5th Intercontinental Geoinformation Days

igd.mersin.edu.tr

Landslide susceptibility assessment employing machine learning ensemble models: a study in the most severely battered district of the Southern Western Ghats

Rajendran Shobha Ajin*[1], Hamza Vijith [1], Megha Krishna Prasad [2], Jayan Bindu Jibitha [3], Romulus Costache [4,5,6]

[1]Kerala State Emergency Operations Centre (KSEOC), Kerala State Disaster Management Authority (KSDMA), Thiruvananthapuram, Kerala, India
[2]Bharathidasan University, Department of Remote Sensing, Tiruchirappalli, Tamil Nadu, India
[3]University of Kerala, Department of Environmental Sciences, Thiruvananthapuram, Kerala, India
[4]National Institute of Hydrology and Water Management, Bucharest, Romania
[5]Department of Civil Engineering, Transilvania University of Brasov, Brasov, Romania
[6]Danube Delta National Institute for Research and Development, Tulcea, Romania

Keywords
GIS
Idukki
Landslides
Machine learning
REPTree
Western Ghats

Abstract

Landslides are one of the natural catastrophes which are frequently reported in the Western Ghats region and result in severe loss. This modelling intends to demarcate susceptible zones in one of the most impacted districts in the southern Western Ghats. Idukki, being the worst affected district, reported more than 2000 landslides in the year 2018. Two machine learning ensemble models and geoinformation techniques have been employed to identify the susceptibility. Twelve landslide conditioning factors have been utilized for this study. The ROC curve-based validation technique ascertained good and fair prediction capability for the created maps, with AUC scores of 0.821 and 0.776 for the MB-REPTree and AB-REPTree models, respectively. From the validation scores, it is found that the MB-REPTree model is more efficient and of good operational use. The study found 7.81% of the district as very highly susceptible and 16.06% as highly susceptible. So, this study suggests that the MB-REPTree model is the best model to demarcate susceptible zones, not just in the Western Ghats but also in other places with similar climatic and terrain conditions.

1. Introduction

The Western Ghats region of Kerala has been severely battered by landslides, especially during the 2018 monsoon season (Ajin et al. 2022a; Ajin et al. 2022b; Hao et al. 2020; Thomas et al. 2021). A total of 4728 landslides have been reported in Kerala in the year 2018 alone (Hao et al. 2020). The Pettimudi disaster, which resulted in the deaths of 70 people (Achu et al. 2021), was the most disastrous, with the highest death toll reported in Kerala, followed by the Kavalappara disaster with 59 deaths (Ajin et al. 2022a). According to the landslide incidence data collected from the Bhukosh portal (https://bhukosh.gsi.gov.in/Bhukosh/Public) of the Geological Survey of India (GSI), Idukki district witnessed 1304 landslides during the 2018 monsoon (Ajin et al. 2022b; Thomas et al. 2021). Recent studies revealed that apart from heavy downpour, development activities and unplanned modification of hill slopes were the major causes of landslides in Idukki district (Abraham et al. 2019, 2021; Ajin et al. 2022b). This underlines the need for a susceptibility map that is validated by employing efficient statistical methods and verified incidence data.

This research is an attempt to identify landslide susceptible zones in Idukki district employing two machine learning (ML) ensemble models such as AdaBoost Reduced-Error Pruning Tree (AB-REPTree) and MultiBoost Reduced-Error Pruning Tree (MB-REPTree) and to assess the efficacy of these models to identify the best model among these two. The analysis used twelve conditioning factors that can induce landslides, including slope, distance from the road, curvature, lineament density, topographic position index (TPI), soil types, lithology, land use/land cover (LULC), stream power index (SPI), aspect, elevation, and rainfall.

* **Corresponding Author**

*(ajinares@ieee.org) ORCID ID 0000-0003-3073-4390
(vijithh@gmail.com) ORCID ID 0000-0002-1064-2088
(meghakprasad99@gmail.com) ORCID ID 0000-0002-9890-1207
(jibithajb@gmail.com) ORCID ID 0000-0001-9748-300X
(romuluscostache2000@yahoo.com) ORCID ID 0000-0002-6876-8572

Cite this study

Ajin RS, Vijith H, Prasad MK, Jibitha JB, & Costache R (2022). Landslide susceptibility assessment employing machine learning ensemble models: a study in the most severely battered district of the Southern Western Ghats. 5thIntercontinental Geoinformation Days (IGD), 26-29, Netra, India

2. Method

2.1. Study Area

Idukki, Kerala's second-largest district, is covered by forests on more than half of its land (Abraham et al. 2019). This hilly district is located in the Western Ghats region and has a total area of 4358 km^2 (Abraham et al. 2021). The main landforms are structural and denudational hills, as well as some mountains that are higher than 2000 metres (Abraham et al. 2021). Figure 1 depicts the location of the study area.

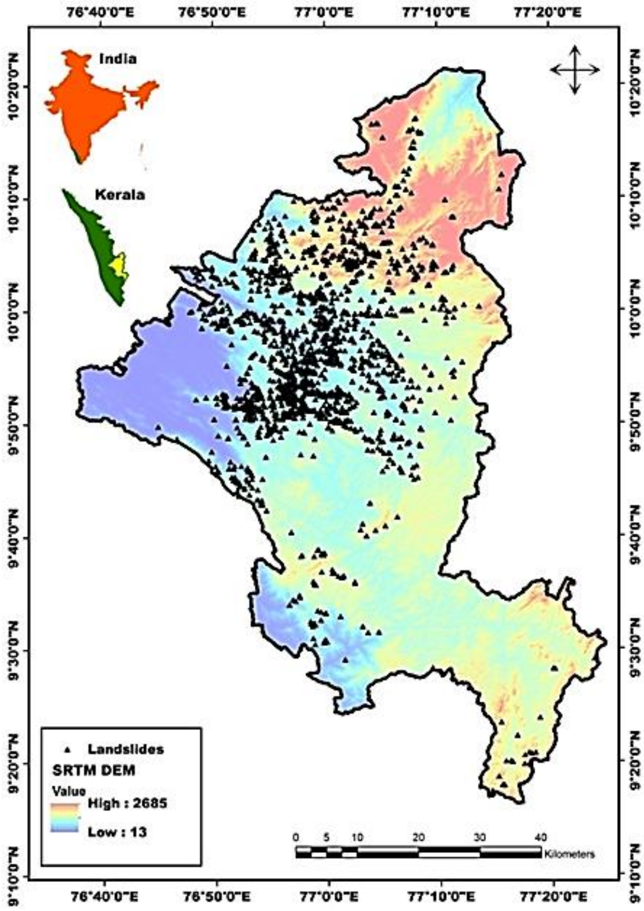

Figure 1. Location of the study area

2.2. Construction of landslide inventory

This study has utilized the landslide inventory created by Hao et al. (2020), which comprises 2219 landslides. The data has been split into training (80%) and validation (20%) datasets (Gautam et al. 2021). The training dataset comprises 1775 landslide locations and the validation dataset comprises 444 landslide locations, respectively.

2.3. Derivation of factors

The factors such as slope, aspect, and elevation were derived from the SRTM DEM utilizing the ArcGIS 10.4 spatial analyst tools. The curvature, SPI, and TPI were derived from the DEM by employing spatial analyst and raster calculator tools. The soil types were extracted from the map published by the National Bureau of Soil Survey and Land Use Planning (NBSS&LUP), whereas the lithology and lineaments were derived from the map published by the GSI. The line density tool in ArcGIS was used to compute the lineament density. The LULC types were extracted from the Landsat 8 OLI images using ERDAS Imagine 9.1 software. The road networks were extracted from the topographic map and Google Earth, and the Euclidean distance tool was used to derive the distance from the road layer. The rainfall distribution was extracted from the World Climate Report portal (https://www.worldclim.org/). The R 4.2.1 software was utilized for computing the weights. The landslide susceptibility of the district is categorized into five zones employing the Natural Breaks method (Babitha et al. 2022).

2.4. REPTree model

The reduced-error pruning tree (REPTree) is an ensemble model of the decision tree (DT) and reduced error pruning (REP) techniques that is used to generate a DT model by reducing the variance and can successfully manage missing data (Bui et al. 2020; Vishwakarma et al. 2022). It utilizes information gain ratio (IGR) values to construct a regression/decision tree and employs reduced-error pruning to prune the tree (Al Snousy et al. 2011).The IGR values were computed by applying Equation 1 and the entropy (E) function (Bui et al. 2020).

$$IGR(x,S) = \frac{E(S) - \sum_{i=1}^{n} \frac{E(S_i)|S_i|}{|S|}}{-\sum_{i=1}^{n} \frac{|S_i|}{|S|} log_2 \frac{|S_i|}{|S|}} \tag{1}$$

Where S = training dataset, S_i = subset (i = 1, 2, 3, …. n), E = entropy function

2.5. AB-REPTree model

AdaBoost (AB) is an ensemble approach that performs boosting, which integrates weak classifiers in series to produce a strong classifier, and trains and deploys trees in series (Misra and Li 2020). AB reduces variance and bias (Jukic et al. 2020), but it performs poorly in noisy environments and takes longer to train (Misra and Li 2020). The weights were computed by applying Equation 2 (Wang et al. 2021).

$$F_n(x) = F_{m-1}(x) + argmin_h \sum_{i=1}^{n} L\left(y_{i,} \middle| F_{m-1}(x_i) + h(x_i)\right) \tag{2}$$

where $F_n(x)$ is the overall model, $F_{n-1}(x)$ is the overall obtained in the previous round, $y_i$is the prediction result of the i-th tree, and $h(x_i)$ is the newly added tree (Wang et al. 2021).

2.6. MB-REPTree model

MultiBoost (MB) is an ensemble model of AdaBoost and Bagging, and can reduce both bias and variance, thereby reducing the errors to a greater extent (Webb 2000, 2011). The base classifier errors were computed employing Equation 3 (Shirzadi et al. 2018).

$$e = \frac{\sum_{x_j \in S', C_t(x_j) \neq y_j} weight(x_j)}{m} \quad (3)$$

where e = errors of base classifiers, S' = dataset; and x_j and y_j = elements of datasets (Shirzadi et al. 2018).

2.7. Validation of the models

The created maps were validated by employing the ROC curve technique (Metz 1978) and the validation dataset. The AUC scores between 0.5-0.6, 0.6-0.7, 0.7-0.8, 0.8-0.9, and 0.9-1.0 depict failure, poor, fair, good, and excellent prediction capabilities, respectively (Lüdemann et al. 2006). The IBM SPSS Statistics 23.0 software was utilized for creating the ROC curves and determining the AUC scores.

3. Results

The landslide susceptibility maps created employing two different ML ensemble models are depicted in Figure 2. The validation of the models revealed that the MB-REPTree model (0.821, or 82.1%) has a better prediction capability than the AB-REPTree model (0.776, or 77.6%) (Figure 3). According to the MB-REPTree model, a total of 7.81% of the district is categorized as a very high-susceptible zone. The percentage of each susceptible zone is depicted in Table 1.

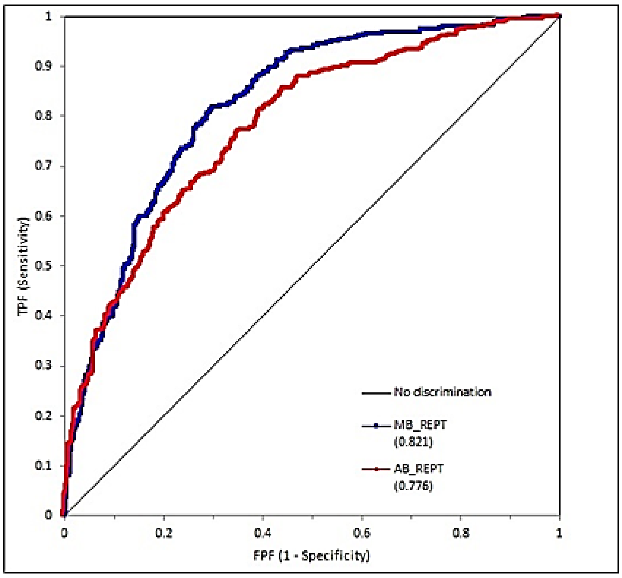

Figure 3. The ROC curves

4. Conclusion

This study effectively categorized the landslide susceptible zones in Idukki district utilizing ML ensemble models and identified the MB-REPTree model as the best model among these two models. The created maps will help decision makers and officials of the emergency management department implement ideal mitigation measures that will help in reducing disaster risks in the future. Also, these types of validated susceptibility maps will help find important buildings and infrastructure that are in critical zones.

Table 1. Percentage of landslide susceptible zones

Susceptible zones	Percentage of susceptible zones AB-REPTree	MB-REPTree
Very low	10.08	21.49
Low	27.41	32.32
Moderate	30.76	22.32
High	21.66	16.06
Very high	10.09	7.81
Total	**100**	**100**

References

Abraham, M. T., Pothuraju, D., & Satyam, N. (2019). Rainfall thresholds for prediction of landslides in Idukki, India: An empirical approach. Water, 11(10). https://doi.org/10.3390/w11102113

Abraham M. T., Satyam, N., Shreyas, N., Pradhan, B., Segoni, S., Maulud, K. N. A., & Alamri, A. M. (2021). Forecasting landslides using SIGMA model: a case study from Idukki, India. Geomatics, Natural Hazards and Risk, 12(1), 540-559. https://doi.org/10.1080/19475705.2021.1884610

Achu, A. L., Joseph, S., Aju, C. D., & Mathai, J. (2021). Preliminary analysis of a catastrophic landslide event on 6 August 2020 at Pettimudi, Kerala State, India. Landslides, 18, 1459–1463. https://doi.org/10.1007/s10346-020-01598-x

Ajin, R. S., Nandakumar, D., Rajaneesh, A., Oommen, T., Ali, Y. P., & Sajinkumar, K. S. (2022a). The tale of three landslides in the Western Ghats, India: lessons to be learnt. Geoenvironmental Disasters, 9. https://doi.org/10.1186/s40677-022-00218-1

Ajin, R. S., Saha, S., Saha, A., Biju, A., Costache, R., & Kuriakose, S. L. (2022b). Enhancing the accuracy of the REPTree by integrating the hybrid ensemble meta-classifiers for modelling the landslide susceptibility of Idukki district, South-western India. Journal of the Indian Society of Remote Sensing 50, 2245–2265. https://doi.org/10.1007/s12524-022-01599-4

Al Snousy M. B., El-Deeb, H. M., Badran, K., & Al Khlil, I. A. (2011). Suite of decision tree-based classification algorithms on cancer gene expression data. Egyptian Informatics Journal, 12(2), 73-82. https://doi.org/10.1016/j.eij.2011.04.003

Babitha, B. G., Danumah, J. H., Pradeep, G. S., Costache, R., Patel, N., Prasad, M. K., Rajaneesh, A., Mammen, P. C., Ajin, R. S., & Kuriakose, S. L. (2022). A framework employing the AHP and FR methods to assess the landslide susceptibility of the Western Ghats region in Kollam district. Safety in Extreme Environments 4, 171–191. https://doi.org/10.1007/s42797-022-00061-5

Bui, D. T., Shirzadi, A., Amini, A., Shahabi, H., Al-Ansari, N., Hamidi, S., Singh, S. K., Pham, B. T., Ahmad, B. B., & Ghazvinei, P. T. (2020) A hybrid intelligence approach to enhance the prediction accuracy of local scour depth at complex bridge piers. Sustainability, 12(3). https://doi.org/10.3390/su12031063

Gautam, P., Kubota, T., Sapkota, L. M., & Shinohara, Y. (2021). Landslide susceptibility mapping with GIS in high mountain area of Nepal: a comparison of four methods. Environmental Earth Sciences 80. https://doi.org/10.1007/s12665-021-09650-2

Hao, L., Rajaneesh A., van Westen, C., Sajinkumar K. S., Martha, T. R., Jaiswal, P., & McAdoo, B. G. (2020). Constructing a complete landslide inventory dataset for the 2018 monsoon disaster in Kerala, India, for land use change analysis, Earth System Science Data, 12, 2899–2918. https://doi.org/10.5194/essd-12-2899-2020

Jukic, S., Saracevic, M., Subasi, A., & Kevric, J. (2020). Comparison of ensemble machine learning methods for automated classification of focal and non-focal epileptic EEG signals. Mathematics, 8(9). https://doi.org/10.3390/math8091481

Lüdemann, L., Grieger, W., Wurm, R., Wust, P., & Zimmer, C. (2006). Glioma assessment using quantitative blood volume maps generated by T1-weighted dynamic contrast-enhanced magnetic resonance imaging: a receiver operating characteristic study. Acta Radiologica, 47(3), 303-310. https://doi.org/10.1080/0284185050053

Metz, C. E. (1978). Basic principles of ROC analysis. Seminars in Nuclear Medicine, 8(4), 283-298. https://doi.org/10.1016/S0001-2998(78)80014-2

Misra, S., & Li, H. (2020). Chapter 9 - Noninvasive fracture characterization based on the classification of sonic wave travel times, In: Misra, S., Li, H., He, J. (Eds) Machine Learning for Subsurface Characterization, Gulf Professional Publishing, pp. 243-287. https://doi.org/10.1016/B978-0-12-817736-5.00009-0

Shirzadi, A., Soliamani, K., Habibnejhad, M., Kavian, A., Chapi, K., Shahabi, H., Chen, W., Khosravi, K., Pham, B. T., Pradhan, B., Ahmad, A., Ahmad, B. B., & Bui, D. T. (2018). Novel GIS based machine learning algorithms for shallow landslide susceptibility mapping. Sensors, 18(11). https://doi.org/10.3390/s18113777

Thomas, A. V., Saha, S., Danumah, J. H., Raveendran, S., Prasad, M. K., Ajin, R. S., & Kuriakose, S. L. (2021). Landslide susceptibility zonation of Idukki district using GIS in the aftermath of 2018 Kerala floods and landslides: a comparison of AHP and frequency ratio methods. Journal of Geovisualization and Spatial Analysis 5. https://doi.org/10.1007/s41651-021-00090-x

Vishwakarma, D. K., Ali, R., Bhat, S. A., Elbeltagi, A., Kushwaha, N. L., Kumar, R., Rajput, J., Heddam, S. & Kuriqi, A. (2022). Pre- and post-dam river water temperature alteration prediction using advanced machine learning models. Environmental Science and Pollution Research. https://doi.org/10.1007/s11356-022-21596-x

Wang, C., Xu, S., & Yang, J. (2021). Adaboost algorithm in artificial intelligence for optimizing the IRI prediction accuracy of asphalt concrete pavement. Sensors, 21(17). https://doi.org/10.3390/s21175682

Webb, G. I. (2000). MultiBoosting: A technique for combining boosting and wagging. Machine Learning 40, 159–196. https://doi.org/10.1023/A:1007659514849

Webb, G. I. (2011). MultiBoosting. In: Sammut, C., Webb, G. I. (Eds) Encyclopedia of Machine Learning. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-30164-8_567

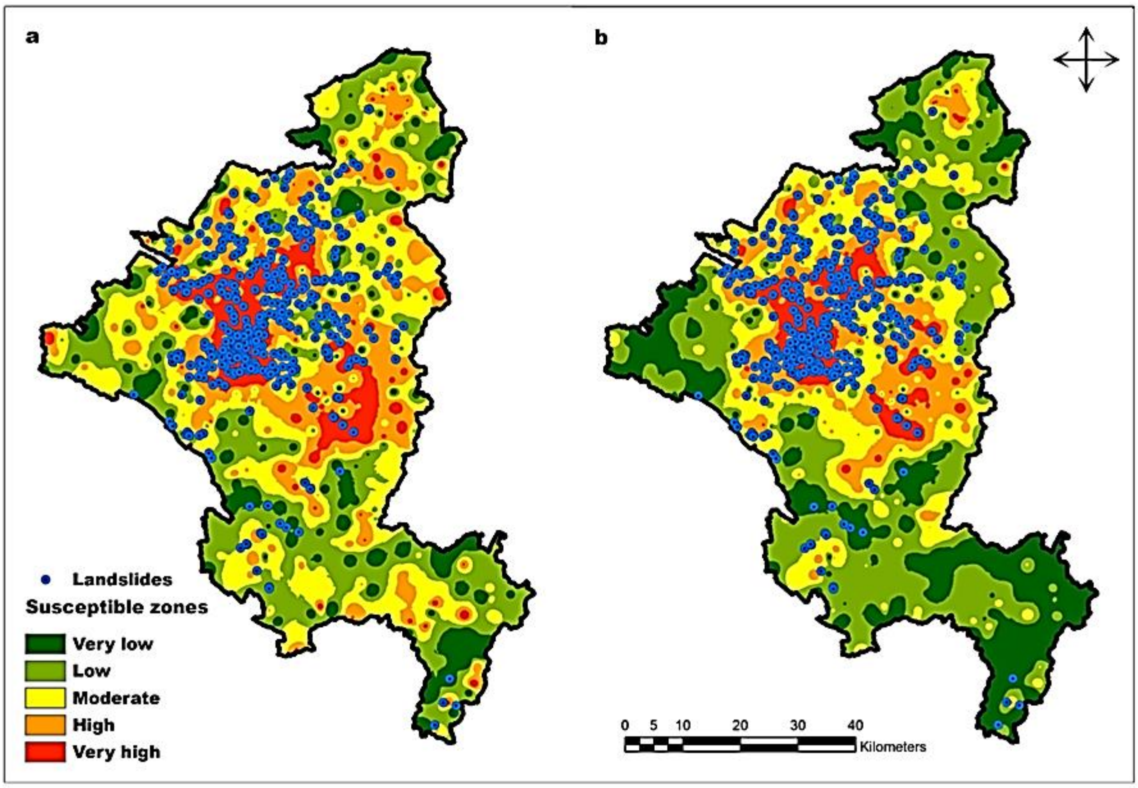

Figure 2. Landslide susceptible zones **a.** AB-REPTree model **b.** MB-REPTree model