Implementation of Landslide Susceptibility Model Using Machine Learning for Semi Detailed Map Scale in Mountainous Region of Java Island

Imam A. Sadisun,1,2,* Rendy D. Kartiko,1 and Indra A. Dinata,1
1 Applied Geology Research Group, Faculty of Earth Sciences and Technology, Institut Teknologi Bandung, Indonesia
2 Research Centre for Disaster Mitigation, Institut Teknologi Bandung, Indonesia
*Corresponding author: sadisun@gmail.com

Abstract. Landslide susceptibility modeling using neural network (ANN) are applied to semi detailed volcanic-sedimentary water catchment. Annually landslide occurred in catchment area frequently in unconsolidated and weathered material combined with uncertainty in rainfall pattern that complicated landslide occurrence. Data used for analysis including landslide inventory, geology, digital elevation related data, distance to stream, and several other available data. Results show that machine learning method yield fair result data based on evaluation on Area under Curve (AUC). Thus, it can be suggested that machine learning methods for landslide susceptibility model could still be develop to produce robust prediction model with different characterization of parameter data and machine learning parameters.

Keywords: landslide, neural network, support vector machine, AUC, SCAI.

1. Overview
Lately in second decade of second millennium, machine learning and artificial intelligence have become hot issues and commodities. A lot of fields try to grab these concepts and adopt it, either successfully or only partially successful. Research in landslide as part of normal science in predicting occurrence of landslide event is also part of this trend. Latest comparative review [1] with paper citation develop exponentially until around 2018. At least 11 machine model architectures are reviewed by Merghadi, with ensemble (combining methods) have higher overall performance. AUC (area under curve) as one indicator of model performance showed variation of 0.816 to 0.894 based on Merghadi’s validation. Meanwhile, other researchers [2, 3, and 4] also has quite high (even excellent) AUC value, range from 0.87 to 0.982. This pique our interest to see how well this new trend when applied to Indonesian case.

2. Study Area
Small subset of water catchment in West Java had been picked for study area as seen in Figure 1. This area is part of Cidadap dan Cisokan rivers in West Java. Older volcanic peaks are distributed south of study area. Intense geomorphological process occurred in this area, whether rainfall and also as an active tectonic region.
2

3. Analysis
Input data consisted of scale value data (slope angle, NDVI), ordinal data (distance to structure, distance to stream), and also nominal data such as lithology. Input data distribution can be seen in Figure 2. Landslide in the study area consisted of 212 location which are distributed within 94 km².

Artificial Neural Network classification are applied using SPSS multi layers perceptron. Two hidden layers are used for classification. All parameters (6 parameters) are included without any statistic test filter to find overall effect on method. Two main scenarios are conducted in this analysis, first with pure data set and second combined with interval scaled data. As an example, lithology is put as nominal scale value in the first scenario, while in second scenario lithology is weighted with frequency ratio value from landslide distribution within lithology zone.

Binary logistic regression is performed for controlling factor to see how accurate artificial neural network compared to other method, in this case with multivariate method. Logistic regression also uses pre-weighted value for nominal parameters, such as lithology.

4. Result
Area under curve value (AUC) for three scenarios as seen in Figure 3, Figure 4, Figure 5 show that Logistic Regression has the highest AUC with 0.694, followed with pre-weighted ANN (second scenario) with 0.682, and unweighted ANN (first scenario) with 0.612.
Visually, gradation within second ANN scenario is quite jarring as seen in the Figure 4 with distance to structure and lithology has over control for whole region, which made very distinct zones. Gradation in first ANN scenario show graded zone with still distinct zone area easily recognized due to distance from stream parameter. Logistic regression gradation shows more graded zone than both ANN scenario with slope parameters as dominant control of susceptibility zoning.

Figure 3. Landslide susceptibility with ANN first scenario (without pre-weighting parameters).

Figure 4. Landslide susceptibility with ANN second scenario (with pre-weighting parameters).

Figure 5. Landslide susceptibility with Binary Logistic Regression.

5. Discussion
Result of landslide susceptibility within this small set watershed show only moderate AUC (or ROC) value, 0.6 to 0.7. Highest AUC is not even ANN method. This does not directly show that ANN are less powerful in predicting landslide than multivariate method.

Some approach should be analysed later on ANN method, especially in preparing data for analysis and internal adjustment within ANN method itself. It seems that data preparation using pre-weighted data will resulted in higher AUC value but must be inspected visually not just statistically. ANN method itself perhaps need several adjustments such as learning time value, number of hidden layers, and also added more parameters. Thus, this preliminary result perhaps can be expanded later on, and will yield in better AUC and visual distribution of landslide susceptibility zones.

6. Conclusion
• The dominant control in landslide susceptibility in three scenarios is slope which has 0.668 AUC value in unweighted ANN (first scenario), 0.667 AUC value in pre-weighted ANN (second scenario), and 0.668 AUC value in Logistic Regression.
• Area under curve value (AUC) for three scenarios show that Logistic Regression has the highest AUC with 0.694, followed with pre-weighted ANN (second scenario) with 0.682, and unweighted ANN (first scenario) with 0.612.

References
[1] Merghadi A, Yunus A P, Dou J, Whiteley J, Thai P. B, Bui D T, Avtar R, and Abderrahmane B 2020 Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance Earth-Science Reviews 207 103225
[2] Arabameri A, Pradhan B, Rezaei K, and Lee C W 2019 Assessment of landslide susceptibility using statistical-and artificial intelligence-based FR–RF integrated model and multiresolution DEMs Remote Sensing 2019 11 999

[3] Hong H, Liu J, and Zhu A X 2019 Landslide susceptibility evaluating using artificial intelligence method in the Youfang District (China) Environmental Earth Sciences 15 488

[4] Phong T V, Phan T T, Prakash I, Singh S K, Shirzadi A, Chapi K, Ly H B, Ho L S, Quoc N K, and Pham B T 2019 Landslide susceptibility modeling using different artificial intelligence methods: A case study at Muong Lay district, Vietnam Geocarto International 1–24