Int. J. of GEOMATE, Dec. 2011, Vol. 1, No. 2 (Sl. No. 2), pp. 99-104
Geotec., Const. Mat. and Env., ISSN:2186-2982(P), 2186-2990(O), Japan

Landslide Susceptibility Mapping by Using Logistic Regression Model with Neighborhood Analysis: A Case Study in Mizunami City

Liangjie Wang, Kazuhide Sawada and Shuji Moriguchi
Department of Civil Engineering, Gifu University, Japan

ABSTRACT: Landslides which affect human lives and economic losses are always attracted a lot of concerning in modern society. In order to identify the potential hazardous areas related to landslides, three methods have been used, such as qualitative or knowledge-based method, deterministic method and quantitative-based method. Geographical information system (GIS) technology and high computing ability provide a convenient tool to deal with landslide triggering factors and make the quantitative-based method achieve effectively.
In this study, landslide-related factors such as topographical elevation, slope angle, slope aspect, topographical wetness index (TWI) and stream power index (SPI), were employed in the landslide susceptibility analysis. The logistical regression was used to obtain the relationships for landslide susceptibility between landslides and causative factors. The distributions of observed landslides were used to evaluate the performance of the susceptibility map. The approaches described in this paper showed us that the logistical regression and neighborhood can be used as simple tools to predict the potential landslide locations. This map will be helpful for city planning, infrastructure construction and agriculture developments in the future.

Keywords: Landslide, Susceptibility map, Logistic regression, GIS

1. INTRODUCTION

Landslides which cause the loss of human life and the damage to the social economy were attracted a lot of attention over the last decades. According a previous study Schuster (1996) [1], landslides will increase in the next decades due to continued deforesting and the changing climatic patterns in landslide-prone areas. In order to identify the landslide-related area, there are three main approaches to assess the landslide susceptibility: qualitative methods, deterministic methods and quantitative methods. In the late 1970s, qualitative approaches were widely applied by engineering geologists to evaluate landslide. Deterministic approaches focus on slope geometry, shear strength data, and pore-water related data (Netra et al., 2010) [2] but lack of taking climatic and human induced factors into accounted. Nowadays, the rapid developments of computer technology and geographic information system (GIS) provide a convenient tool to deal with landslide triggering factors and make the quantitative-based method achieve effectively. Among a lot of quantitative methods, logistic model was recognized as the suitable approach to assess landslide susceptibility because it is free of data distribution and can handle a variety of datasets (Nandi et al., 2009) [3].
In this study, neighborhood analysis “seed cell approach” proposed by (Suzen and Doyuran., 2004) [4] and logistic model were applied to create the relationship between landslides and controlling factors with GIS in Mizunami city ,Gifu prefecture where is well know for landslide hazards.

2. DATABASE CONSTRUCTION

In order to detect the landslide-related factors, some prior knowledge and experiments should be prepared. The main factors include three aspects such as topographic factors water-related factors and human being activity factor are applied in this study. The approach mentioned is based on the assumption that past environmental conditions at the time of landslides can be keys to evaluation the potential sites for landslide in the future. Then, all the landslide-related factors were created and store in the spatial database by using a GIS software with the pixel size or mesh 10×10m.

2.1 Landslide Related Factors

The first database is topographic factors. We create digital elevation model (DEM) derive from a triangulated irregular network (TIN) using elevation points from GSI (Geological Survey of Japan AIST) with spatial resolution 10m. The parameters such as elevation, slope angle, aspect, plan curvature, profile curvature were constructed from DEM. The second database is water-related factors. Those parameters were also obtained from DEM. The topographical wetness index (TWI) and stream power index (SPI) are important parameters of wetness and stream power [5] (Moore et al., 1991) and these are defined as following formulas.

$$TWI = \log_e(A / b \tan\alpha) \tag{1}$$

$$SPI = A \tan\alpha / b \tag{2}$$

Where A (m^2) is the upstream area, b is the size of the mesh expressed as m. and α is the slope gradient. The third database is the human actively factors. The proximate to the highway was used to indicate the human actives induce the landslides. The table 1 shows the significant of the all parameters related to landslides and their spatial distribution patterns are shown in Fig.1.

2.2 Landslide Inventory and Neighborhood Analysis

The landslide inventory map of the study area was downloaded from the organization of NIED (National Research Institute for Earth Science and Disaster Prevention).
The sizes of landslides rage from 0.00002 km^2 to 1.2 km^2, which cover 13.6% of the study area. According to Varnes’s (1978) definition, mass movement like soil slides, debris slides, rock slides and debris flows are incorporated into the term landslides.

Table 1 Landslide causative factors [6]

Data type	Factors	Significance
	Elevation	Climate, Vegetation
	Slope	sub-surface flow velocity
	Aspect	Evapo-transpiration
Topographic	Plan curvature	Converging, soil water content ,soil characteristics
	Profile curvature	Erosion ,Geomorphology
	Tangent curvature	Erosion and deposition
	Solar radiation	Weathering soil moisture
	Flow length	Run off velocity, run off volume
	Flow direction	Runoff velocity. potential energy
	Flow accumulation	Runoff velocity and potential emerge
Water-related	Stream power index	Erosive power of water flow
	Distance to rivers	Susceptible to hill slope undercutting
	Topographic wetness index	Soil water content
Anthropogenic	Distance to highway	Landslide triggering by road construction and vibration

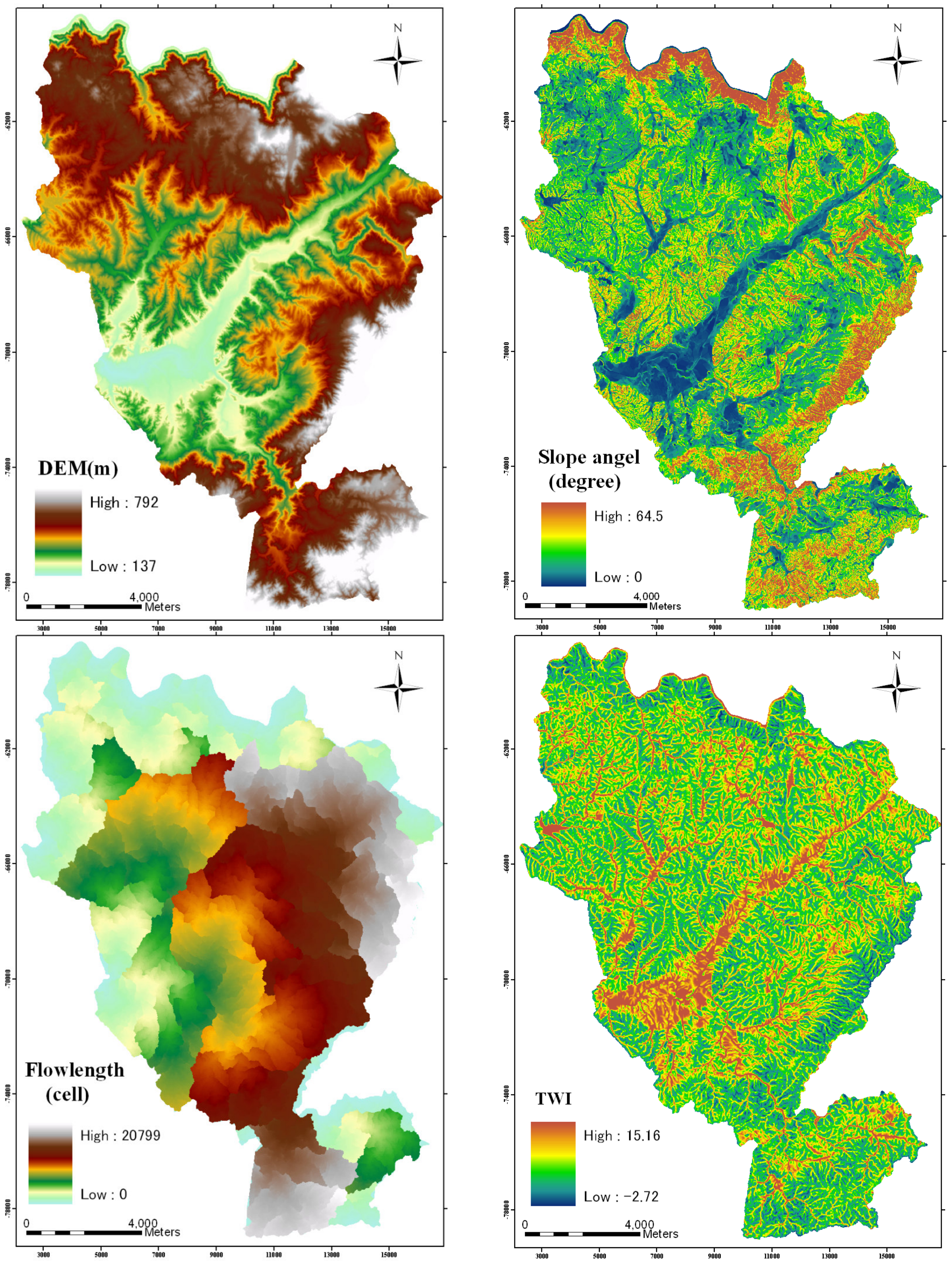

Fig.1. Distribution of landslide causative factors

Neighborhood analysis is a spatial analysis tool used to identify the “seed cells” which represent the undestroyed morphologic conditions before the landslides occurred (Suzen and Doyuran., 2004) [4]. These “seed cells” can be extracted from the boundaries of landslides by buffering zone to the landslides. Former studies show us that the buffer zones should be range from 100m to 150m. In this study, 120m buffer zone is chose to be the “seed cells”

3 LANDSLIDE SUSCETIBILITY MAPPING

3.1 Logistic Regression Model

Logistic regression model is applied to establish the relationship between a dependent variable and independent variables (Atkinsm and Massari., 1998) [7]. This model is careless about the distribution pattern of the independent variables. Most of the topographic factors don’t have normal distributions. Therefore, logistic regression model could get better results comparing other mathematic models. The predicted values range from 0 to 1 can be defined as landslide hazard index. The index can be expressed as following formula.

$$P = 1/(1+e^{-z}) \tag{3}$$

$$z = b_0 + b_1x_1 + b_2x_2 + \ldots + b_nx_n \tag{4}$$

Where P is the probability of landslide occurrence (landslide hazard index), z is the linear logistic mode, b_0 is the intercept of the model, n is the number of landslide causative factors, b_n is the weight of the each factor and x_n is the landslide causative factors.

Our purpose of this research is to propose a model that could be applied in some place where people could not easily get access to or some cities which lack of basic datasets (land cover dataset or digital geological dataset). In this study, most of the landslide-relative factors derive form the DEM that could be easier get from nowadays spatial technology. The causative factors are continuous that would avoid the occurrence of huge redundant data.

3.2 Landslide Points Sampling

Former studies show us that it is the best sampling pattern if the ratio of landslide points to landslide free points is equal to 1(Suzen and Doyuran.,2004) [4]. For this purpose, the ratio is equal to 1 between “seed cells” points that used to extract the attributes from the landslide-causative factors and the random landslide-free points. All of the 11266 points were chosen to create the relationship between landslides and causative factors in logistic regression model. The model would lose his statistical significance if there are no validate points used to identify the efficient of the result. Therefore, the 3518 landslide points occurred in the past were employed to validate the accuracy of the landslide hazard index map. Fig.2 shows us the distribution of landslide points, “seed cells” and landslide free points in Mizunami city.

3.3 Model Results and Validation

The forward stepwise logistic regression model was applied in this study. The probability for stepwise was range from 0.05 to 0.1. And then, $\chi 2$ value of the Hosmer and Lemeshow Test, Cox & Snell R Square and Nagelkerke R Square were applied to evaluate the effectiveness of training datasets. The table 2 shows that the result of the model had statistic significance and the independent variables could explain the dependent variables in this model. During the process of model calculation, the Wald value decreed from 565.8 to 5.3 in all 8 steps. The distance to highway which had the highest value of Wald value implied that this landslide-related factor was the most significant value for explaining landslide occurred, followed by distance to rivers, flow length, elevation, slope angle, solar radiation, and flow accumulation. The relationship between the causative factors and landslide could be shown in table 3. According to table 3, slope angle, flow length and solar radiation were positive relationship with the occurrence of a landslide whereas flow accumulation, distance to highway, elevation and distance to rivers indicated negative relate to the occurrence of a landslide.

Using the weights shown in table 3, the landslide susceptibility index (LSI) of landslides was calculated (Fig.3).

The landslide hazard index decreased from 0.94 to 0.06. If the value is close to 1, it indicates landslides are more likely happen in this area.

The accuracy of the logistic regression model of landslide susceptibility was evaluated by calculating relative operative characteristic (ROC) and percentage of observed landslides points in various susceptibility categories [3] (A.Nandi and A.shakoor., 2009). The area under ROC cure (AUC) value which represents the quality of the probabilistic model by describing its ability to predict the occurrence or non-occurrence of an event was applied in this paper (Yesilnaca and Topal., 2005) [8]. The value of AUC is range from 0.5 to 1. If the AUC value close to 1, that means high accuracy of prediction model and if the AUC value close to 0.5 which indicates the inaccuracy of the model (Fawccett. 2006) [9]. In this study, we obtained the AUC value was 0.69, which implied the reliable correlation between causative factors and landslides. Also the observed landslide points were used to identify the accuracy of the landslide probability map. We divided the landslide hazard index map into five chasses (Fig.4): very low ($0.05<LHI\leq0.1$), low ($0.1<LHI\leq0.3$), medium ($0.3<LHI\leq0.5$), high ($0.5<LHI\leq0.7$) and very high ($LHI>0.7$).

From Fig.5, 9.55% of observed landslide points were found in very high susceptibility class. 6.76% of the landslide points occur in the areas with very low and low susceptibility classes. Near 60% of the landslide points were concentrated in the area with high susceptibility class. The hypothesis of the method was that a landslide would occur in area with at least medium values of susceptibility or high and very high susceptibility values (Bai et al., 2010) [10]. 93.2% of landslide points appeared in the area with susceptibility classes from medium to very high. Among the whole Mizunami city, 12% of the study area was designated as very high susceptibility zone which had the

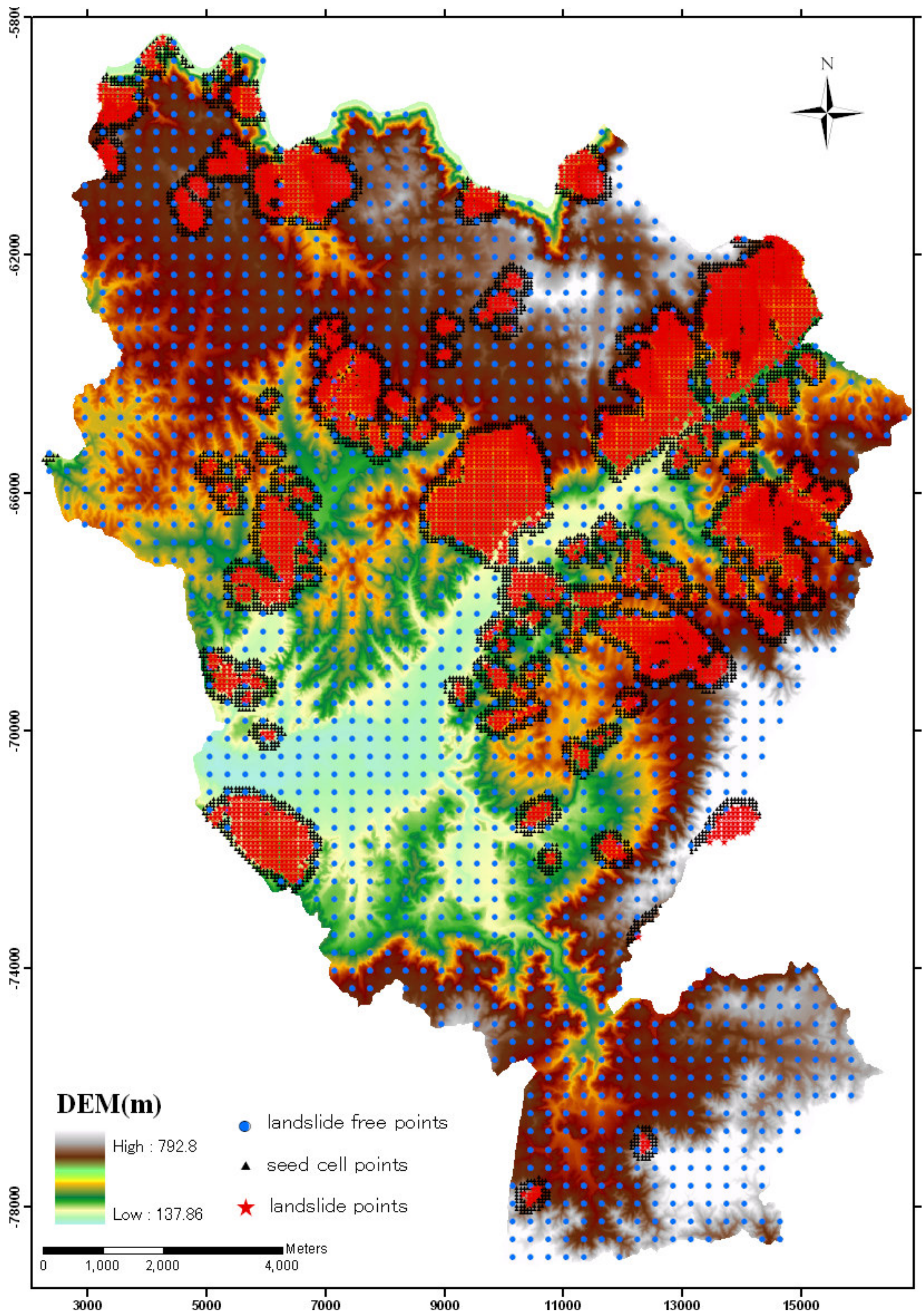

Fig.2. Distribution of sampling points

Table 2 Statistics of logistic regression model with landslide causative factors

Hosmer and Lemeshow Test			-2 Log likelihood	Cox & Snell R^2	Nagelkerke R^2
χ^2	Sig.	Sig.			
173.664	8	.025	14598.819	0.286	0.315

Table 3 Coefficients of the logistic regression model for the seven landslide-related factors

Independent variables	Coefficient
Slope angle	0.01877
Flow length	0.00006
Flow accumulation	-0.000003
Solar radiation	0.000002
Distance to highway	-0.00007
elevation	-0.00289
Distance to rivers	-0.00044
Constant	-0.19102

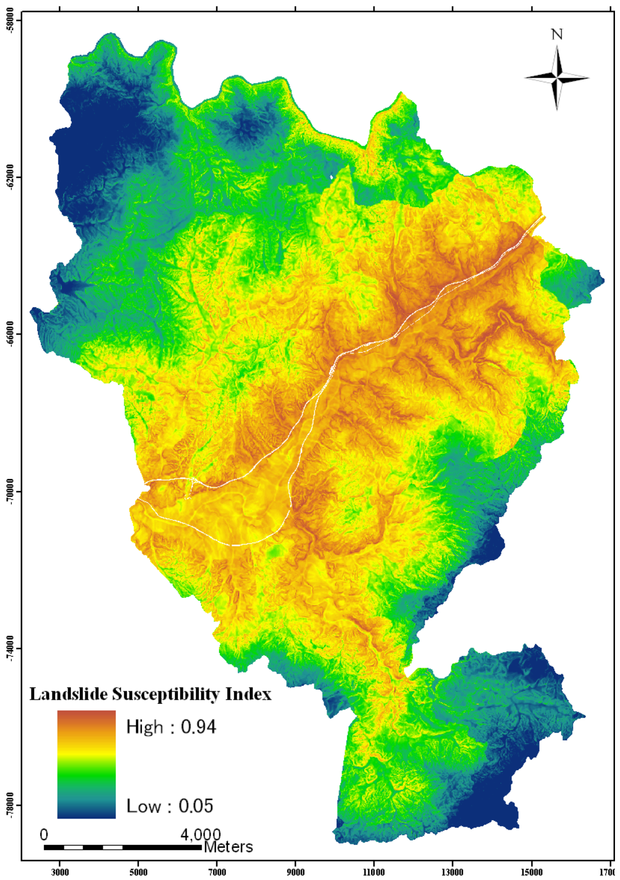

Fig.3. Landslide susceptibility index map

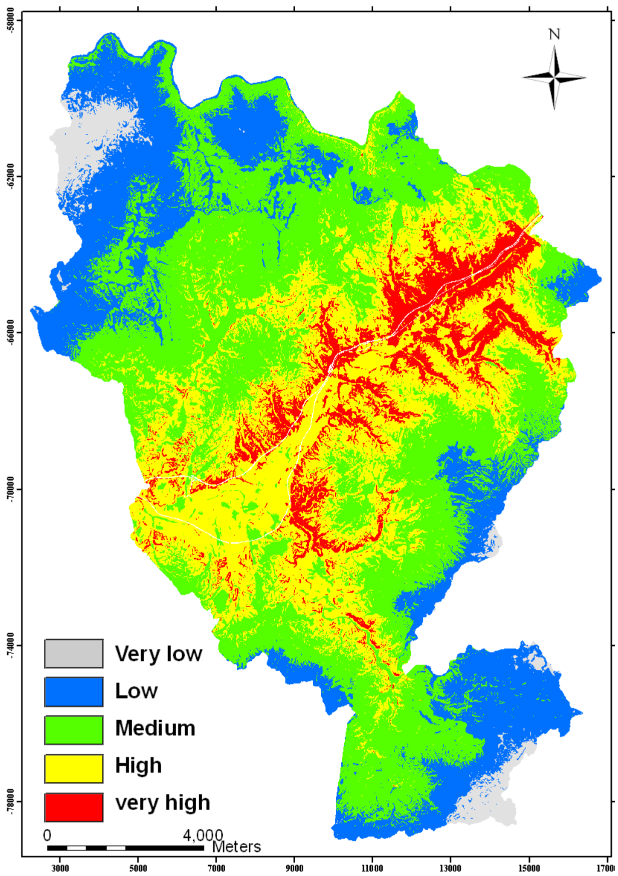

Fig.4. Reclassified landslide susceptibility map

approximately equal percentage of the former landslide area (covering 13% of the study area) indicated that the dividing approach was appropriate in Mizunami city.

4 CONCLUSION

Landslides are very complexity phenomenon in the nature world. The reasons cause landslides appearance are hardly be understood. Although some landslide causative factors could be obtained by exploiting advanced spatial technology, some factors remain unknown. There are many quantitative methods used to create the relationship between landslides and causative factors. And then the landslide susceptibility maps are applied for mitigating the geo-hazard in the world. In this paper, logistic regression model which is simple and careless about the distribution pattern of the independent variables was closed to build the landslide susceptibility map by using GIS tools. Most of the causative factors were derived from DEM which was easier acquired nowadays. Seed cells which extracted from the boundaries of landslides were employed to represent undestroyed geomorphologic conditions. The seed cells points were entered into the logistic regression model as training datasets and the observed landslides points were applied as the test dataset to evaluate the accuracy of the model. The results of landslide susceptibility analysis showed us that 70% of the observed landslides points were located in high and very high susceptibility categories. The model seems to be reliable in Mizunami city. The accuracy and prediction ability of the landslide susceptibility map could offer us crucial information for city planning, infrastructure construction and agriculture developments in the future or in other area with similar conditions.

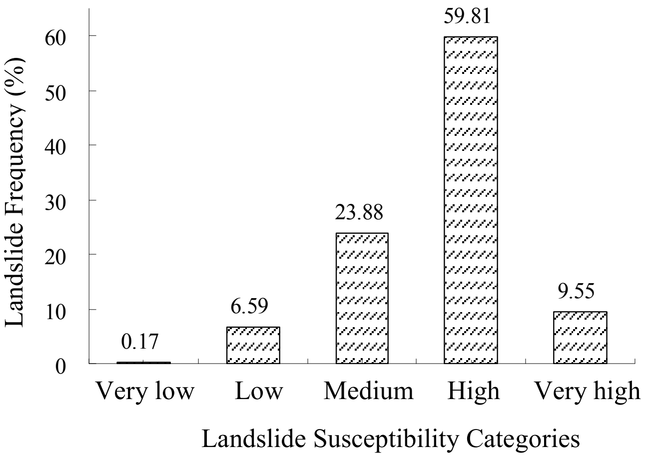

Fig.5. percentages of observed landslide points falling into different susceptibility categories

5 REFERENCES

[1] Schuster, R., Socioeconomic significance of landslides. In: Turner, A.K., Schuster, R.L. (Eds.), Landslides: Investigation and Mitigation, Special Report, vol. 247. National Academic Press, Washington, DC, 1996, pp. 12–36.

[2] Netra, R. R., John, R.G., John, D.V., Assessing susceptibility to landslide: Using models to understand observed changes in slopes. Geomorphology, 2010, 122 (4), 25–38.

[3] Nandi, A., Shakoor, A., A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses. Engineering Geology, 2009, 110, 11–20.

[4] Süzen, M.L., Doyuran, V., Data driven bivariate landslide susceptibility assessment using Geographical Information Systems: a method and application to Asarsuyu catchment, Turkey. Engineering Geology, 2004, 71, 303-321.

[5] Moore, I.D., Grayson, R.B., Ladson, A.R., Digital terrain modeling – a review of hydrological, geomorphological, and biological applications. Hydrological Processes, 1991, 5, 3-30.

[6] Regmi, N.R., Giardino, J.R., Vitek, J.D., Modeling susceptibility to landslides using the weight of evidence approach: western Colorado, USA. Geomorphology, 2010, 115, 172–187.

[7] Atkinson, P.M., Massari, R., Generalized linear modeling of susceptibility to landsliding in the central Apennines, Italy. Computers and Geosciences, 1998, 24, 373-385.

[8] Yesilnacar, E., Topal, T., Landslide susceptibility mapping: a comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). Eng. Geol. 2005, 79, 251-266.

[9] Fawcett, T., An introduction to ROC analysis. Pattern Recognition Letters, 2006, 27, 861-874.

[10] Bai, S.B., Wang, J., Lü, G.N., Zhou, P.G., Hou, S.S., Xu, S.N., GIS-based logistic regression for landslide-susceptibility mapping of the Zhongxian segment in the Three Gorges Area, China. Geomorphology, 2010, 115, 23-31.

International Journal of GEOMATE , Dec. 2011, Vol. 1, No.2 (Sl. No. 2)
MS No. 2c received Dec.1, 2011, and reviewed under GEOMATE publication policies.

Pertinent discussion including authors' closure, if any, will be published in the Dec. 2012 if the discussion is received by June 2012.
Corresponding Author: Liangjie Wang