Catena 163 (2018) 399–413

Contents lists available at ScienceDirect

Catena

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

Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China)

Haoyuan Hong[a,b,c], Junzhi Liu[a,b,c], Dieu Tien Bui[d], Biswajeet Pradhan[e,f], Tri Dev Acharya[g], Binh Thai Pham[h,i], A-Xing Zhu[a,b,c,*], Wei Chen[j,**], Baharin Bin Ahmad[k]

[a] *Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, Nanjing 210023, China*
[b] *State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing 210023, China*
[c] *Jiangsu Center for Collaborative Innovation in Geographic Information Resource Development and Application, Nanjing, Jiangsu 210023, China*
[d] *Geographic Information System group, Department of Economics and Computer Sciences, Telemark University College, Bø i Telemark, N-3800, Norway*
[e] *School of Systems, Management and Leadership, Faculty of Engineering and IT, University of Technology Sydney, CB11.06.217, Building 11, 81 Broadway, Ultimo, NSW 2007, (PO Box 123), Australia*
[f] *Department of Energy and Mineral Resources Engineering, Choongmu-gwan, Sejong University, 209 Neungdong-ro Gwangjin-gu, Seoul 05006, Republic of Korea*
[g] *Department of Civil Engineering, Kangwon National University, Republic of Korea*
[h] *Department of Civil Engineering, Gujarat Technological University, Nr.Visat Three Roads, Visat - Gandhinagar Highway, Chandkheda, Ahmedabad 382424, Gujarat, India*
[i] *Department of Geotechnical Engineering, University of Transport Technology, 54 Trieu Khuc, Thanh Xuan, Ha Noi, Viet Nam*
[j] *College of Geology and Environment, Xi'an University of Science and Technology, Xi'an 710054, China*
[k] *Department of Geoinformation, Faculty of Geoinformation and Real Estate, Universiti Teknologi Malaysia (UTM), Malaysia*

ARTICLE INFO

Keywords:
J48 Decision Tree
AdaBoost
Bagging
Rotation Forest
Guangchang
GIS

ABSTRACT

Landslides are a manifestation of slope instability causing different kinds of damage affecting life and property. Therefore, high-performance-based landslide prediction models are useful to government institutions for developing strategies for landslide hazard prevention and mitigation. Development of data mining based algorithms shows that high-performance models can be obtained using ensemble frameworks. The primary objective of this study is to investigate and compare the use of current state-of-the-art ensemble techniques, such as AdaBoost, Bagging, and Rotation Forest, for landslide susceptibility assessment with the base classifier of J48 Decision Tree (JDT). The Guangchang district (Jiangxi province, China) was selected as the case study. Firstly, a landslide inventory map with 237 landslide locations was constructed; the landslide locations were then randomly divided into a ratio of 70/30 for the training and validating models. Secondly, fifteen landslide conditioning factors were prepared, such as slope, aspect, altitude, topographic wetness index (TWI), stream power index (SPI), sediment transport index (STI), plan curvature, profile curvature, lithology, distance to faults, distance to rivers, distance to roads, land use, normalized difference vegetation index (NDVI), and rainfall. Relief-F with the 10-fold cross-validation method was applied to quantify the predictive ability of the conditioning factors and for feature selection. Using the JDT and its three ensemble techniques, a total of four landslide susceptibility models were constructed. Finally, the overall performance of the resulting models was assessed and compared using area under the receiver operating characteristic (ROC) curve (AUC) and statistical indexes. The result showed that all landslide models have high performance (AUC $>$ 0.8). However, the JDT with the Rotation Forest model presents the highest prediction capability (AUC = 0.855), followed by the JDT with the AdaBoost (0.850), the Bagging (0.839), and the JDT (0.814), respectively. Therefore, the result demonstrates that the JDT with Rotation Forest is the best optimized model in this study and it can be considered as a promising method for landslide susceptibility mapping in similar cases for better accuracy.

1. Introduction

A Landslide is a manifestation of slope instability causing different kinds of damage affecting life and property (Dehnavi et al., 2015; Gutiérrez et al., 2015; Pham et al., 2017b). To reduce these damages, the assessment of slope conditions that have potential for landslides

* Correspondence to: A-X. Zhu, Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, Nanjing 210023, China.
** Corresponding author.
E-mail addresses: hong_haoyuan@outlook.com (H. Hong), azhu@wisc.edu (A.-X. Zhu), chenwei0930@xust.edu.cn (W. Chen).

https://doi.org/10.1016/j.catena.2018.01.005
Received 26 July 2017; Received in revised form 30 December 2017; Accepted 3 January 2018

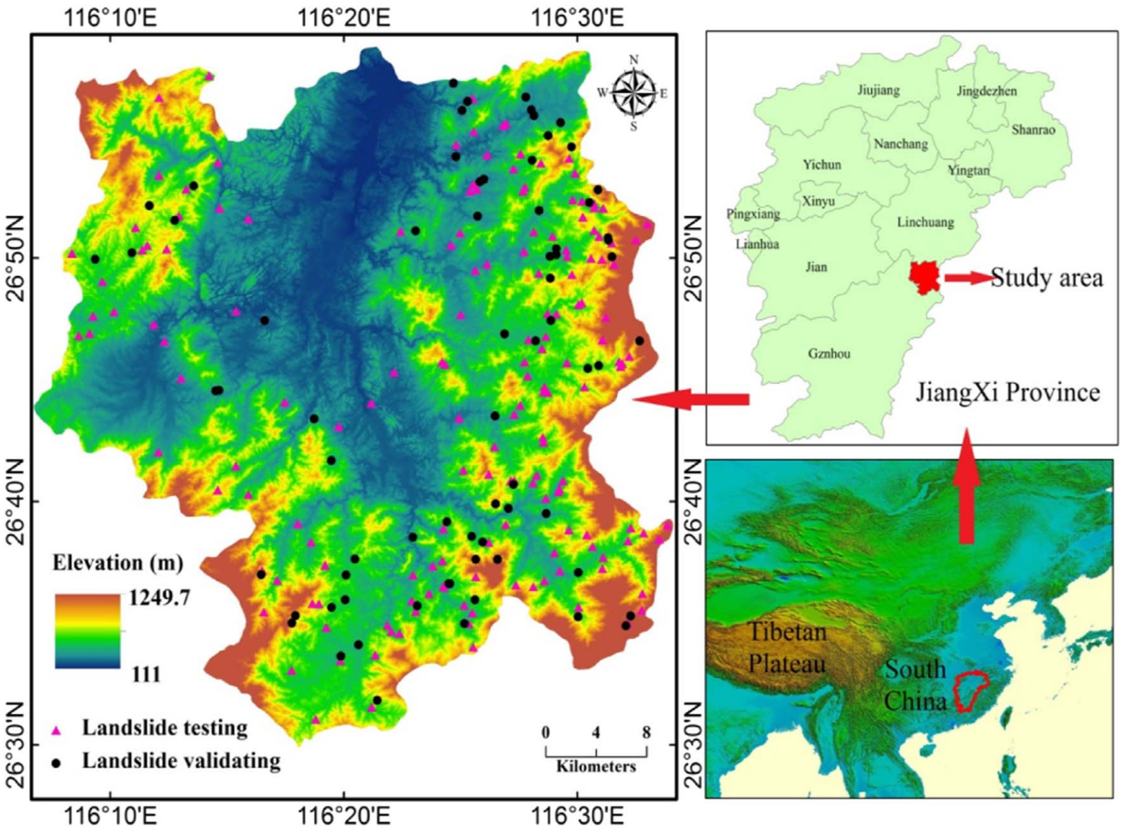

Fig. 1. Geographical position of study area with training and validation landslides.

should be carried out and the results can be used to develop strategies for landslide hazard prevention and mitigation (Havenith et al., 2015; Komac and Hribernik, 2015; Lacasse and Nadim, 2009). However, the quality of the slope assessment is one of the most debated subjects during the last few decades (Safran et al., 2015; Sewell et al., 2015; Vallet et al., 2015). Therefore, studies on the improvement of model prediction capability have long attracted researchers in the field of landslides (Chen et al., 2017e; Kavzoglu et al., 2014; Ma et al., 2015; Mathew et al., 2014; Zhao et al., 2015).

Since landslides are complex processes that relate to various geo-environmental variables and triggering factors, the modeling of a landslide is not a simple task. Although various methods and techniques have been proposed in recent years, these methods can be divided into knowledge based conventional and data driven statistical methods (Alimohammadlou et al., 2014; Kavzoglu et al., 2014; Lee et al., 2013; Pradhan, 2013). Conventional methods are subjective whereas statistical methods are highly dependent on data and applied models. With advances in computing technology, statistical methods have been evolved into machine learning techniques such as, artificial neural networks (ANN), logistic regression (LR), decision trees (DT), and support vector machines (SVM) and has been reported to outperform conventional methods (Alimohammadlou et al., 2014; Althuwaynee et al., 2014; Chen et al., 2017c; Chen et al., 2017d; Chen et al., 2017f; Chen et al., 2017i; Conforti et al., 2014; Felicisimo et al., 2013; Hong et al., 2017b; Hong et al., 2017c; Hong et al., 2017d; Hong et al., 2016; Hong et al., 2015; Kavzoglu et al., 2014; Lee et al., 2013; Pradhan, 2013; Regmi et al., 2014; Tien Bui et al., 2015). These landslide models also require constantly evaluation as landslide information and their related factors change. But, with limited available information on landslide and factors, prediction power and robustness are the two main aspects in selection of techniques for better modeling results.

Ensemble methods are machine learning techniques in which a prediction model is formed from a combination of various base classifiers. Ensemble methods have shown significant improvement over the individual models (Kuncheva, 2014; Ozdemir and Altural, 2013; Shahabi et al., 2014; Umar et al., 2014). The ongoing development of data mining has introduced various ensemble frameworks such as i.e., Stacking (Wolpert, 1992), Bagging (Breiman, 1996), AdaBoost (Freund and Schapire, 1997), Random Subspace (Ho, 1998), MultiBoost (Webb, 2000), Random Forests (Breiman, 2001), DECORATE (Melville and Mooney, 2003), and Rotation Forest (Rodriguez et al., 2006). Rotation Forest is one of the ensemble methods that outperforms the others in terms of overall performance in various datasets (Rodriguez et al., 2006; Zhou et al., 2013).

Use of these classifier ensemble approaches has shown enhancement of results (Chen et al., 2017a; Chen et al., 2017b; Kakuwa et al., 2014; Šilhán and Stoffel, 2015; Tien Bui et al., 2013, 2014; Trigila et al., 2015). Tien Bui et al. (2013) assessed the prediction performance of landslide models using Bagging and AdaBoost ensembles with J48 Decision Tree (JDT) and reported that the Bagging ensemble showed the highest prediction capability along the National Road 32 of Vietnam. In another study, Tien Bui et al. (2014) compared Bagging ensample with JDT and fuzzy rule-based classifier for landslide modeling at the Lang Son city area (Vietnam), and stated that the Bagging ensemble have significantly improved predictive power for the individual base classifiers for landslide modeling. Despite such promising results, studies on the application of these frameworks for landslide predictions are still rare (Mohammady et al., 2012; Pham et al., 2017a; Xu et al., 2015; Yenes et al., 2015). Hence, exploration of potential application of these frameworks in different areas is essential for landslide studies.

The Guangchang district of Jiangxi province is a mountainous region in southern part of China. It has experienced various landslides during the last ten years (Fan et al., 2012; Lin et al., 2012; Xu et al., 2014; Yin et al., 2010). These landslides were mainly the result of torrential rainfall during the rainy season (Weiguang et al., 2012).