Optimizing Machine Learning for Landslide Susceptibility Mapping in Northern Vietnam

Tran Tung Lam, Tatsuya Nemoto, Truong Xuan Quang, Sambuddha Dharand and Venkatesh Raghavan

Department of Geoscience, Osaka Metropolitan University

Geoinformatics Lab

Introduction

Van Yen (VY) district in Yen Bai province is highly susceptible to landslides. Landslide Susceptibility Mapping (LSM) provides a practical and cost-effective tool for zoning areas prone to landslides. LSM expresses in the form of a probability of landslide risk in each pixel. This study implements Machine Learning (ML) models to generate Susceptibility Maps in Van Yen (VY).

Four ML models—Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), and Extreme Gradient Boosting (XGBoost)—were trained in VY area, then validated on the dataset in another area. Evaluation process was conducted using Frequency Ratio (FR), Accuracy Evaluation and Efficient Global Optimization (EGO).

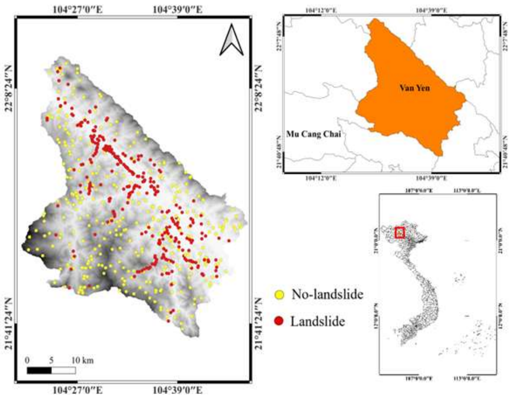

Figure 1: The Study Area

Aim

1. Compare the performance of different ML models for binary classification task, which can be applied for other types of hazard susceptibility mapping.

2. An open-source and Python-based framework for binary classification analysis, including data optimizing and methods for validating ML performance.

Methodology

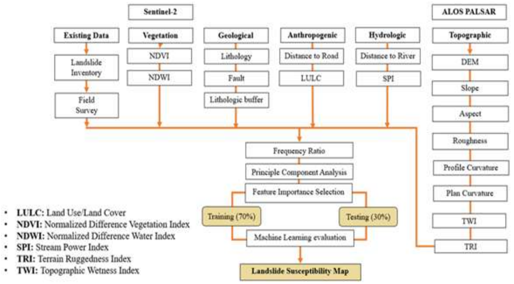

Figure 2: Machine Learning framework

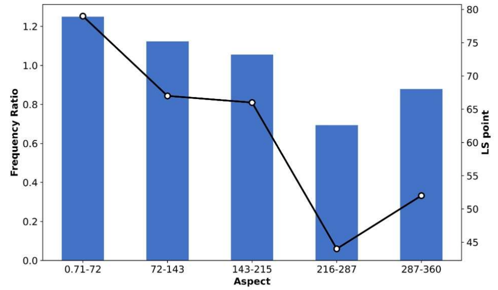

Figure 3: Frequency Ratio compares the density of landslide pixels within a specific factor class to the overall density of landslides

Results

Table 1: ML performance before and after FR

Models	Accuracy Score	Kappa Score	Sensitivity	Specificity
Before				
RF	0.83	0.66	0.76	0.90
SVM	0.88	0.76	0.84	0.92
LR	0.81	0.62	0.72	0.94
XGBoost	0.84	0.68	0.75	0.92
After				
RF	0.83	0.66	0.79	0.87
SVM	0.82	0.63	0.74	0.89
LR	0.82	0.63	0.70	0.94
XGBoost	0.84	0.69	0.87	0.82

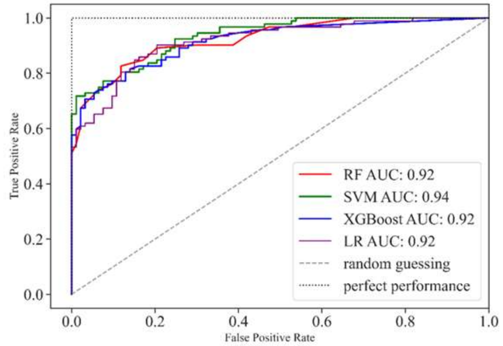

Figure 4: ROC curves for the ML models after applying FR standardization

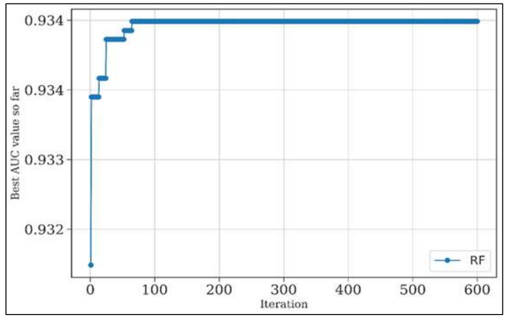

Figure 5: Convergence rate of RF

The Gaussian process provides a Bayesian posterior probability distribution that describes potential values for f(x) at a candidate point x

$$EI_n(x) = E_n\left[\left(f_*^n - f(x)\right)^+\right]$$

Where:

- EI is the Expected Improvement
- f_*^n is the best value so far
- $f(x)$ is the value the function at the candidate point x
- $\left(f_*^n - f(x)\right)$ is positive improvement:

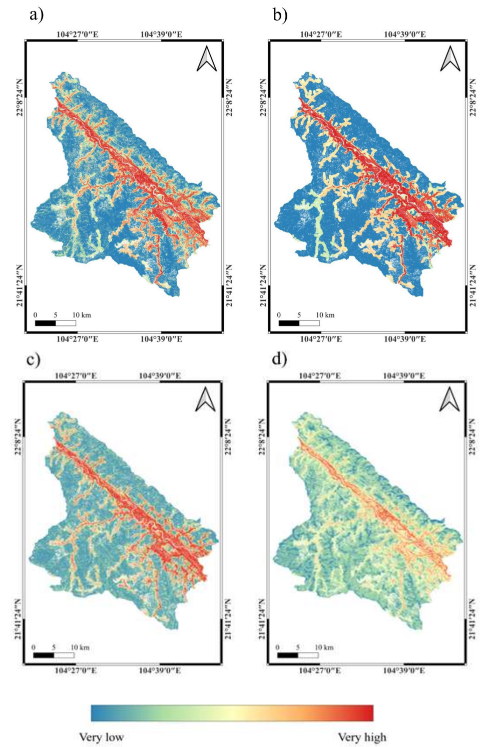

Figure 6: LSM models from: a) RF; b) XGBoost; c) SVM and d) LR

Conclusions

- RF and XGBoost models achieved the highest performance of the four models tested.
- FR method helps prevent situations when models have inconsistent performance.
- The framework can be applied on any other areas for susceptibility analysis.

References

Tran, T., L. Landslide-susceptibility-mapping.GitHub. [Computer software]. Available: https://github.com/Ttunglam/Landslidesusceptibility-mapping-Vietnam. [Accessed: April 15, 2025].