Environmental Modelling & Software 95 (2017) 229–245

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

A novel hybrid artificial intelligence approach for flood susceptibility assessment

Kamran Chapi [a, *], Vijay P. Singh [b], Ataollah Shirzadi [a], Himan Shahabi [c], Dieu Tien Bui [d], Binh Thai Pham [e], Khabat Khosravi [f]

[a] Department of Rangeland and Watershed Management, Faculty of Natural Resources, University of Kurdistan, Sanandaj, Iran
[b] Department of Biological and Agricultural Engineering, College of Agriculture and Life Sciences, Texas A & M University, College Station, TX 77843-2117, USA
[c] Department of Geomorphology, Faculty of Natural Resources, University of Kurdistan, Sanandaj, Iran
[d] Geographic Information System Group, Department of Business Administration and Computer Science, University College of Southeast Norway, Hallvard Eikas Plass 1, N-3800 Bøi Telemark, Norway
[e] Department of Geotechnical Engineering, University of Transport Technology, 54 TrieuKhuc, ThanhXuan, Hanoi, Viet Nam
[f] Department of Watershed Sciences Engineering, Faculty of Natural Resources, University of Agricultural Science and Natural Resources of Sari, Mazandaran, Iran

ARTICLE INFO

Article history:
Received 11 February 2017
Received in revised form
13 April 2017
Accepted 15 June 2017

Keywords:
Flood susceptibility
Bagging-LMT
Bayesian logistic regression
Logistic model tree
Iran

ABSTRACT

A new artificial intelligence (AI) model, called Bagging-LMT - a combination of bagging ensemble and Logistic Model Tree (LMT) - is introduced for mapping flood susceptibility. A spatial database was generated for the Haraz watershed, northern Iran, that included a flood inventory map and eleven flood conditioning factors based on the Information Gain Ratio (IGR). The model was evaluated using precision, sensitivity, specificity, accuracy, Root Mean Square Error, Mean Absolute Error, Kappa and area under the receiver operating characteristic curve criteria. The model was also compared with four state-of-the-art benchmark soft computing models, including LMT, logistic regression, Bayesian logistic regression, and random forest. Results revealed that the proposed model outperformed all these models and indicate that the proposed model can be used for sustainable management of flood-prone areas.

1. Introduction

Floods cause huge losses of life and property each year in one part of the world or the other (Ceola et al., 2014). reported that in 2011 and 2012 alone, floods affected around 200 million people and caused economic losses of about $95 billion. Floods are more hazardous than other natural catastrophic hazards, such as landslides, earthquakes, and volcanoes (Bolt et al., 2013). Hence, it is of paramount importance to manage floods and reduce their risk, which requires flood prediction and computation of inundation areas.

Flood prediction decreases flood-related fatalities and associated economic losses (Bubeck et al., 2012), and delineation of flood prone areas is a key element in any flood alleviation strategy (Sarhadi et al., 2012). In Iran, especially in northern parts, floods are observed almost every year. During the last decade, economic damages by flood disasters have sharply increased, largely owing to increased urbanization involving conversion of forest lands to residential areas and building of villas (Khosravi et al., 2016; Norouzi and Taslimi, 2012).

Flooding is a complex phenomenon and hence prediction of flood occurrence is difficult (Pappenberger et al., 2006). For predicting the probability of a flood and for mitigating and managing future floods, mapping flood susceptibility is an essential step (Kourgialas and Karatzas, 2011). To that end, modeling of flood hazards, which may entail multi-temporal datasets, is required (Martinez and Le Toan, 2007). In recent years, flood susceptibility and hazard mapping has been done using remote sensing data and GIS tools (Bates, 2004, 2012; Haq et al., 2012; Pradhan et al., 2009,

* Corresponding author.
E-mail address: kamranchapi@gmail.com (K. Chapi).

http://dx.doi.org/10.1016/j.envsoft.2017.06.012
1364-8152/

2014; Rahmati et al., 2016; Tehrany et al., 2015b; Wanders et al., 2014).

Examples of the GIS-spatial models in flood studies, including bivariate and statistical models, have are frequency ratio (Lee, 2012; Tehrany et al., 2015a), analytical hierarchy process (Kazakis et al., 2015; Rahmati et al., 2016), logistic regression (Fekete, 2009; Tehrany et al., 2014a), and weights-of evidence (WOE) (Rahmati et al., 2016; Tehrany et al., 2014b). For mapping flood susceptibility, fuzzy logic (Pulvirenti et al., 2011), neuro-fuzzy logic (Bui et al., 2016a; Mukerji et al., 2009), artificial neural networks (Kia et al., 2012; Nikoo et al., 2016), decision trees (DT) (Tehrany et al., 2013), support vector machines (SVM) (Tehrany et al., 2015a, 2015b), and naive Bayes (Liu et al., 2015) have been used. However, no consensus has yet been reached on the selection of best model for flood susceptibility and new methods for flood susceptibility are therefore needed.

Machine learning ensembles and hybrid methods have recently been shown to perform better than conventional methods (Pham et al., 2016b, 2017); however, these methods have not been used for flood susceptibility modeling. This study therefore aims to propose a new hybrid intelligence model, based on bagging ensemble and logistic model tree, namely "Bagging-LMT" model, for flood susceptibility modeling and identifying areas prone to flooding with a case study of the Haraz watershed, northern Iran. The Logistic Model Tree (LMT) classifier is one of the recent machine learning methods and is a hybrid of Decision Tree (DT) classifier and Logistic Regression (LR) function. Bagging ensemble is an ensemble technique that can improve the accuracy of prediction of a base classifier.

The proposed model is evaluated using training and validation datasets, sensitivity, specificity, accuracy, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Area Under the Receiver Operating Characteristic (ROC) curve (AUROC). Additionally, the proposed model is compared with four state-of-the-art benchmark soft computing machine learning methods: logistic model tree (LMT), logistic regression (LR), Bayesian logistic regression (BLR), and random forest (RF).

2. Study area and flood database

2.1. Description of study area

The study area is the Haraz watershed in northern Iran and lies between longitudes of 51°43′ to 52°36′E, and the latitudes of 35°45′ to 36°22′N. The watershed, with an area of about 4014 km^2 (Fig. 1), is characterized by mountains, hills, rivers, and streams. It has an altitude ranging from 300 m to 5595 m with a mean value of 2674 m and a standard deviation of 699 m. Its slope angle varies from 0° to 66°. Approximately 77.25% of the total area has slopes greater than 30°; while 3.83% of the area has slopes less than 5° and 69.56% of the area has slopes from 5° and 20°.

The average annual rainfall is 723 mm. The majority of rainfall generally occurs in January, February, March, and October, among which October is the wettest month with an average rainfall of 160 mm. The surficial geology is covered by Triassic, Jurassic, Cretaceous, Permian, Tertiary and Quaternary geologic formations. About 62.7% of the region is configured by formations of the Jurassic period (39.3%) and Quaternary period (23.4%). The land use consists of grasslands (pasture), forest lands, wood lands, barren lands, water bodies and residential areas; whereas, grassland covers most the area (92.6%). Floods occur almost every year in the watershed because of high-intensity short duration rainfall events, land use changes from pastures and forests to agricultural and residential areas, and lack of any action plan to prevent occurrences of floods.

2.2. Flood conditioning factors

Factors, that affect flood susceptibility mapping, may vary, depending on the watershed characteristics (Bui et al., 2016a) and determination of these factors is vital for flood modeling (Sanyal and Lu, 2004).

For the Haraz watershed, eleven conditioning factors, including slope, elevation, curvature, NDVI, stream power index (SPI), topographic wetness index (TWI), lithology, land use, rainfall, stream

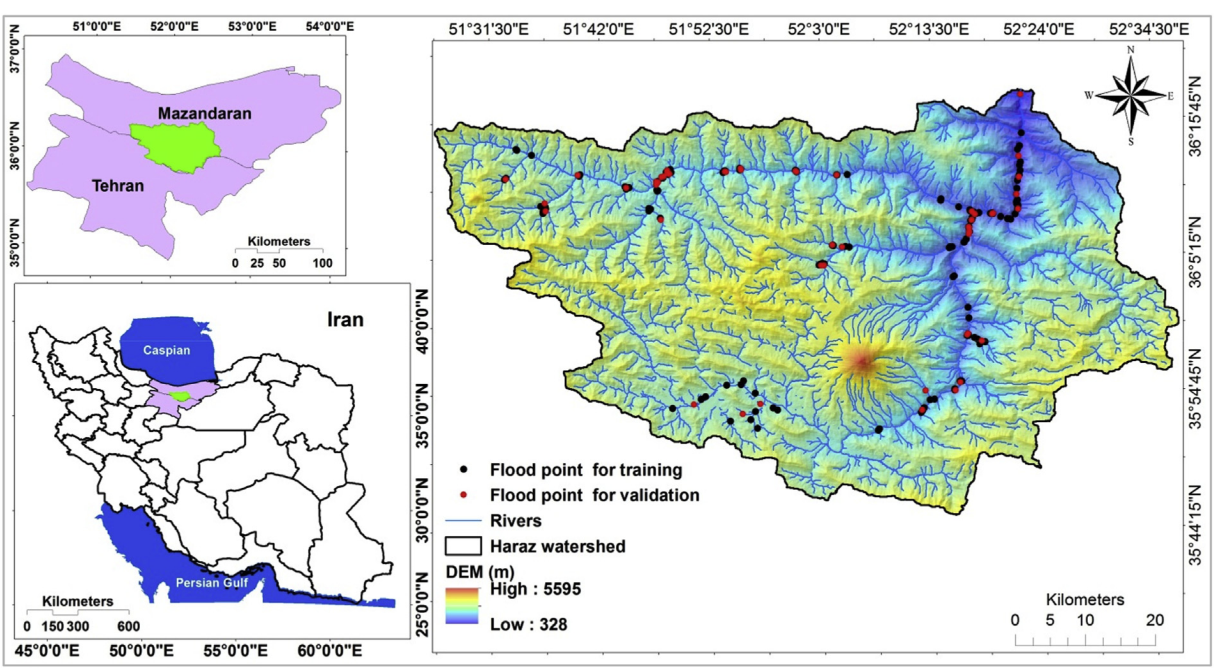

Fig. 1. Location of the study area and flood inventories.