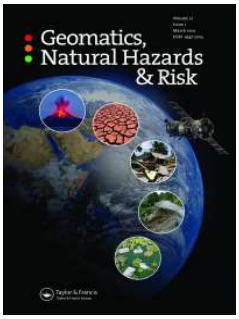

Geomatics, Natural Hazards and Risk

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgnh20

Robustness analysis of machine learning classifiers in predicting spatial gully erosion susceptibility with altered training samples

Tusar Kanti Hembram, Sunil Saha, Biswajeet Pradhan, Khairul Nizam Abdul Maulud & Abdullah M. Alamri

To cite this article: Tusar Kanti Hembram, Sunil Saha, Biswajeet Pradhan, Khairul Nizam Abdul Maulud & Abdullah M. Alamri (2021) Robustness analysis of machine learning classifiers in predicting spatial gully erosion susceptibility with altered training samples, Geomatics, Natural Hazards and Risk, 12:1, 794-828, DOI: 10.1080/19475705.2021.1890644

To link to this article: https://doi.org/10.1080/19475705.2021.1890644

Published online: 10 Mar 2021.

GEOMATICS, NATURAL HAZARDS AND RISK
2021, VOL. 12, NO. 1, 794–828
https://doi.org/10.1080/19475705.2021.1890644

Taylor & Francis
Taylor & Francis Group

Robustness analysis of machine learning classifiers in predicting spatial gully erosion susceptibility with altered training samples

Tusar Kanti Hembram[a], Sunil Saha[a], Biswajeet Pradhan[b,c,d], Khairul Nizam Abdul Maulud[d,e] and Abdullah M. Alamri[f]

[a]Department of Geography, University of Gour Banga, Malda, West Bengal, India; [b]The Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and Information Technology, University of Technology Sydney, NSW, Australia; [c]Department of Energy and Mineral Resources Engineering, Sejong University, Seoul, Republic of Korea; [d]Earth Observation Centre, Institute of Climate Change, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia; [e]Department of Civil Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, UKM Bangi, Selangor, Malaysia; [f]Department of Geology & Geophysics, College of Science, King Saud University, Riyadh, Saudi Arabia

ABSTRACT

The present research intended to assess the robustness of three popular machine learning models, i.e. random forest (RF), boosted regression tree (BRT) and naïve bayes (NB) in spatial gully erosion susceptibility modelling in Jainti River basin, India. A gully inventory map of 208 gullies was prepared through field survey and Google earth imageries. Following the 70/30 ratio, three randomly sampled groups of altered training and validation gully sets G1, G2 and G3 were prepared for modelling gully erosion susceptibility. Using information gain ratio and multi-collinearity analysis, 14 gully conditioning factors (GCF) were selected. The discrimination ability and reliability of the models were measured through Kappa coefficient, efficiency, receiver operating characteristic curve, root-mean-square-error (RMSE) and mean-absolute-error (MAE). The stability of the machine learning models was estimated by comparing the accuracy statistics and the departure in areal outcomes among intra-model and inter-model. RF model was found as the most consistent. With the highest mean AUC (0.903), efficiency (91.17), Kappa coefficient (0.835) and lowest RMSE (0.192) and MAE (0.081), RF was found to be more consistent when the training and validation data sets were altered. The effectiveness of each input GCFs was determined using map removal sensitivity analysis technique. This study could be supportive in ascertaining model deployment for mapping gully erosion and managing the land resource.

ARTICLE HISTORY
Received 17 June 2020
Accepted 10 February 2021

KEYWORDS
Machine learning; gully erosion; GIS; altered gully samples; India

CONTACT Biswajeet Pradhan biswajeet.pradhan@uts.edu.au

scientific reports

Check for updates

OPEN

Measuring landslide vulnerability status of Chukha, Bhutan using deep learning algorithms

Sunil Saha[1], Raju Sarkar[2]✉, Jagabandhu Roy[1], Tusar Kanti Hembram[3], Saroj Acharya[4], Gautam Thapa[5] & Dowchu Drukpa[6]

Landslides are major natural hazards that have a wide impact on human life, property, and natural environment. This study is intended to provide an improved framework for the assessment of landslide vulnerability mapping (LVM) in Chukha Dzongkhags (district) of Bhutan. Both physical (22 nos.) and social (9 nos.) conditioning factors were considered to model vulnerability using deep learning neural network (DLNN), artificial neural network (ANN) and convolution neural network (CNN) approaches. Selection of the factors was conceded by the collinearity test and information gain ratio. Using Google Earth images, official data, and field inquiry a total of 350 (present and historical) landslides were recorded and training and validation sets were prepared following the 70:30 ratio. Nine LVMs were produced i.e. a landslide susceptibility (LS), one social vulnerability (SV) and a relative vulnerability (RLV) map for each model. The performance of the models was evaluated by area under curve (AUC) of receiver operating characteristics (ROC), relative landslide density index (R-index) and different statistical measures. The combined vulnerability map of social and physical factors using CNN (CNN-RLV) had the highest goodness-of-fit and excellent performance (AUC = 0.921, 0.928) followed by DLNN and ANN models. This approach of combined physical and social factors create an appropriate and more accurate LVM that may—support landslide prediction and management.

Among the numerous natural hazards, landslides are considered to be one of the biggest, as they can cause tremendous loss of life and property as well as affect the natural ecosystem. Now attention is being shifted towards the issue of landslides, as increasing developmental works in the Himalayan areas are taking place[1]. The Himalayan region's tectonic fragility is already illustrated, with an overabundance of literature focusing on the broad scale and few on the micro-scale. Being a part of this region, Bhutan has long been known as a place sensitive to natural disasters, such as landslides[2]. Therefore, the need for micro-scale study i.e. province-wise study in Bhutan can be fruitful to act locally viz. the present analysis.

In recent decades, the most suitable way to tackle this hazardous event is the spatial assessment of vulnerability to landslides[3,4]. Assessment of geomorphological, geological, tectonic, climate, vegetation, and land practices may help in identifying the area susceptible and vulnerable to landslides[5,6]. The occurrence of landslides is a natural phenomenon but man-made processes are one of the causes of vulnerability to landslides, which makes it difficult to predict the spatial and temporal occurrence of landslides[7]. In this connection, landslide vulnerability maps (LVMs) may be used as the primary method to classify the high-risk zones which are prone to landslides and also help in identifying the variables responsible for the occurrence of landslides[8]. Presently, with the development of various tools, hardware, and availability of data, it has become easy to produce landslide susceptibility and vulnerability maps[9].

In spatial landslide modelling, the terms "susceptibility" and "vulnerability" are often used interchangeably; however, "susceptibility" often points to causes intrinsic to physical predisposition (e.g., structural and topographical), while "vulnerability" often corresponds to external influences along with causes intrinsic to physical predisposition (e.g., anthropogenic exposure)[10]. Most of the previous landslide studies have not considered the "vulnerability" aspect rather those have modelled the "susceptibility". Therefore, another exclusivity of the present work also lies in it. According to the report (May 2010) of XVIth SAARC (The South Asian Association of Regional Cooperation) summit hosted by the Royal Government of Bhutan, several social factors are also

[1]Department of Geography, University of Gour Banga, Malda, West Bengal, India. [2]Department of Civil Engineering, Delhi Technological University, Delhi, India. [3]Department of Geography, Nistarini College, Purulia, West Bengal, India. [4]National Centre for Hydrology and Meteorology, Thimphu, Bhutan. [5]Royal Government of Bhutan Projects, Phuentsholing, Bhutan. [6]Department of Geology and Mines, Ministry of Economic Affairs, Thimphu, Bhutan. ✉email: rajusarkar@dce.ac.in

ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Advances in Space Research xxx (xxxx) xxx

ADVANCES IN SPACE RESEARCH (a COSPAR publication)

www.elsevier.com/locate/asr

Hybrid ensemble machine learning approaches for landslide susceptibility mapping using different sampling ratios at East Sikkim Himalayan, India

Sunil Saha [a,*], Jagabandhu Roy [a], Biswajeet Pradhan [b,c,d], Tusar Kanti Hembram [e]

[a] Department of Geography, University of Gour Banga, Malda-732103, West Bengal, India
[b] The Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and Information Technology, University of Technology Sydney, NSW 2007, Australia
[c] Center of Excellence for Climate Change Research, King Abdulaziz University, P. O. Box 80234, Jeddah 21589, Saudi Arabia
[d] Earth Observation Centre, Institute of Climate Change, Universiti Kebangsaan Malaysia, Bangi 43600, Malaysia
[e] Department of Geography, Nistarini College, Purulia-723101, West Bengal, India

Received 19 January 2021; received in revised form 12 May 2021; accepted 20 May 2021

Abstract

Landslide is a big problem in the mountainous region all over the world. Sikkim Himalayan region is also suffering from landslide problem. This study's main objective was to generate landslide susceptibility map (LSM) considering the hybrid ensemble of machine learning approaches using different sample ratios. Random Forest (RF) as the base classifier an ensemble with bagging, Rotation Forest (RTF), and Random Subspace (RS) Meta classifiers were used for spatial landslide modeling. First, collected 86 landslides locations through field investigation and from Sikkim district disaster office were mapped as a landslide inventory. Collected landslide locations were categorized into training and testing datasets randomly using four sample ratios (50:50, 60:40, 70:30 and 80:20). Based on the four sampling ratios and fifteen conditioning factors, a total of sixteen LSMs were prepared using RF, Bagging-RF (B-RF), RTF-RF and RS-RF in GIS platform. For assessing the modeling accuracy and comparison among these, the area under the receiver operating characteristics (AUROC) and other statistical methods such as root-mean-square-error (RMSE), mean-absolute-error (MAE) and R-index methods were used. The overall proficiency of RS-RF (AUC = 0.871, 0.847 of 50%:50%, AUC = 0.925, 0.931 of 60%:40%, AUC = 0.933, 0.939 of 70%:30%; AUC = 0.927, 0.933 of 80%:20%) was found to be substantially greater than the results of RF, B-RF, and RTF-RF. The RS-RF model and 70:30 sample ratio had the highest goodness-of-fit and accuracy as per the RMSE, MAE, and R-index methods. Furthermore, the model based on RS-RF is a promising and acceptable way of mapping regional landslides.

Keywords: Landslide susceptibility; GIS; Meta classifier; Ensemble machine learning; East Sikkim Himalaya

1. Introduction

Himalayan region of India is the most natural hazard zone, affected by landslides and earthquakes. Approximately 80% of landslides in India occur in this area (Onagh et al., 2012). Sikkim is the part of the Himalaya where landslide is a reoccurring problem. This area is frequently affected by the landslide every year due to heavy rainfall during the monsoon time (Roy et al., 2019). Recently, landslides have increased in this area, affecting life and property due to much anthropogenic interferences (Mukane, 2014; Sarkar et al., 1995).

* Corresponding author.
E-mail addresses: sunilgeo.88@gmail.com (S. Saha), jagabandhuroy1991@gmail.com (J. Roy), biswajeet.pradhan@uts.edu.au (B. Pradhan), tusarpurulia1991@gmail.com (T.K. Hembram).

https://doi.org/10.1016/j.asr.2021.05.018
0273-1177/