discussed. Finally, some conclusions from the current research are drawn.

2. Materials and methods

2.1. Study area

The Khiyav-Chai watershed is located in Ardabil Province in north-western Iran. This watershed is an important tourist area that generates economic and ecological benefits for both the government and the local people (Porkhial et al., 2008). The watershed area is 126 km^2 and is located between 47°38′34″ to 47°48′18″ east and 38°12′30″ to 38°23′51″ north (Fig. 1). This area has a rugged topography and an elevation between 1372 m at the river outlet and 4353 m upstream. The climate is semi-arid with an average annual precipitation of 368 mm (2000–2016), and an average temperature of 10 °C (Moghaddam, 2006). The average humidity in the Khiyav-Chai watershed is about 58%. The lowest humidity occurs in July at 13% and the highest is in May at 85%. The mean annual discharge at the watershed outlet (Pol-Soltani hydrometric station) was about 1.8 m^3 s^{-1} from 1970 to 2016.

2.2. Methodology

The flowchart of the methodology used in this study is shown in Fig. 2. It includes (i) preparing a flood inventory map, (ii) determining the appropriate flood conditioning factors, (iii) modeling the flood susceptibility map using CART, SVM and MDA algorithms, (iv) performing an accuracy assessment of the models, (v) implementing a sensitivity analysis of flood conditioning factors, and (vi) producing a flood susceptibility map using an ensemble modeling approach.

2.2.1. Flood inventory map

The flood inventory map is a basic map for flood susceptibility assessment (Tehrany et al., 2014; Rahmati et al., 2016). Accurate analysis of flood susceptibility requires a precise flood inventory map that shows the locations of flood occurrence. In the current study, 51 flood location points, identified between the years 2010 to 2017, were obtained from