第35回日本情報地質学会講演会 講演要旨集 031-032頁 2024年

Evaluating Machine Learning Models for Landslide Susceptibility Mapping in Yen Bai Province, Vietnam

Tran TUNG LAM*, Tatsuya NEMOTO*, Truong XUAN QUANG** and Venkatesh RAGHAVAN*

*Department of Geosciences, Graduate School of Science, Osaka Metropolitan University, 3-3-138 Sugimoto Sumiyoshi-ku, Osaka, 558-8585, Japan E-mail: sp22872l@st.omu.ac.jp
** Vietnam National University, Hanoi School of Interdisciplinary Sciences and Arts, Hanoi, Vietnam.

Key words: Landslide, Susceptibility, Machine Learning, Open Source

1. Introduction

Although landslides in Vietnam occur in all mountainous regions, the northern mountainous area is, however, recognized as one of the regions that are most susceptible to landslides in the country (Dieu Tien Bui et al., 2016). According to a study by the Ministry of Resources and Environment of Vietnam, over 70% of communes in Yen Bai province face high or very high landslide risk, particularly in Mu Cang Chai (MCC) and Van Yen (VY) districts (Binh Thai Pham *et al.*, 2016). Landslide Hazard Assessment is a crucial tool in managing and mitigating the risks associated with landslides. The primary product of this assessment, a landslide susceptibility map (LSM), serves as a practical and cost-effective method for zoning areas prone to landslides. This research focuses on implementing Machine Learning (ML) models to create robust LSMs that minimize geographical bias, making them applicable in diverse topographic regions, particularly in areas significantly impacted by human activities.

2. Study Area and Methodology

2.1 Study area and data

The Mu Cang Chai (MCC) located between 21°39′N–21°50′ North latitude and 103°56′–104°23′ East Longitude has been selected as the study area. It is located in the northwest of Yen Bai province of Viet Nam, covering an area of about 1196.47 km^2. Most of the study region is covered by forests (61.76 %) (Binh Thai Pham *et al.*, 2016). The landslide inventory utilized the data from previous project (Truong *et al.*, 2023). From that, landslide and non-landslide points are generated from recorded landslide activities in the area.

2.2 Methodology

Utilizing data from previous research and recorded landslide occurrences in Mu Cang Chai (MCC), a balanced training dataset of landslide inventory is established, incorporating in total 16 contributing factors. Principal Component Analysis (PCA) and Pearson correlation assess the independence and correlation of these factors. Feature importance evaluation removes the least significant correlated factor. Subsequently, four machine learning models–Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), and Extreme Gradient Boosting (XG Boost)–are deployed. Ten-fold cross-validation is employed to tune the hyperparameters, while confusion matrices, accuracy score, Kappa score, Receiver Operating Characteristic Curve (ROC) and Area under the ROC Curve (AUC) are utilized to assess model performance. After training with the MCC dataset, external validation for the ML models is conducted in the Van Yen (VY) district with its own landslide inventory. Additionally, the methodology is also performed in an area within Nagaoka City, Niigata Prefecture, Japan, the areas is much smaller in size with much bigger landslide inventory, giving much high density of training points. This test in Nagaoka assesses the reliability of the models in an area with entirely different topographic, hydrologic, geologic, and anthropogenic features, providing a comprehensive evaluation and identifying potential improvements for ML models.

3. Results and Discussion

In MCC, the performance of the three models (RF, SVM, LR) is positive, but not as high as desired, with an overall accuracy score of 0.7 and a Kappa score of 0.5.

The performance of the four models is put into comparison in Figure 1, using the ROC and AUC for better visualization.

When the models are validated on VY area, the Random Forest model achieves a positive accuracy score of 0.8, demonstrating potential for further application.

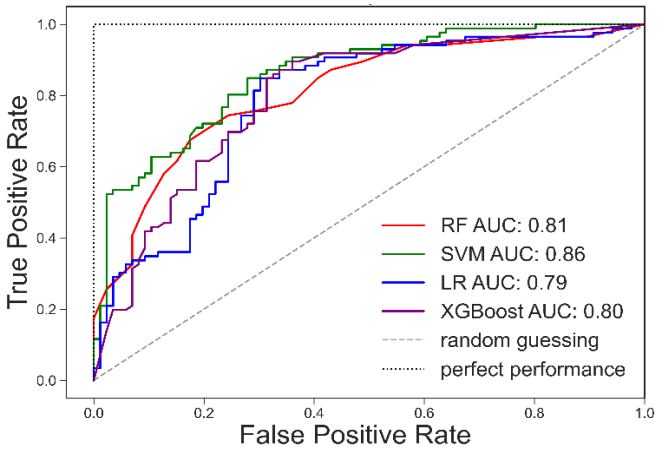

Figure 1. ROC curves for the three models in Mu Cang Chai.

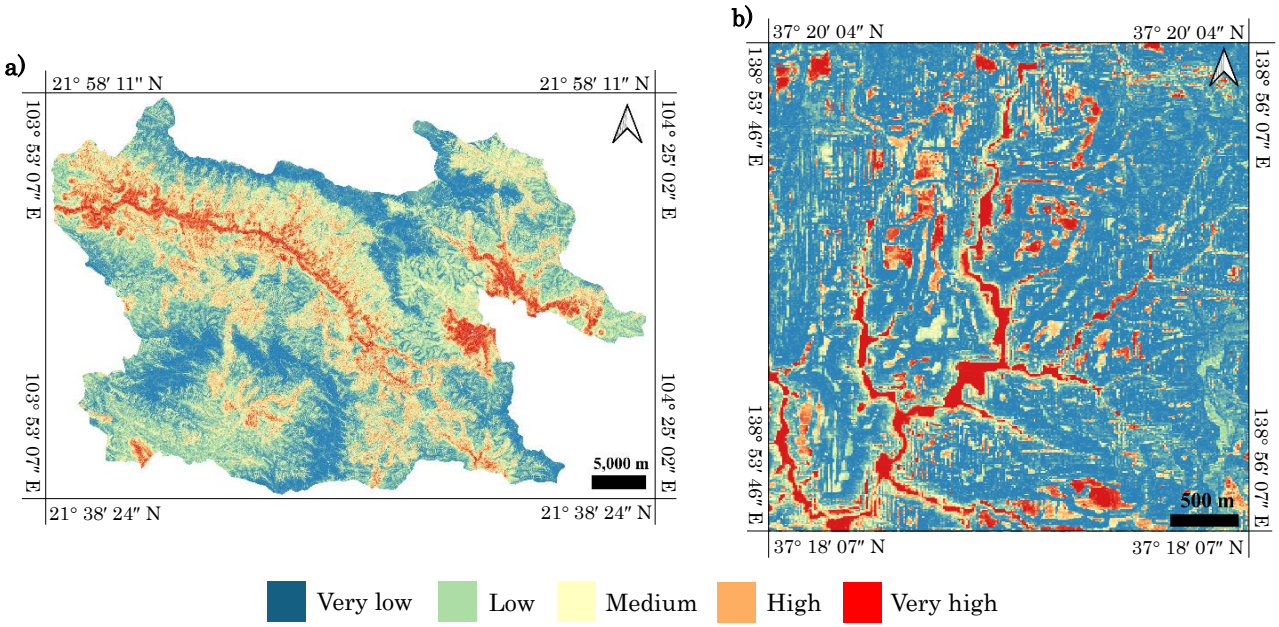

Figure 2. Landslide susceptibility maps for a) Mu Cang Chai; b) The study area within Nagaoka city.

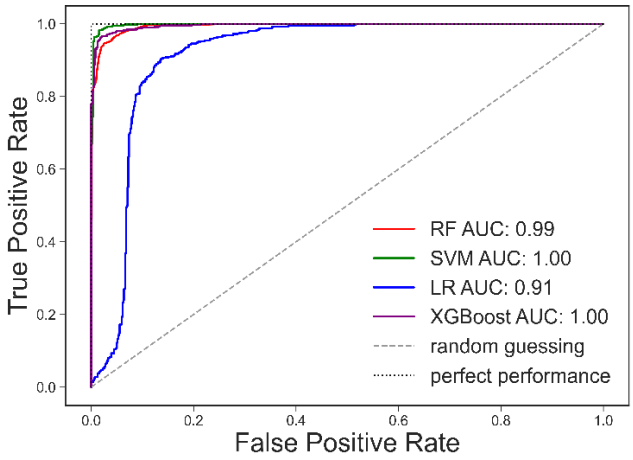

Figure 3. ROC curves for the three models in Nagaoka.

LSM are visualized using equal interval classes across five risk categories: very low, low, medium, high, and very high. This classification approach allows for better understanding of risk distribution (Figure 2a).

In Niigata, Japan, the model performs exceptionally well, with an average accuracy of 0.94 and a Kappa score of 0.87. The evaluation of the ML models is in the figure above (Figure 3).

After validating the framework in Niigata, the ML models trained on Nagaoka data will be applied to another dataset in Ojiya city, Niigata Prefecture. The validation process gives similar results to those obtained in Vietnam, with Random Forest (RF) and Extreme Gradient Boosting (XG Boost) demonstrating consistently positive outcomes. The landslide susceptibility map for Nagaoka is presented in Figure 2b.

ML models offer a promising approach for precise LSM for future landslide prediction and risk mitigation strategies, providing a means to analyze the relationship between different contributing factors. RF is shown to have the most robustness when working with unknown datasets and is the most suited for landslide classification.

4. Summary

Accurate landslide prediction is hindered by the quality of existing inventories, which often rely on past studies. Human activities and urbanization can alter topography, increasing landslide susceptibility, as shown by the importance of road buffers in the PCA biplot and feature selection process. To improve landslide mapping, future research will focus on enhancing inventory accuracy and expanding ML model capabilities by combining different algorithms. Additionally, extensive data validation across diverse geographical areas will provide further insights into model reliability.

References

Bui, D. T., Tuan, T. A., Hoang, N., Thanh, N. Q., Nguyen, D. B., Van Liem, N. and Pradhan, B. (2016) Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. *Landslides*, 1 vol.4, no. 2, pp.447–458. https://doi.org/10.1007/s10346-016-0711-9.

Pham, B. T., Bui, D. T., Pham, H. V., Le, H. Q., Prakash, I. and Dholakia, M. B. (2016) Landslide hazard assessment using random SubSpace Fuzzy rules-based classifier ensemble and probability analysis of rainfall data: a case study at Mu Cang Chai District, Yen Bai Province (Viet Nam). *Journal of the Indian Society of Remote Sensing*, vol.45, no.4, pp.673–683. https ://doi. org/10.1007/s12524-016-0620-3.

Truong, X. Q., H. V., T. Thuy, P. Q. Nhan, P. T. Thanh, T. X. Luan, V. A., B. N. Dung, T. T. Tran, N. T. Thanh, H. Y. N. Thi, D.V. Nam (2023) Integrating Artificial Intelligence and Earth Observation Technologies for Landslide Studying in the Northern Mountain of Vietnam. *Ministry of Science and Technology of Vietnam*. Project No. NĐT/IT/21/14.