1. INTRODUCTION

Landslide is a major geological hazard worldwide, accounts for a high number of human casualties and an enormous amount of property loss, and causes severe damage to natural ecosystems and human-built infrastructures [Dai et al. 2002; Guzzetti et al. 2012]. It is necessary to understand the potential exposure to landslide hazard in the areas of mountainous and hilly terrain. The elucidation of the triggering mechanism, characteristics of movement, soil mechanical properties, and the associated geology of landslides can translate to sufficient geologic investigations, geotechnical engineering practices, and ultimately effective enforcement of land management regulation to reduce landslide hazards.

Over the last two decades, many models for landslide susceptibility mapping have been proposed with the assumptions that landslide susceptibility is related to causative factors and can be evaluated as long as the causal relationship is known [Zhu et al. 2014]. Even though the methods for landslide susceptibility mapping can be qualitative and quantitative, it is important to implement [Clerici et al. 2002; Süzen & Doyuran 2004]: 1) mapping of previous landslide inventory in the target region, 2) creation of geological and geomorphological factors that are directly or indirectly correlated with landslides, 3) estimation of the causative factors with the landslides, and 4) classification of the target region into categorical landslide susceptibility (hazard zoning).

During May to August 2012, high-intensity rainfall in Ambon city triggered 89 landslides, most of these landslides happened in municipality area. The damage was severe in the city and at several sites along the transportation network. The landslides resulted in 167 houses destroyed, including 32 deaths, injured numerous people and 305 people evacuated (Fig. 1). Economic losses caused by the landslide events are estimated to be around 25 million U.S. dollars.

In the present study, we aim to assess landslide susceptibility and optimize causative factors using artificial neural network method in Ambon, Indonesia. Elevation, slope angle, slope aspect, lithology, geological density, proximity to river, proximity to faults, and proximity to road networks were chosen as the causative factors. Based on the obtained results, proximity to river and slope aspect were the least influential causative factors in the study, these two causative factors were then eliminated for the optimized landslide model. Proximity to road and geological density were proved to be the most influential causative factors. The six causative factors landslide susceptibility model returned better accuracy when compared to the eight causative factors landslide susceptibility model. The output susceptibility maps were reclassified into five classes ranging from very low to very high susceptibility using Jenks natural break method. 20% of all mapped landslides were used as the validation of the susceptibility models. Receiver operating curves (ROCs) were calculated, the areas under the curve (AUC) for the success rate curves of six factors landslide susceptibility map and eight factors landslide susceptibility maps were 0.770 and 0.734, respectively. The AUC for the prediction rate curve for the six factors and eight factors landslide susceptibility maps were 0.777 and 0.717, respectively.

Key words: landslide susceptibility, Ambon, artificial neural network, causative factors optimization
This study aims to comparatively evaluate the usage of artificial neural network (ANN) to optimize causative factors in landslide susceptibility assessment in Ambon, Indonesia. In this regard, the occurrences of landslides were identified in the study area by field surveys and satellite imagery.

2. STUDY AREA

The study area was located in the Ambon Island at the 3°-4° S and 128°-129° E extending to an area of 377 km² (Fig. 2). The study was conducted in all area of Ambon City which includes five subdistricts: Nusaniwe subdistrict, Sirimau subdistrict, South Leitimur subdistrict, Baguala subdistrict, and Ambon Bay subdistrict totaling to 50 villages.

Ambon has tropical monsoon climate when dry season occurs during December-March, while rainy season occurs during May-October. During the last decade, June has the highest average monthly rainfall counting up to 674.7 mm. When floods occurred in July and August 2013, rain fell every day during the month of July with total monthly rainfall in July 2013 was 1928 mm and maximum daily rainfall was 432 mm.

Varied topography characterizes Ambon; this is shown with elevation variation of 0 m.a.s.l on coastline area and > 900 m.a.s.l for the inner mountainous area. This topographical factor affects the slope distribution throughout Ambon. Slope class of 0 to > 40° is distributed in the entire area of Ambon.

The geology of Ambon is represented by many rock conglomerates, such as alluvial, Kanikeh formation, ultrabasic, Ambon volcanic ash, and Ambon granites (Table 1). The fault lines direction is mainly from northwest to southwest and northeast to southeast. The structural element that was made by the tectonic process is the fault lines, reverse fault lines, and strike-slip fault. Earthquake occurs at the depth 0-99 km with the magnitude up to 7 Richter scale.

3. METHODOLOGY

3.1 Datasets

Independent variables were generated from datasets using ArcGIS® 10.1 SP 1 for desktop. The databases containing the landslides causative factors were prepared in raster format using identical spatial projection and cell size (30 x 30 meter).

Based on the availability of data in the study area, eight landslide causative factors, slope angle (in
degree), slope aspect, elevation (meter above sea level),
geology, geological density (km/km²), proximity to fault lines (meter), proximity to river (meter), and proximity to road networks (meter) were taken into consideration in this study (Fig. 3).

The geomorphological factors were derived from ASTER GDEM including, elevation, slope angle, slope aspect. The geological factors were obtained from the Geological Research and Development Center. The anthropological factors have been collected from Indonesia Geospatial Information Agency. The landslide inventory was produced from the combination of an intensive field survey conducted in January 2015 and analyses from high-resolution satellite imagery.

The total landslide number in the study area is 282 cases which translated to 841 pixels of landslide area. The landslide inventories were divided into training data (80% of total landslide cases, 741 pixels of landslide area) and validation data (20% of total landslide cases, 100 pixels of landslide area).

3.2 Artificial neural networks

Artificial neural networks (ANN) are computational information processing units inspired by the structure and behavior of real biological neurons whose architecture mimics the knowledge acquisition and organizational skills of human brain cells. According to Cilliers [1998], the importance of ANN can be summarized in the following features. They conserve the complexity of the systems they model because they have complex structures themselves. The advantages include that they recognize different sets of data within a whole data set, do not require pre-existing knowledge or experience, do not need a statistical pre-existing model in order to train data and finally, they give reasonable results even when data are inaccurate and incomplete [Jing & Hudson, 2002; Jing 2003].

In this study, multi-layer perceptron (MLP) was applied. MLP is the most popular and most widely used ANN architecture which consists of input layer, output layer, and hidden layers in between. Each layer in a network contains a sufficient number of neurons. The output layer produces the neural network’s results.

![Fig. 2 Landslide training and validation data in the study area.](image)

| Table 1 Ambon regional geology description. |
|---------------------------------------------|
| **Formation** | **Composition** | **Deposition** | **Age** |
| Ambon volcanic rocks | Extrusive; intermediate; lava | Volcanism; subaerial | Pliocene |
| Coral limestone | Sediment; chemical; limestone | Sedimentation; neritic; shallow | Holocene |
| Alluvial | Sediment; clastic; alluvium | Sedimentation; alluvial | Holocene |
| Kanikke formation | Sediment; clastic; sandstone offshore | Sedimentation; neritic; Late | Triassic |
| Ambon granite | Intrusive; granitoid ophiolite | Plutonism; batholith | Phocene |
| Ultrabasic rocks | Tectonite; terrestrial | Sedimentation; Jurassic | |
Thus, the number of neurons in the input and output layers typically depends on the problem which the network was designed for Fig. 4. Each hidden and output layer neuron processes its inputs by multiplying each input ($x_i$) by a corresponding weight ($w_i$), summing the product (Eq. 1), and then processing the sum (if that exceeds the neuron threshold, the neuron is then activated) using a non-linear activation function (Eq. 2) to produce a result ($y_i$), which is the output node.

Fig. 3 Distribution of the eight causative factors used as independent variables in this study. a) elevation; b) slope; c) slope aspect; d) lithology; e) geological density; f) proximity to faults; g) proximity to river; h) proximity to road networks.
with bias vectors $b^{(1)}, b^{(2)}$; weight matrices $W^{(1)}, W^{(2)}$ and activation functions $G$ and $s$. A three layer feed-forward ANN was built. Optimum network architecture (8 inputs x 17 hidden neurons x 2 outputs) was selected following trial and error and considering the minimum mean square error. Initial weights were randomly initiated in a small range. The parameter of learning rate was set to 0.01. In this study, the network training activation transfer function for all layers was hyperbolic tangent.

Causative factors optimization was conducted by eliminating two factors with the least influential effects by the normalized importance value obtained from the ANN analysis.

4. RESULTS AND DISCUSSION

4.1 Artificial neural networks

The normalized importance of each landslide causative factor is presented in Table 2. The ANN analysis was performed with a mean square error of 0.02 in the training set. The final landslide susceptibility map was produced by multiplying each causative factor with the independent variable importance calculated through the ANN analysis (Eq. 3) and then an overlay of these layers was performed.

$$LS_{ANN} = \sum_{j=1}^{n} f_{wi} \times w_{i,j}$$

where $LS_{ANN}$ is the final landslide susceptibility index calculated for each pixel, $f_{wi}$ is the weight of each causative factor and $w_{i}$ is the normalized weight for the category $j$ of the factor $i$.

Based on the normalized importance value obtained from the ANN analysis, the slope aspect and proximity to river were eliminated for the optimized six causative factors landslide susceptibility analysis. Previous study by Pradhan & Lee [2010] showed that ANN techniques is suitable for optimizing causative factors in landslide susceptibility assessment by eliminating the least influential factors.

The produced landslide susceptibility maps were categorized using Jenk’s natural break method into five classes (very low to very high). The produced landslide susceptibility maps are shown in Fig. 5

4.2 Models validation

Susceptibility maps were validated by comparing landslide areas to susceptibility classes, the likelihood of landslide occurrences in a particular region. It was observed that the smaller degree of fit was distributed in the low and very low susceptibility classes. The higher values of the degree of fit were found to be in the high and very high susceptibility classes for the landslide susceptibility maps produced by the three models.

The comparison between the validation and training set overlay analysis of the ANN-derived susceptibility maps show similar results across all susceptibility categories (Fig. 6).

For the 8 factors LS models, the training set data indicates that 79% and 0% of the landslide pixels occur in the very high and very low susceptibility classes respectively and that 21% in the high, moderate, and low susceptibility categories respectively. The validation set shows a total of 63% and 0% of pixels occurring in the very high and very low susceptibility classes and 47% occur in the high, moderate, and low susceptibility classes respectively. It is evident that there is a variation between results produced by the validation and training set for the very high susceptibility class, which shows a difference of 13% whereas a total of 10% represents the difference in landslide pixels in the moderate class.

Table 2 The importance and normalized importance value derived from artificial neural network (ANN).

| Causative factors        | ANN importance | ANN normalized importance |
|--------------------------|----------------|--------------------------|
| Proximity to road        | 0.23           | 100.0%                   |
| Geological density       | 0.19           | 82.8%                    |
| Slope                    | 0.12           | 51.7%                    |
| Proximity to fault       | 0.11           | 47.4%                    |
| Elevation                | 0.11           | 46.2%                    |
| Lithology                | 0.10           | 41.3%                    |
| Aspect                   | 0.09           | 38.1%                    |
| Proximity to river       | 0.06           | 26.8%                    |
susceptibility class. 8% of the landslides in the training data belong to the low categories. These landslide events are located mainly in the outlier of the inner mountainous area of the study area which represents a small number of landslides.

In the case 6 factors LS model, the training set data indicates that 92% and 0% of the landslide pixels occur in the very high and very low susceptibility classes respectively and that 8% in the high, moderate, and low susceptibility categories respectively. The validation set shows a total of 82% and 0% of pixels occurring in the very high and very low susceptibility classes and 18% occur in the high, moderate, and low susceptibility classes respectively. It is evident that there is a variation between results produced by the validation and training set for the very high susceptibility class, which shows a difference of 7% whereas a total of 4% represents the difference in landslide pixels in the moderate susceptibility class.

4.4 Receivers Operating Curves (ROC)

A standard validation analysis to compare prediction performance of various classifiers is the Receiver Operating Curve (ROC) and the calculation of Area Under Curve (AUC) [Akgun et al. 2012; Tien Bui et al. 2012]. The ROC is a useful method for representing the quality of deterministic or probabilistic landslide susceptibility model classifiers. In the ROC graph, the sensitivity of the model which is determined as the percentage of the correctly predicted landslide pixels by the model is plotted against specificity, which is the proportion of predicted landslide pixels over the total study area. The AUC represents the quality of the models to reliably predict the occurrence or the non-occurrence of landslides.

A good fit model has an AUC value from 0.5 to 1.0. The ideal model performs an AUC value close to 1.0 (perfect fit), whereas a value close to 0.5 indicate inaccuracy in the model (random fit), [Carvalho et al. 2014].

Fig. 7 show the ROC of 6 factors and 8 factors ANN models for the training and validation sets. The measurement of how well the model performs is represented in the success rate curve (training data) while the capability of the model to predict is represented in the prediction rate curve (validation data).

It is observed that all the models have good success rate with the highest one being the 6 factors ANN model (AUC : 0.770), while the 8 factors model returns AUC of 0.734. In the case of the prediction rate curve, 6 factors ANN model shows the highest value (AUC : 0.777) whereas the 8 factors model returns AUC of 0.717.

4.5 Landslide causative factors relationship with landslide occurrence

The relationship between the eight causative factors and landslide occurrences are shown in Fig. 8. The landslide frequencies are plotted in the primary axis while the landslide densities are plotted in the secondary axis.

In this study, ANN method was applied to optimize the causative factors of landslide occurrences and to compare the produced LS maps. The general trends, however, dictate that the geological factors and topography be amongst the most significant [Sakellariou & Ferentinou 2005]. The geomorphology
and slope morphology has a definite influence on slope stability and is considered as the most important causative factors related to mass movement activity. Parameter class relationships are evident through the assessment of landslide densities with the occurrence of observed landslides Fig. 8.

Lithology varies regarding chemical and physical properties, which give rise to different levels of susceptibility to landslide occurrence. In this study area, the highest landslide occurrence is found in the Ambon granite formation which comprises of biotite granite and biotite cordierite granite. Ultrabasic formation which comprises of harzburgite, dunite, serpentinite and gabbro follows as the second highest density in the lithology parameters.

In term of geological density the parameter class >4 km/km² showed the highest landslide occurrence which suggests the interlayers of geological structures affects slope stability in a negative way. A consistent increase of landslide density is observed in geological density parameters suggesting that geological density showed high correlation with landslide occurrences. Based on the normalized importance value obtained from ANN (Table 2), geological density is the second most important parameter after proximity to road network (normalized importance value : 0.19). A previous study by Kawabata and Bandibas [2009] suggested that landslide occurrences increase as the density of geological boundaries increases due to unstable relationship between geological substrata. Higher geological density correlates with weaker slope stability which corresponds to landslide occurrences. This phenomenon is evident in the southern part of the study area which has the highest geological density where most of the identified landslides are located.

The highest landslide occurrence is associated with elevation ranging from 200–300 m which occurs mostly along the northeastern and southwestern part of the study area. Steeper slope angles indicate larger driving forces relative to resisting forces which result in a greater potential to fail. In the studied area, most landslides are associated with slope angles ranging from 30°–40° where some municipalities are located. The western-facing slope shows the highest landslide

![Fig. 6](image)

Fig. 6 6 factors and 8 factors LS models relationship with landslide occurrences (a) training data ; (b) validation data.

![Fig. 7](image)

Fig. 7 Receiver operating curves 6 factors and 8 factors artificial neural network methods. Figure 7 a indicates success rate curves and Figure 7 b indicates the prediction rate curve.
density compared to other class in the slope aspect classes. Past study in the area showed similar findings that most of the landslide occurrence happened in the west slope due to soil resistivity contrast between adjacent rocks [Souisa et al. 2015].

Drainage networks impact negatively on landslide susceptibility due to their abrasion processes along the base of slopes, which also result in the saturation of associated material thereby reducing the stability of the slope [Demir et al. 2013]. It is thus expected that more landslides should occur within a limited distance from the stream network. However, this study findings show that the greatest landslide occurrence is associated with the class representing 1000–1250 m from the river network. This might be because of underreported landslide cases found in natural slopes.
Underreported landslide cases might be due to lack of resources to cover extensive and wide landslide area especially in the rainy season of the study area.

The class of > 1500 m of proximity to faults shows the highest landslide occurrences. Even though realistically not associated with the occurrence of landslides, but statistically show high values of occurrence. Closer distances to faults do not show high landslide densities, which may be attributed to that fact that most observed and mapped landslides occur far from the fault lines suggesting that the statistical relationship is based solely on landslide events recorded around a single fault in the interior only.

The highest occurrence of landslides occurs within 250–500 m from the road network. This pattern is expected since the excavation of road cuts reduce the lateral support of material and may trigger landslides. Moreover, this process alters the natural terrain and drainage system. Proximity to road network is regarded as the most important parameter in this study by the ANN model (normalized importance value: 0.23).

The produced LS maps shows good accuracy of > 70%. It is evident from the obtained results that 6 factors ANN model shows higher accuracy for both success rate and prediction rate curves, AUC of 0.770 and 0.777, respectively. In the case of 8 factors ANN model shows returns 0.734 and 0.717 for the success rate and prediction rate curves, respectively. These results agree with past study by Pradhan and Lee, [2010] showing that the soft-computing performance of ANN could be utilized as causative factors optimization method in landslide susceptibility assessments. The relationship between training data and validation data to landslide occurrences also shows acceptable data goodness of fit demonstrated by the consistent increases of landslide pixels per susceptibility classes. Most of the landslide pixels are identified in very high and high susceptibility class in both 6 factors and 8 factors models.

5. CONCLUSION

In conclusion, the current paper provides a comparative evaluation of landslide susceptibility model using artificial neural network by employing optimization on the causative factors. The landslide locations mapped in the studied Ambon city were substantial to the spatial prediction of future landslides. Their relationship with various causative factors has proven to be that of a critical combination. The identification of areas susceptible to the occurrence of landslides is important as it could serve as a preliminary tool in future development planning and for identifying priority areas for early warning systems against potential damage.

Geological factors proved to be critical for all models where they are represented differently being lithological type, proximity to faults, and geological density. In detail, Ambon granite and ultrabasic rocks are exposed to high degree on the susceptibility maps. Proximity to road network was found to be amongst the most influential causative factor with the center portion of the region exhibiting higher densities. In all three maps, the highest susceptibility to landslides was seen in the southwestern part of the city and on the outskirts of the eastern side. These areas are defined by subdistricts such as Nusaniwe and Leitimur Selatan. The resulting susceptibility maps were classified using Jenk’s natural break into five classes (very low to very high susceptibility).

Model performance was tested using an independent validation set comprising 20% of all mapped landslides. For verification of the model performance, receiver operating curves (ROCs) were calculated and the areas under the curve (AUC) for success rate curve were 0.770, and 0.734 for optimized 6 factors and 8 factors respectively. The prediction rate curves AUC were 0.777, and 0.717 for optimized 6 factors and 8 factors respectively.

The results revealed that models showed promising results for shallow landslide susceptibility modeling since they all give accuracies greater than 70%, but the 6 factors model proved to be superior in representing landslide susceptibility throughout the study area. The map derived from this approach is best suited to aid in land-use planning and landslide mitigation. Furthermore, this information can be employed to validate and verify any results acquired at regional and national scale. However, it must be noted that all results obtained are a function of the accuracy of the original database including input causative factors and the landslide inventory. In the study area data availability is one of the major issues in landslide susceptibility assessment. Present study employed eight landslide causative factors, future works would include more data collection to be used as inputs to improve the robustness of landslide susceptibility models in the study area.

ACKNOWLEDGMENT : The author would like to thank Ms. Eva Talakua from Ambon Disaster Management Agency for providing data used in this study.

REFERENCES
Akgun, A., Kincal, C. & Pradhan, B. (2012) : Application of remote sensing data and GIS for landslide risk assessment as an environmental threat to Izmir city (west Turkey). Environmental Monitoring and Assessment, Vol. 184, No. 9,
Carvalho, L. F. et al. (2014): Digital signature of network segment for healthcare environments support. Irbm, Vol. 35, No. 6, pp. 299-309.

Cilliers, P. (1998): Complexity and postmodernism: Understanding complex systems. South African Journal of Philosophy, Vol. 18, No. 2, pp. 258-274.

Clerici, A. et al. (2002): A procedure for landslide susceptibility zonation by the conditional analysis method. Geomorphology, Vol. 48, No. 4, pp. 349-364.

Dai, F. C., Lee, C. F. & Ngai, Y. Y. (2002): Landslide risk assessment and management: An overview. Engineering Geology, Vol. 64, No. 1, pp. 65-87.

Demir, G. et al. (2013): A comparison of landslide susceptibility mapping of the eastern part of the North Anatolian Fault Zone (Turkey) by likelihood-frequency ratio and analytic hierarchy process methods. Natural Hazards, Vol. 65, No. 3, pp. 1481-1506.

Guzzetti, F. et al. (2012): Landslide inventory maps: New tools for an old problem. Earth-Science Reviews, Vol. 112, No. 1-2, pp. 42-66.

Jing, L. (2003): A review of techniques, advances and outstanding issues in numerical modelling for rock mechanics and rock engineering. International Journal of Rock Mechanics and Mining Sciences, Vol. 40, No. 3, pp. 283-353.

Jing, L. & Hudson, J. A. (2002): Numerical methods in rock mechanics. International Journal of Rock Mechanics and Mining Sciences, Vol. 39, No. 4, pp. 409-427.

Kawabata, D. & Bandibas, J. (2009): Landslide susceptibility mapping using geological data, a DEM from ASTER images and an Artificial Neural Network (ANN).

Received: 24 June, 2016
Accepted: 6 December, 2016