



GIS based Landslide Susceptibility Mapping using Artificial Neural Network (ANN) Model in South Sulawesi Province, Indonesia

ABDUL RACHMAN RASYID¹, NETRA P. BHANDARY², RYUICHI YATABE² AND IHSAN¹

¹Department of Urban and Regional Planning Faculty of Engineering Hasanuddin University INDONESIA

²Geo Disaster Laboratory Graduate School of Science and Engineering Ehime University JAPAN

Email: rachman_rasyid@yahoo.com

Abstract: Besides the Area Under Curve (AUC) of Receiver Operating Characteristic (ROC) curve to check the accuracy of the models and the predictive curve, the percentage of pixels of landslide susceptibility classes and landslide data for validation that lied within different susceptibility classes is calculated. The weight of causal factors was counted using Artificial Neural Network (ANN) Model and derived by SPSS software. Six causal factors which are curvature, aspect, slope angle, lithology, distance from river and distance from fault were used as independent variables subsequently landslide inventories for dependent variables. Those variables were processed and converted into pixel format using ArcGis software. This study applied ten times trials to find the best weight of causal factors. The results show satisfactory and therefore, the landslide susceptibility map (LSM) can be used to mitigate landslide and landuse planning managements.

Keywords: GIS, ANN model, ROC Curve, Landslide susceptibility, South Sulawesi Province, Indonesia.

1. Introduction

Landslide disasters occur almost every year, especially during the rainy season in Indonesia. Particularly at southern of South Sulawesi province, landslide nearly occurs along rainfall period in Lompobattang Bawakaraeng mountains (Fig. 1). The area of both Mountains is essential to contribute agriculture and hydrology system. After the landslide hit at that Mountain in 2004 and 2006, the concerning to reduce the loss of the catastrophes is increasing, and research related to creating the landslide susceptibility maps (LSM) is an effort towards a solution. In a study of LSM, awareness to find a fit method and simple procedures of analysis became an obsession for the researchers. The probability of landslides occurrence is one of the significant components of the analysis. Susceptibility and hazard assessment are the first step towards risk analysis in an area. Qualitative and quantitative methods have been developed, proposed, and applied for assessing landslide susceptibility and hazard. Determination of a method of analysis depends on the needs of the users, the goal of analysis, the scale of work, quality of input data, availability of time and funding [1]. Qualitative methods are subjective and portray the susceptible and hazard zones in descriptive terms. On the other hand, quantitative methods produce numerical estimates (probabilities) of the occurrence of landslides phenomena in any susceptible and hazard zones. Currently, one of statistical model, which is Artificial Neural Network (ANN), is very popular approaches to create the LSM. The ANN model is inspired by biological systems and consists of three layers (input layer, hidden layer, and output

layer) then sigmoid function as transfer function are selected for mapping.

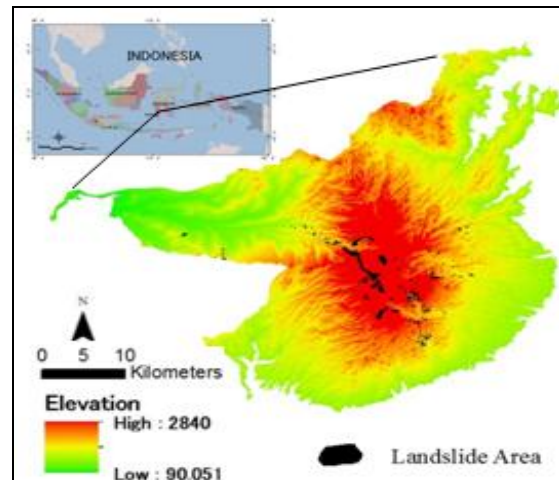


Figure 1: Location map of the study area

Moreover, the best result of the predictive landslide is a requirement when most of the landslide data validation lay on the high class in susceptibility map. The aim of the study is to create the LSM using ANN model at Bawakaraeng Lompobattang Mountain. The results of this study can therefore be used to mitigate landslide and landuse planning managements.

2. Material and Method

Topographic data namely, curvature, aspect, slope angle were derived from ASTER DEM with a spatial resolution of 30 m. Lithology, distance from river and distance from fault, which were collected from the respective governmental institutions. Those data were

referred to as landslide causal factors, and they were used as independent variables. Dependent variable was set up by using landslide inventories from google earth image interpretation and they were divided into two portions. First portion, which were used to create the models, and the second portion were used as data validation. All parameters and landslide distribution map were digitized and then processed by converting all the raster and vector maps into a raster format with 30 m pixel size using ArcGIS 10.0 software. The total numbers of cells of study area are 1,528,838 pixels. The number pixels of landslide data were using to create the models are 6,728 pixels and to validate the models are 1,449 pixels. The ANNs, which is a useful technique for regression and classification problems, has been successfully applied in other fields, and promises to be suitable for the delineation of areas prone to landslide activity. It has been found that ANNs have several advantages for LSM, as these are non-linear and thus has the capability to analyze complex data patterns. The characteristic of artificial neural networks seems to relate to the term of “learns” [2]. It comes up from its ability to learn by adjusting the weights between the neurons in response to the errors between actual output values and target output values. At the end of the training phase from an input value, the neural network represents a model, which should be able to predict a target value [3].

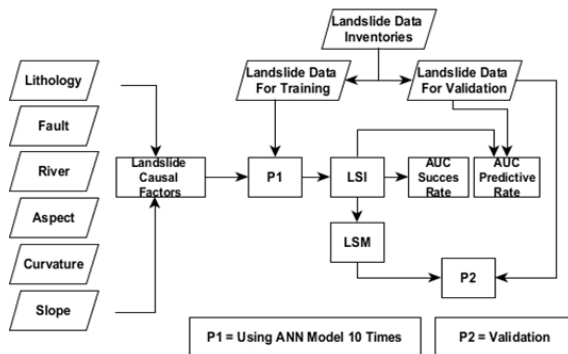


Figure 2: Flow chart showing the whole process using ANN Model and validation

A neural network is a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. Neural networks, with their outstanding capacity to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. Trained neural networks can be thought of as an “expert” in the category of information it has been given to analyze. Other advantages include: adaptive learning, self-organization, real-time operation and fault tolerance via redundant information coding. The ANN can process data at varied measurement scales such as continuous, ordinal and categorical data, a scenario which is often encountered in LSM [4]. ANNs are an attempt, in the simplest way, to imitate the neural

system of the human brain. Besides the ability to handle imprecise and fuzzy data, in the analysis process, ANN deals with continuous or discrete, categorical or any combination of both types (advantages compared with statistical approaches) and binary data without violating any assumptions. It may be considered an ideal application for ANNs because as assessment of probability for landsliding is performed through the forecast of future events from experience of past landslides [2]. Another advantages in analysis process of ANN model in creating landslide susceptibility mapping are there is no need a specific statistical variable for evaluation of model [3], and accurate analysis is possible through a few training dataset because of pixel-based calculation. The application of an ANN starts by identifying the kind of problem that is going to be modelled. In general, ANNs are applied to classification or regression problems. The susceptibility of the territory to landsliding can be seen as a classification problem for most applications. In this way, the ANN outputs can be considered as a sort of degree of membership of each terrain unit to the class of landslide. The ANNs trained could be usefully applied in large areas, in which geological and geomorphological conditions are very similar to those of the test site [5]. The study area was located at southern of South Sulawesi Province in Indonesia and they are famous Mountains. They were surrounded by eight regions and play an important role for agricultural area especially in foot or flat slope. Quarter lompobattang volcanic (qlv) was found as a dominant geological formation and the youngest volcanic products in this area are the Lompobattang Volcanic, which form a broad stratovolcano located in the southeast of South Sulawesi [10].

3. Results and Discussion

In the study of susceptibility mapping, it is important to assume that future landslides will occur in the same condition that caused the past landslides. In general, the procedures using ANN models in landslide susceptibility mapping was begin; 1) the study area was selected 2) Landslide inventories was prepared and its causal factors as well were collected 3) artificial neural networks were trained 4) landslide susceptibility was analyzed, 5) the result was verified and 6) the landslide susceptibility map was created. In this study, we attempt to validate the LSM by overlaying landslide for verification into the map as a final verification. The aim of this process is to calculate the ratio of future landslide fall in high to very high class in the LSM.

In the process of analysis using an ANN model, the merge data of landslide and causal factors will randomly be selected by SPSS software as for training and testing. The training sample comprises the data records used to train the neural network; some percentage of cases in the dataset must be assigned to the training sample in order to obtain a model. The

testing sample is an independent set of data records used to track errors during training in order to prevent overtraining. It is highly recommended that to create a training sample then network training will generally be most efficient if the testing sample is smaller than the training sample. The holdout sample is another

independent set of data records used to assess the final neural network; the error for the holdout sample gives an "honest" estimate of the predictive ability of the model because the holdout cases were not used to build the model. The default values mode are 70% for training and 30% for testing.

Table 1: Weights derived from ANN model from ten trials

Sampling	ILR	Mom	As	Curv	Li	Rv	Sl	Flt	AUC Model	AUC Predictive	(X)	(Y)
1	0.4	0.9	0.064	0.069	0.211	0.119	0.350	0.186	0.948		69.63	7.87
2	0.3	0.6	0.072	0.037	0.342	0.099	0.223	0.228	0.939			
3	0.5	0.7	0.032	0.036	0.220	0.123	0.378	0.211	0.943			
4	0.65	0.9	0.083	0.049	0.260	0.157	0.218	0.234	0.939			
5	0.8	0.5	0.037	0.024	0.262	0.140	0.299	0.237	0.940			
6	0.7	0.4	0.068	0.027	0.220	0.118	0.366	0.201	0.948	0.891	69.70	8.51
7	0.5	0.5	0.038	0.073	0.246	0.124	0.311	0.208	0.935			
8	0.6	0.6	0.015	0.027	0.252	0.150	0.336	0.219	0.934			
9	0.7	0.7	0.049	0.051	0.210	0.118	0.365	0.206	0.939			
10	0.8	0.8	0.127	0.089	0.159	0.101	0.331	0.193	0.940			

ILR is initial learning rate, Mom is momentum, As is aspect, Curv is curvature, Li is lithology, Rv is distance from river, Sl is slope, Flt is distance from fault, (X) %Landslide Fall H+VH, (Y) % Area H+VH

The network used in this study consisted of three layers. The first was the input layer, where the nodes were the elements of a feature vector, the second was the internal or hidden layer, and the third was the output layer that presented the output data. Each node in the hidden layer was interconnected to nodes in both the preceding and following layers by weighted connections [6]. The hidden layer contains unobservable network nodes (units) and serves as the weighted sum of the inputs. The function is the activation function, and the values of the weights are determined by the estimation algorithm. In the process of analysis, the activation function "links" the weighted sums of units in a layer to the values of units in the succeeding layer. The sigmoid function was used in activation and output layer, and takes real-valued arguments and transforms them to the range from 0 to 1. In the case of landslide susceptibility, the input layer includes 6 (six) factors influencing landslide such as slope, aspect, curvature, lithology, distance from river and distance from fault. Figure 2 describes the whole process in this study.

The output layer includes landslide data inventories that contain a single neuron that represents the presence or absence of landslide. The hidden layer is introduced to improve the network ability to analyze complex function. The number of hidden layer may be one or more but generally, one hidden layer is enough to approximate the continuous function in literature. The number of neurons in the hidden layer is not easy to set, but the SPSS software package provides an option to compute the optimum number of nodes in the hidden layer, and it was applied in this study.

To develop an artificial neural network in analysis, the procedure commonly needs to partition the data

into at least two subsets such as training and test data. It is expected that the training data include all the data belonging to the problem domain. Certainly, this subset is used in the training stage of the model development to update the weights of the network.

On the other hand, the test data should be different from those used in the training stage. The main goal of this subset is to check the network performance using untrained data and to confirm its accuracy. Furthermore, no exact mathematical rule to determine the required minimum size of these subsets exists [7]. In this study, the training dataset was set in 80% of the parent database and 20% for testing. These combinations were recommended by [7] and [8]. Back-propagation learning rule in ANN is influenced by initial weights, learning rate, and momentum [9]. Learning rate is put in the neural network process to speed up or slow down the learning that nominally equal to one. There is no universal guideline to select suitable learning rate, and in most of the cases, it is determined by trial and error. A higher learning rate means that the network will train faster, possibly at the cost of becoming unstable and the momentum term helps to prevent instabilities caused by a too-high learning rate. In this study, by using SPSS software to calculate the weight of causal factors, a number of trials were performed to obtain a perfect of learning rate and momentum. According to these trials, learning rate 0.65 and momentum 0.9 was found stable and they used for ANN analysis. The initial weights were set between -0.5 and 0.5. The input data are spatial database of frequency ratio maps. ANN model will produce the weights of landslide parameters that depict the relative importance of each causative factors of landslide. Finally, landslide susceptibility index (LSI) was calculated by summation of each factor's rating

multiplied by the weight of each of the factors in ArcGIS, as seen in Eq. 1.

$$LSI = \sum W \cdot FR \quad (1)$$

Where W is weights of each factor and FR is frequency ratio of each factor class.

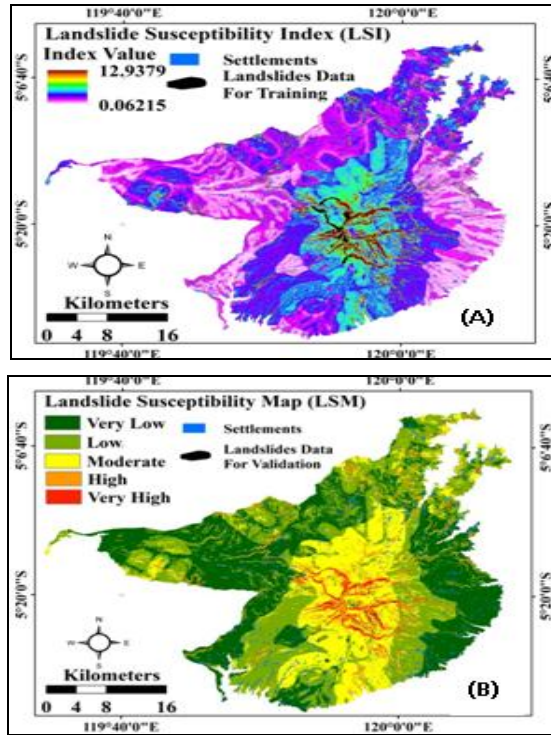


Figure 3: (A) LSI and (B) LSM

The selection of best learning rate and momentum values was started from the LSI. When creating the independent variable of importance in SPSS, the rate AUC of ROC predictive was produced.

The higher rate of AUC curve of model was chosen as the best fit of both learning rate and momentum values. To obtain the best fit of weights of landslide causal factors, we proposed to repeat ten times using different initial learning rate and momentum. One of ANN output is the importance of independent variable, which represents the weight of landslide causative factor, as seen in Table. The results showed that sampling 1 and 6 have higher of AUC curve of model than other sampling that 0.948 which mean 94.8% accuracy. In this case, because the AUC curve is similar, then the process was continuing to the next validation. The sampling no.6 has the best result because the ratio of the landslide fall into the area of high to very high class susceptibility is slightly higher than sampling no.1. It indicates that around 8.51% area of high to very high susceptibility of landslide, the ratio of landslide data for validation which fall was 69.70%. The LSI (Fig. 3a) was produced using the weights values of landslide causal factors from sampling number 6 as in equation below

$$LSI = 0.366 (slope) + 0.220 (lithology) + 0.201 (fault) + 0.118 (river) + 0.068 (aspect) + 0.027 (curvature) \quad (2)$$

The LSI was reclassified to create landslide susceptibility map (LSM) in Fig. 3b using natural breaks method or Jenks optimization method into five classes include very low, low, moderate, high, and very high. To complete the ROC test, AUC predictive was calculated and the results were 0.891 which means 89.1 % accuracy. [11] Classified the AUC value for accepting or rejecting ANN model. The model fails if the area under curve is less than 0.50, acceptable: 0.70-0.80, excellent: 0.80-0.90, and outstanding: 0.90-1.00. Based on the AUC of model and validating datasets or predictive rate, it is easy to conclude that ANN models are acceptable for mapping landslide susceptibility in this study area.

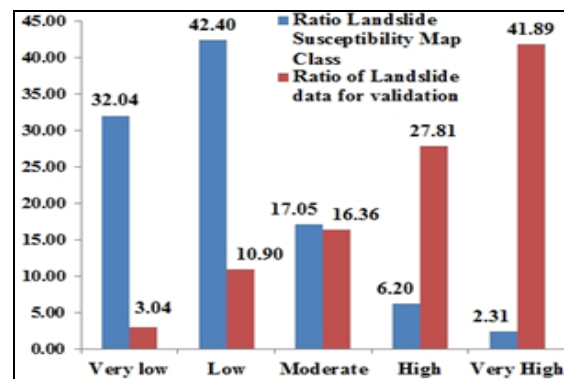


Figure 4: Ratio of landslide susceptibility classes and ratio of landslides data for validation on landslide susceptibility map

Finally, the ratio of landslide susceptibility classes and the ratio of landslide data for validation that lay on susceptibility class illustrate the accuracy of the result (Fig. 4). The reason is that the accuracy of the predicted future landslide that laid on the LSM must have a small ratio value (8.51%) in the susceptibility class of low to very low and higher in the high to very high of susceptibility class (69.70%).

4. Conclusions

In this recent study, landslide susceptibility map was produced by using Artificial Neural Network (ANN) model. The ANN model was based on trial and error to get the best result, therefore this study conducts ten times trials using different initial learning rate and momentum value in process of calculating. Overall, the best result was found in two validation stages. First, by using Receiver Operational Characteristic (ROC) and Area under curve (AUC) of the model. The second stage is the ratio of landslide data for validation that lay on the very high and high class of landslide susceptibility map. All validation tests showed satisfy output. The tests showed that the differences among them are greatly small ranging from 0.934 to 0.948 on the AUC test for landslide data training.

5. Acknowledgements

This research is conducted under collaboration between Hasanuddin University Makassar, INDONESIA and Ehime University JAPAN.

References

- [1] Corominas, J. & Moya, J., 2008. A review of assessing landslide frequency for hazard zoning purposes. *Engineering Geology*. doi:10.1016/j.enggeo.2008.03.018, Volume 102, pp. 193-213.
- [2] Go´mez, H. & Kavzoglu, T., 2005. Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin, Venezuela. *Engineering Geology* 78, pp. 11-27.
- [3] Lee, S., Ryu, J.-H., Min, K. & Won, J.-S., 2003. Landslide susceptibility analysis using gis and artificial neural network. *Earth Surf. Process. Landforms* 28, pp. 1361 -1376
- [4] Kanungo, D., Arora, M., Sarka, S. & Gupta, R., 2006. A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. *Engineering Geology*, Volume 85, pp. 347-366.
- [5] Ermini, L., Catani, F. & Casagl, N., 2005. Artificial Neural Networks applied to landslide susceptibility assessment. *Geomorphology* , Volume 66, p. 327–343
- [6] Choi, J., Oh, H-J., Lee, H-J., Lee, C., Lee, S., 2012. Combining landslide susceptibility maps obtained from frequency ratio, logistic regression, and artificial neural network models using ASTER images and GIS. *Engineering Geology* 124, pp. 12-23.
- [7] Nefeslioglu, H., Gokceoglu, C. & Sonmez, H., 2008. An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. *Engineering Geology* 97, p. 171
- [8] Pradhan, B., Lee, S. & Buchroithner, M. F., 2010. A GIS-based back-propagation neural network model and its cross-application and validation for landslide susceptibility analyses. *Computers, Environment and Urban Systems* doi:10.1016/j.compenvurbys.2009.12.004, Volume 34, p. 216–235.
- [9] Yesilnacar, E. & Topal, T., 2005. Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). *Engineering Geology* 79, pp. 251-256.
- [10] Sompotan, A. F., 2012. *Struktur Geologi Sulawesi*. Bandung Indonesia: Perpustakaan Sasins Kebumian ITB Bandung.
- [11] Hosmer, D. W. & Lemeshow, S., 2002. *Applied logistic regression*. ISBN; 0-471-35632-8. Second ed. USA: Jhon Wiley & Sons Inc.