GIS-BASED LANDSLIDE SUSCEPTIBILITY ASSESSMENT AND FACTOR EFFECT ANALYSIS BY CERTAINTY FACTOR IN UPSTREAM OF JENEBERANG RIVER, INDONESIA

P. F. Nurdin[a], T. Kubota[b] iD

[a] *Graduate School of Bio-Resources and Environmental Science, Kyushu University, Japan*
[b] *Faculty of Agriculture, Kyushu University, 6-10-1 Hakozaki Higashi-Ku Fukuoka 812-8581, Japan*

Abstract: *This study aimed to assess landslide susceptibility by employing certainty factors model (CF) to select the causative factors for landslide susceptibility mapping in Upstream of Jeneberang River, South Sulawesi. Indonesia. The landslide causative factors were: soil, slope angle, aspect, elevation, lithology, land use, distance to the river, drainage density, and precipitation. For validation purpose, landslide inventory map was randomly partition into two groups, 30% for the validation and 70% for the training. Landslide susceptibility maps were produced by logistic regression using original factor (all nine factors) and selected factor (four factors with positive CF value). The result of certainty factor analysis shows CF value is positive for elevation, land use, slope and drainage density. The accuracy of two landslide susceptibility maps were evaluated by calculating the area under the curve of Receiver Operating Characteristic (ROC) curves. The result shows the the success rate curve for nine factor map (80.2%) is higher than four factor map (78%). But in case of closeness between success rate curve and predictive rate curve, certainty factors model has a closer distance. In this study, effect analysis studies show how the accuracy changes when the input factors are changed.*

1. INTRODUCTION

A landslide occurs worldwide. However, the impact is greater in developing country. The spatial probability of landslide occurrence, also known as susceptibility (Brabb, 1985), is the probability that any given region will be affected by landslides, given a set of environmental conditions (Guzzetti et al., 2005). Many landslide causative factors have been considered in the literature for landslide susceptibility mapping, but it is not certain which factors produce the optimal result for an area under analysis. With the availability of increasing number of landslide causative factors, finding the best combination of factors has become an important research issue. Currently, there are no universal guidelines for the selection of landslide causative factors. Determining the causative factors are a difficult task, Van Westen et al. (2003) stated that every study area has its particular set of factors, which triggering landslides. A factor can be a contributing one for landslide occurrence in an area but not in another one.

Landslide causative factors could have different degrees of effects on the accuracy of landslide susceptibility maps that has been investigated by several types of research in the literature. Cuesta, Sánchez, & Garci'a (1999) assessed 209 landslide events from 1980 to 1994 in the Cantabrian Mountains in northwestern Spain and found that precipitation was the most influential causative factor. Moreiras (2005) considered lithology and slope as the most influential factors in landslide mapping based on the study area of the Rio Mendoza Valley in Argentina. Glenn et al. (2006) stated that topographic factors are highly influential parameters in landslide studies. They assessed the efficiency of laser scanning data (LiDAR)-derived topographic factors in characterizing landslide morphology and activity. For instance,

Costanzo et al. (2012) analyzed the relationships between a priori ranking of controlling factors and predictive performances. The results showed that slope angle, roughness, land use and topographical wetness index were main causative factors.

On 26th March 2004, the huge collapse occurred in northern caldera wall of Bawakaraeng caldera (Elev. 2,830 m). The collapsed area has caused a ridge including Mt. Sarongan (Elev. 2,514 m), and the collapsed mass volume was estimated more than 200 million m^3. The collapsed mass has been running down as debris flow, it is reached the Bili-Bili reservoir and treated life time of the dam. The collapse was predicted to be caused by a combination of several factors; such as weak geological structures, steep high walls of the caldera and high rainfall intensity (Tsuchiya et al., 2004). The main portion of the collapse is two twin ridges of the caldera. These ridges were initially formed by the steep slope surface slipping down or creeping; the creep zone eventually expanded and settled in an unstable condition (CTI, 2006). Low cementation rocks, stock and second deposits of volcanic rocks are distributed at the base of Mt. Bawakaraeng forming a low strength base for the high walls of the caldera (CTI, 2006).

The scope of this study aims to investigate the geo-environmental factors that contribute to landslide and assess the most significant causative factors to generate the landslide susceptibility map with better accuracy. The outline of this study is to select the most significant causative factors by certainty factor analysis, produce landslide susceptibility map using the selected and original causative factors by logistic regression and do the comparison between two models.

2. DATA AND METHODS

2.1. Study area

The study area (Figure 1) is located in the upstream of Jeneberang River (Lengkese Sub-watershed). The study area is bounded by the latitude of 05°18 '10" and longitude 119°53'20" with an area about 128.40 km^2 and most of the terrain is mountainous with highest peaks exceeding 2,795 m. The study area located around 70 km east of capital city of South Sulawesi Province. Jeneberang River rises in Mt. Bawakaraeng, which has an elevation of 2,833 m above MSL and flows from the east to the west. Sulawesi Island has a tropical climate with characteristics of two seasons within a year, rainy seasons from November to May and a dry season from June to October. The precipitation is more than 700 mm in the month of February and rises to 900 mm in January (Tsuchiya et al., 2009). This area is a productive land but regularly experience a small to a big landslide, especially during the rainy season. Due to increasing of rainfall intensity, the probability of landslide occurrence, particularly shallow landslides increases and it is very sensitive to short lasting, high intensive rainfall (Hasnawir and Kubota, 2012).
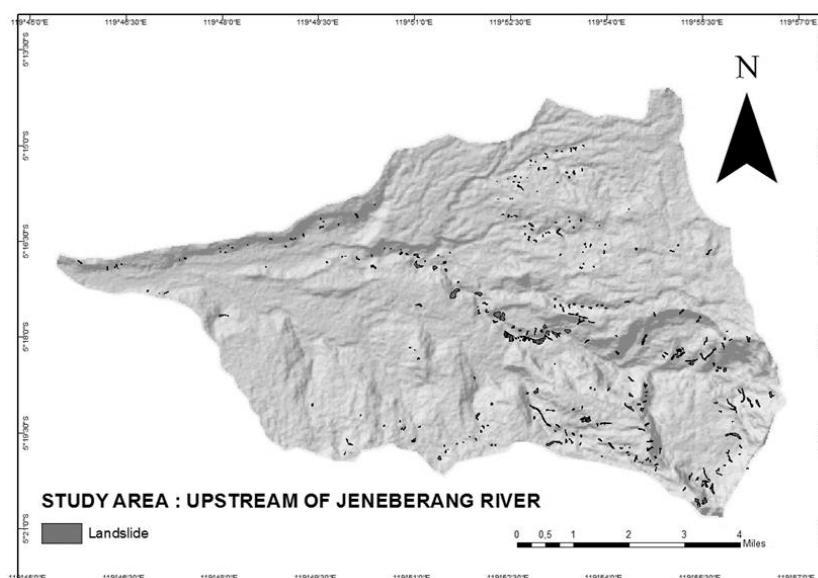

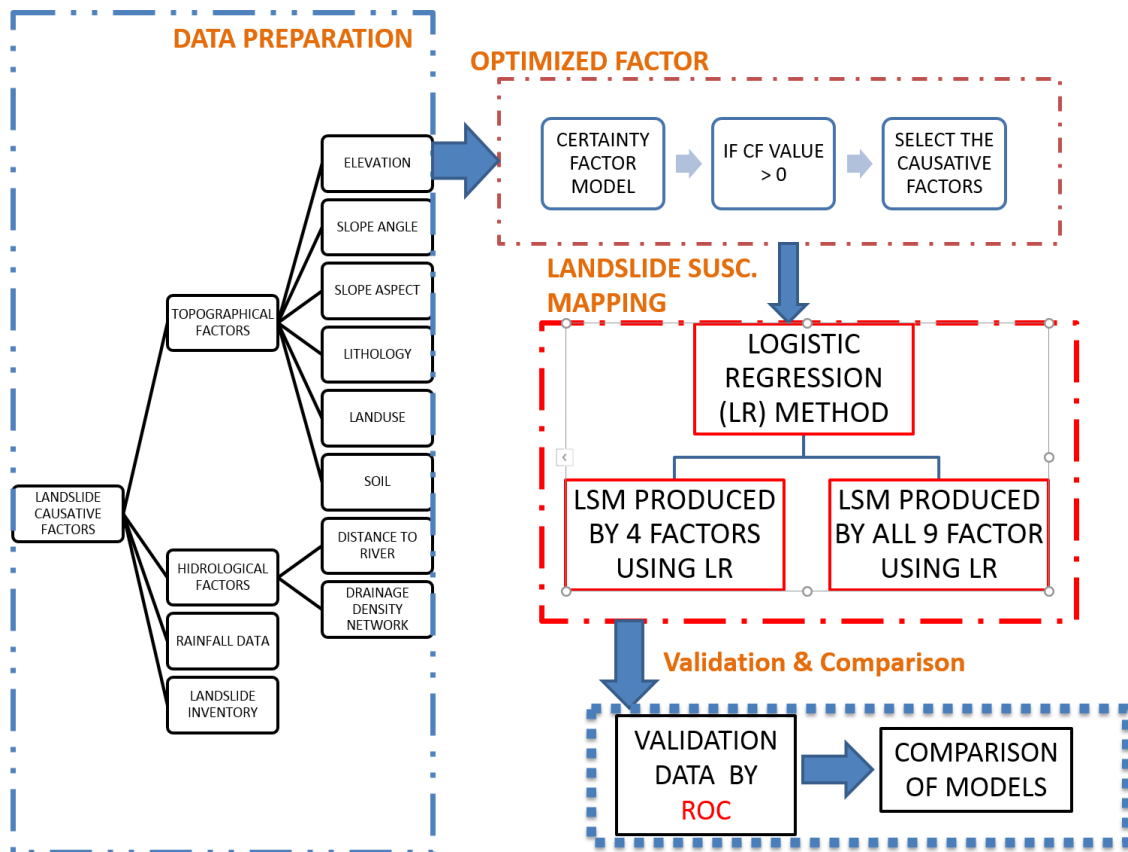**Figure 1.** Hillshaded map of the study area and the landslide inventory

Figure 2. Flow chart of the study.

The lithological layer of the study area was created from geology map of Indonesia with scale 1:250,000. The land use layer was based on Indonesia Ministry of Forestry 2014. River and soil layers were based on BIG (Badan Informasi Geospasial / Geospatial Information Agency). Another factor such as mean annual rainfall data, collected from rainfall gauge stations was available around the study area. Landslide inventory of the area was detected by satellite image interpretation and verified by field investigation.

2.1.1. Landslide inventory map

According to Guzzetti et al. (1999), landslides which occurred in the past and present are keys to predicting landslides happening in future. However, due to the absence of historical records of landslides and their triggering factors, insufficient information and heterogeneity of subsurface conditions and lack of knowledge about their behavior make it very difficult to predict spatial and temporal probabilities of landslides. Indirect mapping methods use either statistical models or deterministic models to predict landslide prone areas, based on information obtained from the interrelation between landslide causative factors and the landslide distribution. Therefore landslide inventory is an essential component of landslide hazard zonation techniques.

A total of 380 known landslides with a total of 9,901 pixels were prepared from field observations and remote sensing of the study area (Figure 1). Some of the archived landslide inventory databases were also used in the previous research to produce landslide hazard map. Landslides in the area include rotational slides and translational slides. For building the susceptibility map models, the landslide inventory was randomly partitioned into two groups: a training data set (70%, 266 landslides) and a validation data set (30%, 144 landslides).

Figure 3. Landslide site in study area

2.2. Geospatial database of geo-environmental factors influencing landslides

Data preparation is the first fundamental and important step for landslide susceptibility analysis. To mapping the potential landslide in sub-watershed Jeneberang, it first conducted studies on the factors that cause landslides. Factors that caused landslides are well known by experts or landslide researchers, but the main factor of avalanche between one watershed and others will vary due to the different of biophysical conditions. In this study, nine landslide causative factors were used, namely: elevation, slope aspect, slope angle, lithology, drainage density, distance to river, soil texture, mean annual rainfall and land cover or land use (Figure 4). Each category was divided into different classes by its value or feature.

Digital elevation model (DEM), remotely sensed imagery and geological maps of the study area were used to create maps of the factors that were employed in subsequent stages. Digital Elevation Model (DEM) was the key to generate various topographic parameters related to landslide activity of the study area. With cell size 10 x 10 meter, elevation (<820 – 2,795 m), slope angle (0 - >45 degrees) and slope aspect layers have been extracted. The resolution and accuracy have a direct influence on the quality of these factors (Lee, 2005).

2.3. Probabilistic analysis

The probabilistic analysis is performed using a methodology integrating the results into a spatial database using GIS.

2.3.1. Certainty Factor Analysis

The certainty factor (CF) model is a method for managing uncertainty in rule-based systems. Shortliffe and Buchanan (1975) developed the CF model in the mid-1970s for MYCIN, an expert system for the diagnosis and medical treatment. In this study, CF is applied to selecting the optimal causative factor related to landslide occurrence. Certainty factor can be calculated using the following functions:

$$CF = \begin{cases} \dfrac{PP_a - PP_s}{PP_a(1 - PP_s)} & \text{if } PP_a \geq PP_s \\[2ex] \dfrac{PP_a - PP_s}{PP_s(1 - PP_a)} & \text{if } PP_a < PP_s \end{cases} \qquad (1)$$

Where PP_a is the conditional probability of having a number of landslides event in a class of parameter and PPs is the prior probability of a total number of landslides in the study area. For each of the causative factors, the weights and contrast were calculated using the certainty factor method. The CF approach transforms each class into interval between -1 and 1, and it indicates a measure of belief and disbelief. A CF value of -1 indicates that an increasing uncertainty of landslide occurrence or the certainty of the hypothesis being true is very small, as compared with a high CF near to 1 means that decreasing uncertainty or the indication strongly supports the hypothesis as true. A value close to 0 means that there is no information of the landslide occurrence. The ppa and pps values were determined by overlaying each parameter layer with the landslide inventory layer in ArcGIS and landslides falling in each parameter class were determined. These values were used to determine the CF value of each classes.

$$Z = \begin{cases} CF1 + CF2 - CF1\,CF2 & CF1,\ CF2 \geq 0 \\ CF1 + CF2 + CF1\,CF2 & CF1,\ CF2 < 0 \\ \dfrac{CF1 + CF2}{1 - \min(|CF1|,\ |CF2|)} & CF1,\ CF2,\ \text{opposite signs} \end{cases} \qquad (2)$$

2.3.2 The Multivariate Approach: Logistic Regression (LR)

It is admitted that among the wide range of statistical methods proposed in landslide susceptibility mapping, Logistic Regression analysis has proven to be one of the most reliable approaches (Ayalew et al., 2005). Logistic Regression analysis relates the probability of landslide occurrence (having values from 0 to 1) to the "logic" Z (where $-1 < Z < 0$ for higher odds of non-occurrence and $0 < Z < 1$ for higher odds of occurrence). In the LR formula, the probability of landslide occurrence is expressed by:

$$P(Y = 1 | x) = \frac{\exp(\sum bx)}{1 + \exp(\sum bx)} x \qquad (3)$$

where P is the estimated probability of landslide occurrence and ranges from 0 to 1; Y is an indicator variable, X is the independent variables (landslide causative factors), $X = (x_0, x_1, x_2, \ldots x_n)$, $x_0 = 1$; b is regression coefficient. To linearize the mentioned method as well as remove the 0/1 boundaries for the original dependent variable, the estimated P probability is transformed by the following formula:

$$P' = \ln\left(\frac{P}{1 - P}\right) \qquad (4)$$

The alteration is referred to as the logit transformation. Theoretically, the logit transformation of binary data can ensure that the dependent variable is continuous and the logit transformation is boundless. Moreover, it can ensure that the probability surface will be continuous within the range [0, 1]. Using the logit transformations, the standard linear regression models can be obtained as follows:

$$P' = \ln\left(\frac{P}{1 - P}\right) = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n + \varepsilon \qquad (5)$$

Here, b_0 is the constant or intercept of the formula, b_1, b_2, . . . bn represents the slope coefficients of the independent parameters, x_1, x_2, . . . xn in the logistic regression and ε is the standard error. Multivariate regression analysis plays a central role in statistics that cause one of the most powerful and commonly used techniques (McCullagh and Nelder, 1989).
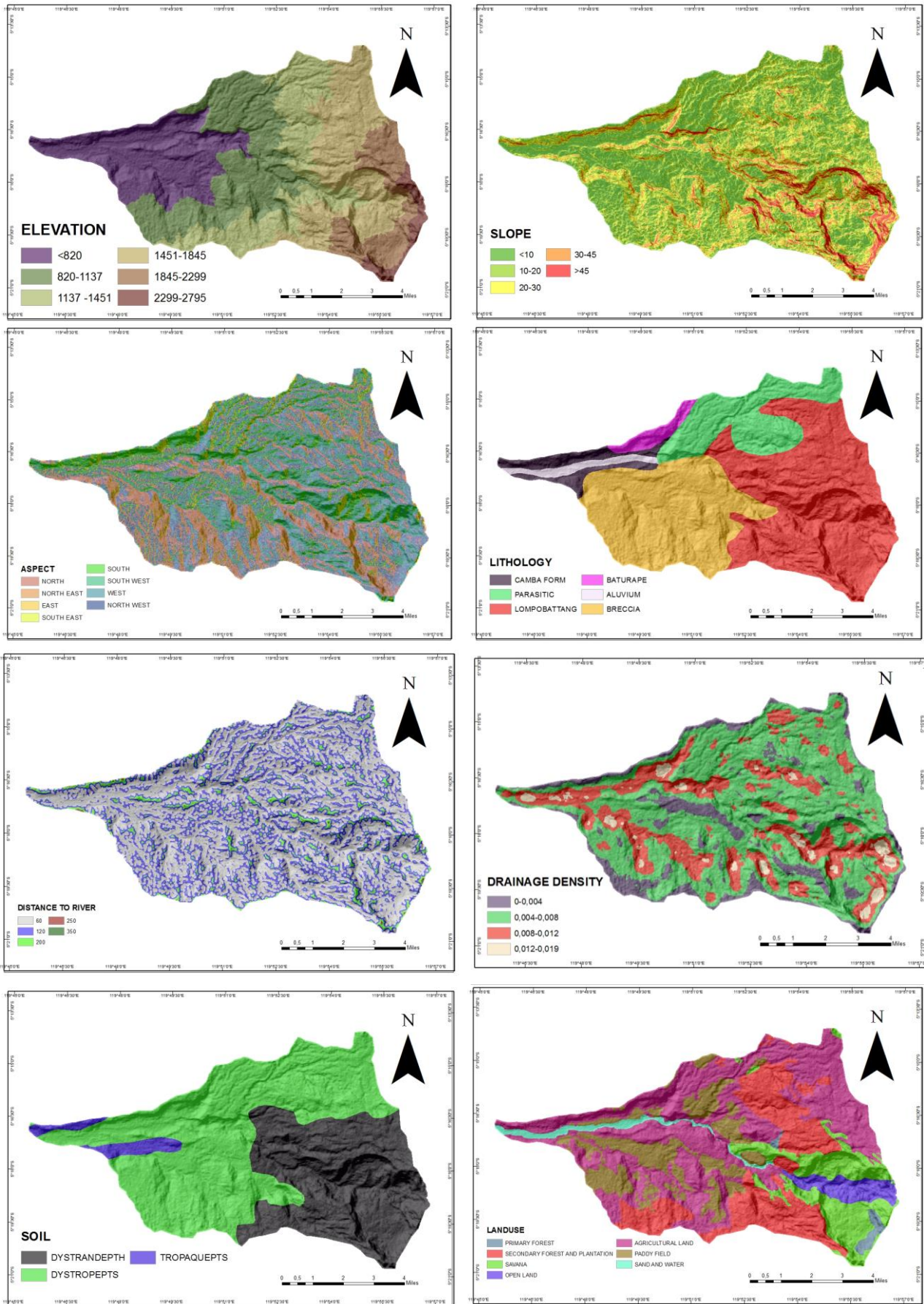
Figure 4. Thematic maps used in this study, (a) Elevation; (b) Slope; (c) Aspect; (d) Lithology; (e) Landuse; (f) Soil; (g) Distance to river; (h) Drainage density; (i) Precipitation.

3. RESULTS AND DISCUSSION

3.1. Physical Factor of Lanslide

3.1.1. Elevation

Elevation is commonly used to assess landslide susceptibility. The variation in elevation may be related to several environmental settings such as rainfall and vegetation variety. A digital elevation model (DEM) can categorize the local relief and locate points of maximum and minimum heights within the terrain. The elevation of the study area is 820 to 2,795 meters. Figure 3 shows that landslides mostly occurred at 820-1,137 m (26%). One of the reason is the land use at those elevation is dominated by agriculture which mostly is paddy field.

3.1.2. Slope angle

Slope is often used to study landslide probability, and all studies into the probability of landslides consider slope (Dai et al., 2001; S. Lee and Talib, 2005). Highly sloped areas and cleared areas receive exposure to direct sunlight, which dries the soil and increases the chances of landslides. The slope angle is frequently considered to be one of the most influential factors for landslide modeling because it influences the shear forces acting on hill slopes (Dai et al., 2001). In the study area, landslides increased with increasing slope steepness. Most landslides occurred on slopes of 30-45 degrees (20%).

3.1.3. Slope aspect

The slope aspect describes the slope direction, and it identifies the downslope direction of the maximum rate of elevation change. Although the relationship between the aspect and the mass movement has been investigated for a long time, there is no general decision regarding the aspect/landslide relationship (Ercanoglu & Gokceoglu, 2004). However, the aspect is a significant factor in producing landslide susceptibility maps (Saro Lee et al., 2004). The slope aspect also plays an important role in exposing the topography to sunlight and drying winds, which control the soil moisture. This is an important factor in landslide studies (Magliulo et al., 2008). According to the number of pixels affected by slope failure, a southwest-facing slope (21,34%) ranks first followed by the north- (19,87%), northwest- (15,91%), and west-facing slopes (14,92%). Moreover, slopes facing south are more prone to landslides because they receive heavy rainfall during the monsoon season.

3.1.4. Lithology

Lithology is the most important parameter in this study of landslides because different lithology units have varying degrees of landslide vulnerability (Cuesta et al., 1999; Dai et al., 2001). The lithological units shown in the surface geologic maps were reclassified according to geology and development center. The result was a generalized geologic map. Finally, the map describes the distribution of six types of lithology:

- TMC (Tertiary Miocene Camba): Marine sediment rocks vary with volcanic rocks, tuff sand vary with sandy tuff and clay stone; and have insertion marl, limestone, conglomerate, volcanic breccias, and coal.
- QLVP, QLV, and QLVB (Quarter Lompobatang Volcanic): Agglomerates, lava, breccias, lahar deposition and tuff.
- TPBV (Tertiary Pliocene Baturape Cindako Volcanic): Lava and breccias, with insertion tuff and conglomerate.
- QAC (Quarte Aluvium): gravel, sand, clay, mud and coral limestone.

The relationship between landslide occurrence and lithological condition showed that QLV has the highest percentage of landslides (63,88%) of the six other lithology classes. QLV is a common volcanic sediment formation in South Sulawesi.
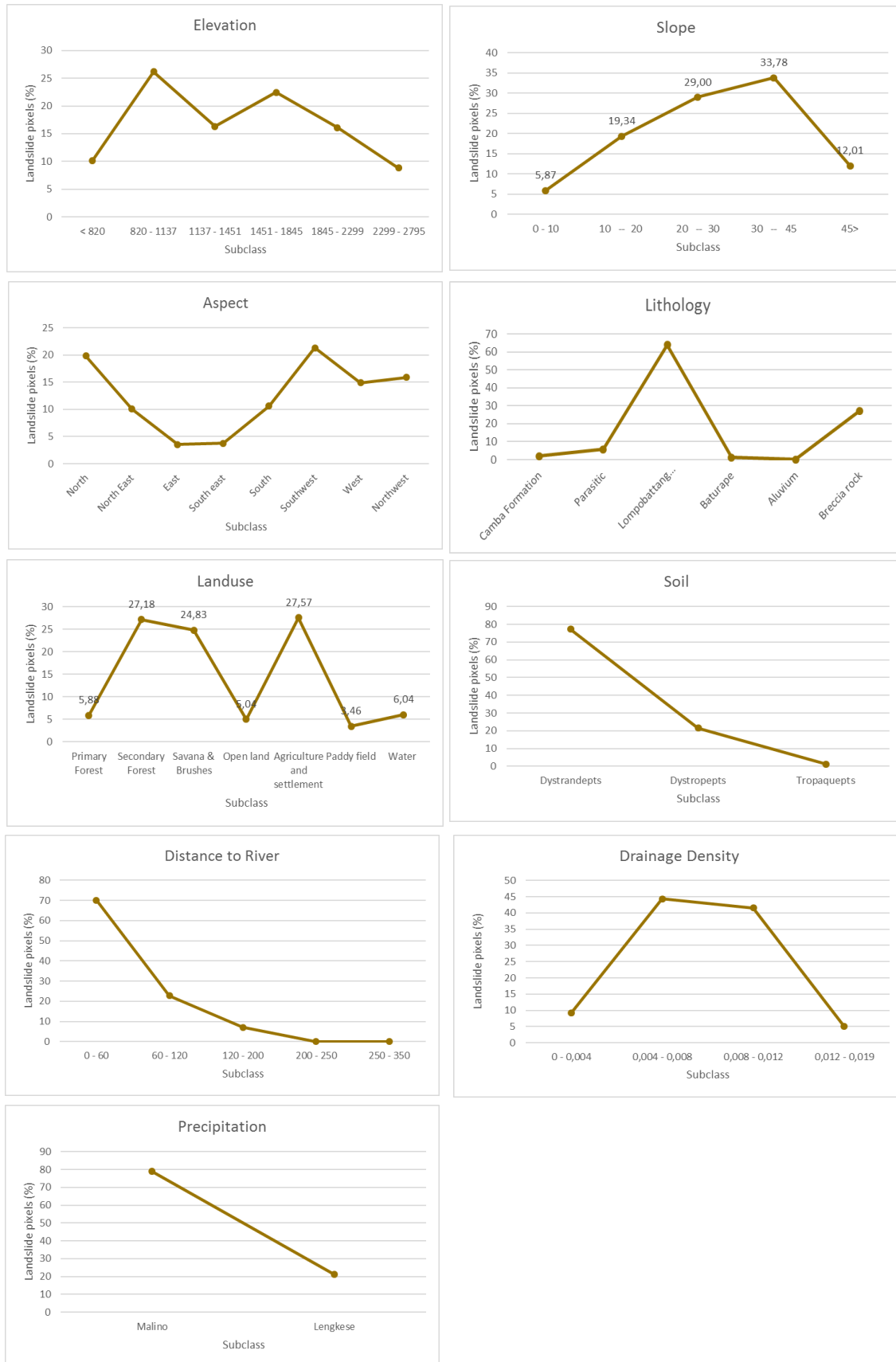
Figure 5. The spatial relationship between landslide occurrence and causative factors. Landslide occurrence correlates strongly with environmental factors that might trigger its mechanism.

3.1.5. Soil

The physical properties of soil are often used for parameter analysis of landslides via a probabilistic approach to soil texture. Soil texture can affect the other physical soil properties such as water infiltration, porosity, and permeability of water and power to pass groundwater. In Indonesia, soil is classified via the United States Department of Agriculture system, the Food and Agriculture Organization of the United Nations (FAO), and the Centre of Soil and Agroclimatic Research. Soil in the study area is from the Andosol family and is divided into three types: Dystrandepth, Dystropepts, and Tropaquepts. Landslides are mostly accumulated on Dystrandepts soil type (77,25%). This might be related to the location of Dystrandepts deposition, which are mostly found at higher altitudes. This class accounts for the largest proportion of the areas.

3.1.6. Drainage density

Drainage density is the total stream length per unit area of a river basin. Hasegawa et al. (2009) noticed that if precipitation increased, then an area with a higher drainage density is more often prone to a shallow landslide. A large-scale landslide is frequent in areas with less drainage density. In this study area, the 0.004-0.008 km subclass has the highest landslide percentage (26%).

3.1.7. Distance to river

Rivers play a major role in modifying the terrain by incising different rocks (Meten et al., 2015). Runoff plays an important role and is a triggering factor in landslides. According to Meten et al. (2015), rivers have a significant role in facilitating landslides. The analysis assessed the influence of distance to river and drainage density on landslide. In the case of the relationship between landslide occurrence and distance to river, subclasses of 0 – 60 have the highest landslide percentage (70%). Gully erosion along the river may initiate landslides. Areas closer to the river network have more erosive forces that erode the base of the slope to a greater degree.

3.1.8. Precipitation

Rainfall is the principal climatic variable that influences landslide distribution. It is affected by topography, elevation, and vegetation—factors that are all interrelated. Mountainous areas cause the air currents to rise and cool resulting in increased precipitation with elevation (Walker and Shiels, 2013). The climate of the study area is tropical humid, and precipitation varies with elevation. The average annual precipitation is between 3,100 – 3,800 mm. The rainfall data area was obtained from two weather stations: Lengkese station and Malino station. Malino has the highest landslide percentage (79%) followed by Lengkese area.

3.1.9. Land use

Land use also plays an important role in the stability of the slope. The land covered by forest regulates continuous water flow. Water regularly infiltrates this area whereas the cultivated land affects the slope stability due to saturation of the covered soil. Land use in the study area is mainly occupied by dry land agriculture, mixed garden, forested area, paddy field, and savannah. Paddy field covered mostly located in the lowland and river floodplain. The landslide area was mostly in agricultural (27,57%) and secondary forest (27,18%).

3.2. Factors selection using Certainty Factor (CF)

The landslide distribution for each class is expressed by the number of occurring pixels and was used to calculate CF values. Table 1 shows the result of Z value of each causative factor. Based on the Certainty Factor method, four causative factors were detected with high influence to slope instability in the study area: land use (z value: 0.82), elevation (z value: 0.56), drainage density (z value: 0.25), and slope angle (z value: 0.30). These four factors have positive relationships with landslide occurrence. Therefore, these four factors were selected for further process to create an optimized landslide susceptibility map. The highest Z value is land use. Landslides especially correspond to the primary forest subclass (CF value: 0.80).

Table 1. The relationship between landslide occurrence and causative factors

Theme	Class	Pixels in class	Landslides	CF	Z
SOIL	Dytstrandepts	508,608.00	7,649.00	0.49	(0.72)
	Dystropepts	732,874.00	2,128.00	-0.63	
	Tropaquepts	42,449.00	124.00	-0.62	
LITHOLOGY	camba formation (TMC)	65,868.00	195.00	-0.62	(1.00)
	parasitic eruption (QLVP)	199,491.00	565.00	-0.63	
	lompobattang volcanic rock (QLV)	572,811.00	6,325.00	0.30	
	baturape volcanic rock (TPBV)	31,343.00	129.00	-0.47	
	aluvium deposition (QAC)	27,945.00	0.00	-1.00	
	breksi (QLVB)	386,473.00	2,687.00	-0.10	
LANDUSE	Primary Forest	15,192.00	582.00	0.80	0.82
	Secondary Forest, forest plantation	279,523.00	2,691.00	0.20	
	Brush & Savana	165,513.00	2,458.00	0.48	
	Open Land	50,810.00	499.00	0.22	
	Agriculture, plantation & settlement	539,506.00	2,730.00	-0.35	
	Paddy Field	201,138.00	343.00	-0.78	
	Water & sand	32,249.00	598.00	0.59	
ASPECT	North	233,598.00	1,967.00	0.08	(0.21)
	Northeast	118,658.00	998.00	0.08	
	East	50,658.00	349.00	-0.11	
	Southeast east	56,486.00	370.00	-0.15	
	Southeast	146,194.00	1,052.00	-0.07	
	South west	198,444.00	2,113.00	0.28	
	West	224,796.00	1,477.00	-0.15	
	Northwest	255,097.00	1,575.00	-0.20	
ELEVATION	<820	281,570.00	1,005.00	-0.54	0.56
	820 - 1,137	323,730.00	2,594.00	0.04	
	1,137 - 1,451	235,103.00	1,615.00	-0.11	
	1,451 - 1,845	278,558.00	2,223.00	0.03	
	1,845 - 2,299	129,473.00	1,592.00	0.38	
	2,299 - 2,795	35,497.00	872.00	0.69	
DISTANCE TO RIVER	0 – 60	818,631.00	6,944.00	0.09	(1.00)
	60 – 120	384,836.00	2,254.00	-0.24	
	120 – 200	78,739.00	697.00	0.13	
	200 – 250	1,579.00	6.00	-0.51	
	250 -350	146.00	0.00	-1.00	
DRAINAGE DENSITY	0 — 0.004	172,338.00	905.00	-0.32	0.25
	0.004 – 0.008	750,490.00	4,384.00	-0.24	
	0.008 – 0.012	319,030.00	4,110.00	0.40	
	0.012 – 0.019	42,073.00	502.00	0.36	
SLOPE ANGLE	0 – 5	115,433.00	157.00	-0.82	1.00
	5 – 10	235,161.00	424.00	-0.77	
	10 – 15	248,266.00	811.00	-0.58	
	15 – 20	209,183.00	1,104.00	-0.32	
	20 – 25	158,647.00	1,293.00	0.05	
	25 – 30	111,267.00	1,578.00	0.46	
	30 – 35	74,494.00	1,373.00	0.59	
	35 – 45	82,741.00	1,972.00	0.68	
	>45	48,739.00	1,189.00	0.69	
Precipitation	malino	947,840.00	7,807.00	0.06	(0.14)
	lengkese	336,091.00	2,094.00	-0.19	

Most landslide cases seen here occurred on natural slopes where vegetation grows perennially. Elevation plays a dominant role in landslide occurrence in the study area. The CF value is positive for intermediate elevation (820-1,137 m) and increased further above 1,451 m. From 2,299 m − 2,795 m, dystrandepts dominate the soil, and landslides are common (CF value: 0,691). It is widely accepted that the slope angle directly influences slope instability. Landslide probability increases from 20 -25 degrees, and it increases further and corresponds to slope angle. The increase in slope angle is affected by the shear stress in the soil or unconsolidated material increases. Gentle slopes generally have a lower frequency of landslides than steep slopes because of the lower shear stresses seen in low gradients (Lee & Talib, 2005). In the study area, most landslides are associated with slope angles greater than 45 degrees (CF value: 0,69).

The drainage density negatively impacts landslide susceptibility due to their abrasive forces along the base of the slope. In the study area, the subclass of 0.08-0.012 km/km^2 has the highest CF value of 0.40 where most shallow landslides are observed. Costanzo et al. (2012) identified the factors based on the ranks associated with the factor's expected contribution to the predictive skill of a multivariable model. Approaches adopting discriminant analysis and logistic regression on the forward selection of variables, however fail when most of the variables are statistically significant. Although this method includes less computation, it requires to categorize the data into landslide and non-landslide groups which is rather exhausted. The proposed model using CF eliminated these limitations because it used only landslide pixels in the computation, and hence is very fast. Prior definition of hazard classes is not required in CF approach and it also supplies advantage of rendering the definition of susceptible classes transparent. Moreover, the proposed model is a relatively straightforward method that allows the causative factors to be ranked according to their certainty values in the range between -1 to 1. It is assumed that positive CF values have a high correlation with the landslide occurrence, and vice versa.

3.3 Landslide susceptibility mapping using logistic regression

In this study, a logistic regression model was developed using an equal proportion of landslide and non-landslide pixels over ten iterations and all non-landslide data as a comparison. The constant and coefficient of independent variables were provided by logistic regression analysis using SPSS. Landslide data were randomly selected by SPSS based on the number of balanced proportion of non-landslide pixels,. Hence, this study proposes to examine ten iterations to get optimal results and a sense of fairness as shown in Tables 2 and 3. By applying a logistic regression model, the landslide occurrence probability was measured. If the values are closer to unity, then landslides are more likely to occur.

**Table 2.** Iteration for all causative factor models

Iteration Data Equal		Drainage density	Distance to river	Soil	Landuse	Lithology	Elevation	Aspect	Precipitation	Slope	Constant	ROC
Iteration 01	B	0.69	1.34	0.87	0.28	-0.42	0.10	0.55	1.44	0.59	-6.06	0.80
Iteration 02	B	0.67	1.03	0.80	0.31	-0.36	0.12	0.74	1.40	0.62	-5.94	0.80
Iteration 03	B	0.60	1.22	0.87	0.31	-0.39	0.08	0.64	1.04	0.64	-6.00	0.80
Iteration 04	B	0.66	1.18	0.89	0.32	-0.37	0.05	0.63	1.47	0.62	-6.10	0.80
Iteration 05	B	0.62	1.09	0.84	0.30	-0.45	0.13	0.71	1.65	0.60	-6.10	0.79
Iteration 06	B	0.67	1.18	0.83	0.30	-0.39	0.12	0.59	1.47	0.62	-6.04	0.80
Iteration 07	B	0.67	1.08	0.86	0.28	-0.41	0.12	0.74	1.55	0.60	-6.11	0.80
Iteration 08	B	0.68	1.06	0.89	0.29	-0.46	0.15	0.72	1.49	0.60	-6.06	0.80
Iteration 09	B	0.69	1.18	0.83	0.31	-0.37	0.12	0.69	1.53	0.60	-6.21	0.80
Iteration 010	B	0.66	1.12	0.84	0.26	-0.46	0.18	0.56	1.38	0.60	-5.77	0.79

Table 3. Iteration for selected causative factor models

Iteration Data Equal		Drainage density	Landuse	Elevation	Slope	Constant	ROC
Iteration 01	B	0.74	0.43	0.33	0.67	-2.62	0.77
Iteration 02	B	0.73	0.45	0.34	0.70	-2.65	0.78
Iteration 03	B	0.67	0.46	0.33	0.72	-2.63	0.78
Iteration 04	B	0.72	0.48	0.31	0.70	-2.66	0.78
Iteration 05	B	0.66	0.44	0.33	0.69	-2.55	0.77
Iteration 06	B	0.85	0.62	0.39	0.71	-3.00	0.77
Iteration 07	B	0.71	0.43	0.35	0.68	-2.61	0.77
Iteration 08	B	0.72	0.44	0.41	0.70	-2.69	0.78
Iteration 09	B	0.74	0.46	0.34	0.68	-2.67	0.77
Iteration 010	B	0.72	0.40	0.41	0.68	-2.65	0.77

Based on logistic regression analysis for all causative factors (Table 2), precipitation has the highest coefficient of 1.47 which means that in the study area precipitation plays the important rules in triggering the landslide. Meanwhile, Table 3 shows that slope plays the most important rules than three others causative factors. In this case, precipitation was excluded because it was not selected by certainty factors analysis (Z= -0.14). The natural break method or Jenks optimization is also known as the goodness of variance fit (GVF) method. It has been used widely especially by planners. It determines the best arrangement of values for the different classes. This method maximizes values between classes and reduces the variance within classes. The five classes include very low, low, moderate, high, and very high values describing the level of landslide susceptibility in the study area. Figure 4 (a) shows the landslide susceptibility map using nine causative factors. The LSM model for nine causative factors was obtained using the coefficient values as the equation 6 :

$$Z = 0.05 \text{ (Elevation)} + 0.62 \text{ (Slope)} + 0.63 \text{ (Aspect)} + (-0.37) \text{ (Lithology)} + 0.32 \text{ (Landuse)} + 0.89 \text{ (Soil)} + 1.18 \text{ (Distance to river)} + 0.66 \text{ (Drainage density)} + 1.47 \text{ (Precipitation)} - 6.10 \quad \text{...........................} \quad \textbf{(6)}$$
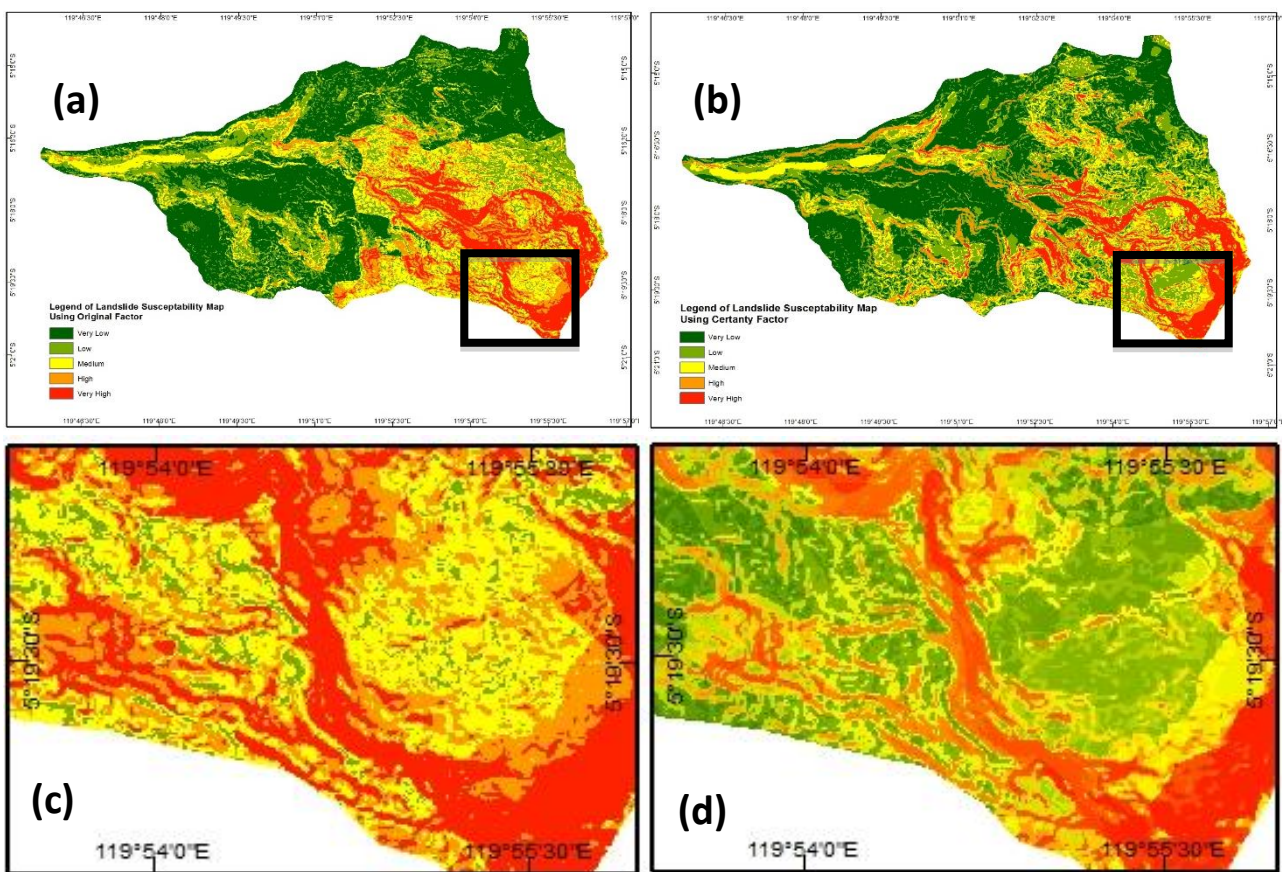
Figure 6. (a) Landslide susceptibility map using nine causative factors generate by logistic regression. (b) Landslide susceptibility map using four causative factors generate by logistic regression. (c) Enlarged image of LSM using nine causative factors. (d) Enlarge image of LSM using four causative factors.

Based on the certainty factor, further LR analysis was conducted using the four highest impact causative factors of landslide occurrences. Optimization was conducted to gain insight into whether the accuracy of the landslide susceptibility map can be increased. The LSM model using four causative factors was obtained using the coefficient values as the equation 7.

$$Z = 0.05 \text{ (Elevation)} + 0.62 \text{ (Slope)} + 0.32 \text{ (Land use)} + 0.66 \text{ (Drainage density)} - 2.63 \quad \text{...........................} \quad \textbf{(7)}$$

Finally, the regression coefficients of predictors were imported to generate the landslide susceptibility map in GIS as shown in Figure 6 (a) and (b). The enlarged images in Figure 6 (c) and (d) facilitate comparison between these two maps. The distribution of medium to high susceptibility areas are much more wide spread in the map of nine causative factors than the four-factor map which is more specific to some location.

3.4. Accuracy assessment of susceptibility maps

Landslide susceptibility maps without validation are less meaningful (Chung and Fabbri, 2012). To validate the landslide susceptibility maps, landslides in the study area were divided into two parts based on random partitions. These partitions divided the area into two groups: prediction (training) and validation (testing). The ROC curve is a graphical representation of the trade-off between the false negative and false positive rates for every possible cut-off value. The area under curve (AUC) is a useful indicator to validate the prediction performance of the model. The success rate curve describes how well the model and controlling factor predict landslides (Chung and Fabbri, 2003). Accuracy is evaluated by the area under the ROC curve. The AUC value lies between 0.5 to 1. An area of 1 represents an excellent classifier and an area of 0.5 represents a worthless classifier.

In this study, both the training data (70% of 380 landslide polygons) and validation (the remaining 30% of 380 landslide polygons) datasets were selected to assess the models. The training data was used for the LSM success rate, and the validation data was used for prediction. The success rate and prediction rate can be obtained by comparing the landslide susceptibility results at known landslide locations. In SPSS software, the AUC of the success rate was derived by linking the landslide index in logistic regression model and the CF model after using landslide data for training. Subsequently, the AUC of predictive rate was obtained using landslide data for validation.
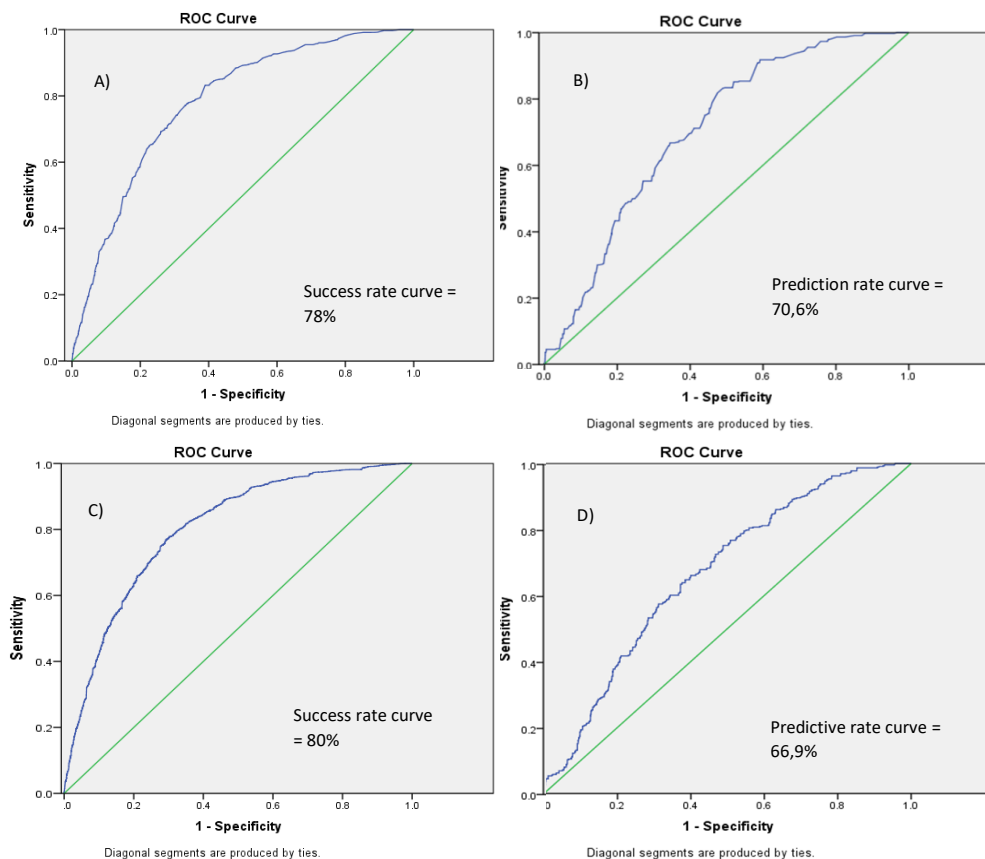

Figure 7. The area under the curve (AUC) represents A) success rate and B) prediction rate curve using four causative factors. C) Success rate and D) prediction rate curve using nine causative factors.

The ROC curves for all nine causative factors or four causative factors are shown in the Figure 6. The AUC of the success rate curve and the predictive rate curve for nine causative factors are 0.80 and 0.67, respectively. It means that the accuracy of the nine causative factors is 80% for training and 66.9% for the validation. In the case of the four causative factors model, the success rate and predictive rate curve return accuracies were 78% and 70%, respectively, with an AUC of 0.78 and 0.70. This proves that the nine causative factor models are better at explaining the cause of landslide occurrences than the four causative factors model.

The AUC curve determined with the validation data set should be approximately equal to the AUC curve determined with the training dataset, but it is generally lower than the success curve because the landslide data about validation areas are not used for modeling (Ngadisih et al., 2013). It is interesting that the similarity of the success rate and the predictive rate values of the four factors models is closer than the nine-factor model. Meten et al. (2015) stated that the proximity of success rate and predictive rate values are also important because it shows how the logistic regression helps predict landslides.

4 CONCLUSION

This study shows the selection of optimum causative factors to build an effective landslide susceptibility map. Four out of nine causative factor was selected by using a certainty factor analysis (elevation, slope, landuse and drainage density). Higher prediction accuracy was obtained from the landslide susceptibility map based on a combination of nine causative factors. The result shows that decreasing the number of causative factors may not always result in higher prediction accuracy. For instance, the combination of nine causative factors showed a higher success rate (80%) than the combination of four landslide factors (success rate 78%). This proved that the landslide susceptibility map from nine causative factors is quite acceptable and should have a greater degree of influence in causing landslides.

There are some limitations and assumptions in this method. The limitations are related to the landslide inventory data. These data did not include the total number of the landslide events within the study area. Furthermore, the output LS map presents only the predicted spatial distribution of landslides and not their temporal probability. Despite these limitations, the produced landslide susceptibility map could be useful to the community and local officials. It could help design future land-use plans and implementation of developments.

5 ACKNOWLEDGMENTS

6 REFERENCES

Ayalew, L., Yamagishi, H., Marui, H., & Kanno, T. (2005). Landslides in Sado Island of Japan: Part {II}. {GIS}-based susceptibility mapping with comparisons of results from two methods and verifications. *Engineering Geology*, 81(4), 432–445. [Crossref]

Brabb, E. E. (1985). Innovative approaches to landslide hazard and risk mapping. *In International Landslide Symposium Proceedings, Toronto, Canada* (Vol. 1, pp. 17–22). [Crossref]

Chung, C.-J. F., & Fabbri, A. G. (2003). Validation of Spatial Prediction Models for Landslide Hazard Mapping. Natural Hazards, 30(3), 451–472. [Crossref]

Chung, C.-J. F., & Fabbri, A. G. (2012). Systematic Procedures of Landslide Hazard Mapping for Risk Assessment Using Spatial Prediction Models. *In Landslide Hazard and Risk* (pp. 139–174). John Wiley & Sons, Ltd. [Crossref]

Costanzo, D., Rotigliano, E., Irigaray, C., Jiménez-Perálvarez, J. D., & Chacón, J. (2012). Factors selection in landslide susceptibility modelling on large scale following the gis matrix method: application to the river Beiro basin (Spain). *Natural Hazards and Earth System Sciences, 12(2), 327–340*. [Crossef]

CTI Engineering. (2006). Report on Urgent Survey for Consulting Engineering Services of Bawakaraeng Urgent Sediment Control Project, Ministry of Public Works, Indonesia.

Cuesta, M. J. D., Sánchez, M. J., & Garci'a, A. R. (1999). Press archives as temporal records of landslides in the North of Spain: relationships between rainfall and instability slope events. *Geomorphology, 30(1–2)*, 125–132. [Crossef]

Ercanoglu, M., & Gokceoglu, C. (2004). Use of fuzzy relations to produce landslide susceptibility map of a landslide prone area (West Black Sea Region, Turkey). *Engineering Geology, 75(3–4), 229–250*. [Crossef]

Glenn, N. F., Streutker, D. R., Chadwick, D. J., Thackray, G. D., & Dorsch, S. J. (2006). Analysis of {LiDAR}-derived topographic information for characterizing and differentiating landslide morphology and activity. *Geomorphology, 73(1–2), 131–148*. [Crossef]

Guzzetti, F., Carrara, A., Cardinali, M., & Reichenbach, P. (1999). Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology, 31(1–4), 181–216*. [Crossef]

Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M., & Ardizzone, F. (2005). Probabilistic landslide hazard assessment at the basin scale. *Geomorphology, 72(1–4), 272–299*. [Crossef]

Hasegawa, S., Yamanaka, M., Mimura, T., Dahal, R. K., & Nonomura, A. (2009). Drainage density as rainfallinduced landslides susceptibility index. In International Seminar on Hazard Management for Sustainable Development in Kathmandu, Nepal (pp. 72–75).

HASNAWIR, & KUBOTA, T. (2012). Rainfall Threshold for Shallow Landslides in Kelara Watershed, Indonesia. *International Journal of Erosion Control Engineering, 5(1), 86–92*. [Crossef]

Lee, C. F., Li, J., Xu, Z. W., & Dai, F. C. (2001). Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong. *Environmental Geology, 40(3), 381–391*. [Crossef]

Lee, S., Ryu, J.-H., Won, J.-S., & Park, H.-J. (2004). Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. *Engineering Geology, 71(3–4), 289–302*. [Crossef]

Lee, S., & Talib, J. A. (2005). Probabilistic landslide susceptibility and factor effect analysis. *Environmental Geology, 47(7), 982–990*. [Crossef]

Magliulo, P., Lisio, A. Di, Russo, F., & Zelano, A. (2008). Geomorphology and landslide susceptibility assessment using {GIS} and bivariate statistics: a case study in southern Italy. *Natural Hazards, 47(3), 411–435*. [Crossef]

McCullagh, P., & Nelder, J. A. (1989). An outline of generalized linear models. *In Generalized Linear Models (pp. 21–47)*. Springer {US}. [Crossef]

Meten, M., PrakashBhandary, N., & Yatabe, R. (2015). Effect of Landslide Factor Combinations on the Prediction Accuracy of Landslide Susceptibility Maps in the Blue Nile Gorge of Central Ethiopia. *Geoenvironmental Disasters, 2(1)*. [Crossef]

Moreiras, S. M. (2005). Landslide susceptibility zonation in the Rio Mendoza Valley, Argentina. *Geomorphology, 66(1–4), 345–357*. [Crossef]

Ngadisih, Yatabe, R., Bhandary, N. P., & Dahal, R. K. (2013). Integration of statistical and heuristic approaches for landslide risk analysis: a case of volcanic mountains in West Java Province, Indonesia. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 8(1), 29–47*. [Crossef]

Shortliffe, E. H., & Buchanan, B. G. (1975). A model of inexact reasoning in medicine. *Mathematical Biosciences, 23(3–4), 351–379*. [Crossef]

Tsuchiya, S., Koga, S., Sasahara, K., Matsui, M., Nakahiro, M., Watanabe, H., … Yoshida, K. (2004). Reconnaissance of the gigantic landslide occurred on Mt. Bawakaraeng in the south Sulawesi state of Indonesia and unstable debris sedimentation (prompt report). *Journal of the Japan Society of Erosion Control Engineering, 57(3), 40–46.*

Tsuchiya, S., Sasahara, K., Shuin, S., & Ozono, S. (2009). The large-scale landslide on the flank of caldera in South Sulawesi, Indonesia. *Landslides, 6(1), 83–88*. [Crossef]

Van Westen, C. J., Rengers, N., & Soeters, R. (2003). Use of Geomorphological Information in Indirect Landslide Susceptibility Assessment. *Natural Hazards, 30(3), 399–419*. [Crossef]

Walker, L. R., & Shiels, A. B. (2013). Large scales and future directions for landslide ecology. *In Landslide Ecology (pp. 227–240)*. Cambridge University Press. [Crossref]