The Use of Machine Learning for Accessing Landslide Susceptibility Class: Study Case of Kecamatan Pacet, Kabupaten Mojokerto

Listyo Yudha Irawan1*, Sumarmi1, Syamsul Bachri1, Damar Panoto1, Nabila1, Irfan Helmi Pradana2, Rahmad Faizal3, Melinda Meganagatha Rosbella Devy4, Widodo Eko Prasetyo1

1Department of Geography, State University of Malang, Jl. Semarang 05, Malang 65145, Indonesia

*listyo.fis@um.ac.id

Abstract. Kecamatan Pacet, Kabupaten Mojokerto is one of an area with many landslide events in East Java Province. As a mitigation effort, this research aimed to map the landslide susceptibility class distribution of the research area. This research applied a machine learning analysis technic which combined Frequency Ratio (FR) and Logistic Regression (LR) models to assess the landslide susceptibility class distribution. FR bivariate analysis is used to normalized the data and to identify the influence significancy on each class of triggering factors. LR multivariate analysis is applied to generate the landslide probability (susceptibility) and to show the influence significancy of each triggering factor to landslide events. There are 12 triggering factors to landslide used in this research, which is: TPI, TWI, SPI, slope, aspect, elevation, profile curvature, distance to drainage, geological unit, rainfall, land use, and distance to the road. This research has 383 landslides and 383 non-landslide events as the data sample based on field survey, BPBD Kabupaten Mojokerto, and Google Earth Pro imagery interpretation. The proportion of dataset training and testing is 70% and 30%, which generated from the data inventory. This research used ROC analysis to validate the landslide susceptibility model. The result showed that the landslide susceptibility model has an AUC value of 0.91, which indicated that the model has high accuracy.

1. Introduction

Indonesia is a country with a high rate of landslide events. In 2017, landslide events were causing 163 fatalities, 185 people injured, and more than 59,641 people relocated nationally [1]. With a total of 848 occurrences in the same year, landslides are the third-highest disaster events in Indonesia after floods (979) and whirlwinds (886) [1]. Respectively, the province with the highest to lowest landslide events in Java Island is Central Java, West Java, East Java, and Special Region of Yogyakarta (D.I. Yogyakarta) Province [1], [2]. Among the four provinces, East Java Province has the highest number of casualties caused by landslides (Figure 1).

![Figure 1. Diagram of Total Fatalities and People Lost Caused by Landslide](Source: BNPB (2018) Modified)

Kecamatan Pacet, Kabupaten Mojokerto is one of an area with many landslide events in East Java Province. Therefore, it needs a landslide detection in Kecamatan Pacet, Kabupaten Mojokerto as a response to the high susceptibility in the area. Furthermore, the detection effort needs to be translated
into a more communicative form for the community. Providing information about landslide susceptibility class can further increase the community preparedness to the events [3-5]. Accordingly, this research tried to map the landslide susceptibility class in Kecamatan Pacet, Kabupaten Mojokerto.

This research applied a machine learning analysis technic to assess the landslide susceptibility class. The analysis technic involved both statistical and soft computing approaches. In general, the principal work of machine learning is collecting data to create a prediction. This paper used machine learning to enhance the accuracy of landslide susceptibility class assessment. It is because of the ability of machine learning technic to analyze the complex relationship between landslide susceptibility and its triggering factors [5], [6], [7], [8], [9], [10]. Some of the machine learning analysis technic model to look landslide events probability (landslide susceptibility class) are; artificial neural network [9], [10], support vector machine [8], [11], random forest [6], [12], logistic regression [8], [13], and others.

The logistic regression model is used in this research given the high accuracy to assess a landslide susceptibility. As a consequence, this model is the most used model to assess landslide susceptibility in the world [14]. The recent trends in similar research also include the data normalization process. The process is applied to increase the prediction result [8], [11], [13], [15]. Bivariate analysis is used to normalized the data thus the influence of every triggering factor to landslide can be seen. To validate the result of landslide events probability prediction, ROC curved is used. All of the mentioned processes are used to generate appropriate landslide modeling.

2. Data Preparation and Methods
There are four phases to generate landslide events probability (susceptibility) using machine learning, which are: 1) spatial dataset preparation, 2) data normalization, 3) landslide modeling, and 4) model validation (Figure 2).

2.1 Study Area
The study area (Kecamatan Pacet) area covers an area of 17,300 km², composed of volcanic landform (Figure 3). Kecamatan Pacet located within a typical tropical monsoon region with 2,518 mm/years. The BPBD of Kabupaten Mojokerto stated that there were 45 landslide events from 2015 to 2020 in the area. There were several impacts regarding the events which varied in different parts of the research area, such as; cut transportation connections, destroyed farmlands, and damaged houses.
2.2 Spatial Dataset Preparation

2.2.1 Landslide Data Inventory

There were three ways to do the landslide data inventory in this research, which are: 1) field survey in 2020, 2) landslide events records from BPBD Kabupaten Mojokerto data collection and 3) 2011—2020 Google Earth Pro imagery interpretation. The data collection resulted in 383 landslide events shown in Figure 3. There was a variability of landslide length in some of the landslide spots in the field—the shortest was 2 meters while the longest was 17 meters.

The use of machine learning to model the landslide probability can be considered as a typical binary classification [16], [17]. As a consequence, the exact same number of non-landslide spots are also needed [18], [19], [20], [21]. Landslide events are symbolized as “1” and non-landslide events are symbolized as “0”. This research used supervised machine learning analysis to predict the landslide and non-landslide events into the dependent variable.

2.2.2 Landslide Triggering Factor

There were twelve triggering factors to landslide used in this research, which includes topographic, geologic, hydrologic and topographic conditions. Topographic parameters were TWI (Topographic Wetness Index), TPI (Topographic Position Index), SPI (Stream Power Index), slope, elevation, profile curvature and distance to drainage [22], [23], [24]. The geologic parameter was geological unit [24], while hydrologic parameter was rainfall intensity [25]. For the anthropogenic parameter, this research accounted land use and distance from the nearest road. Distance from the nearest road was used on the account of some landslide events which are triggered by a cut slope.

The topographic parameters were generated by ALOS Palsar DEM using Q.GIS 3.21 (elevation, distance to drainage and slope) and SAGA GIS 2.3.2 (TWI, TPI, SPI, aspect and profile curvature). Factor from the lithological condition was generated by geological unit classification from 1:50.000 Geological Map of Pandaan Page, published by PVMBG (Pusat Vulkanologi dan Mitigasi Bencana)

Figure 3. Landslide Distribution in Pacet, Mojokerto
Source: Authors (2020)
Geologi or Center of Volcanology and Geological Hazard Mitigation). Average rainfall intensity for the last 10 years (2010—2019) from Dinas Pekerjaan Umum Kabupaten Mojokerto (The Public Works Service of Kabupaten Mojokerto) was used as the hydrological parameter in this research. Digitation was done in order to generate the land-use factor of the anthropogenic parameter, while the road data of Kabupaten Mojokerto from Ina-Geoportal was used as the basis of distance to road factor.

Each factor should be in a raster format. Therefore, some shapefile data must be converted into a raster with the same spatial resolution of ALOS Palsar DEM which is 12 meters. Not only every raster-format spatial data must encompass the same coordinate system, but also the spatial resolution and pixel numbers [4], [9], [13], [26], [28]. If the spatial resolution and pixel numbers are inconsistent, the analysis process is rejected.

Figure 4. Landslide Triggering Factors (a. TWI; b. TPI)

Source: Authors (2020)

Topographic Wetness Index indicated the hydrological processes related to the water flow accumulation based on its slope control [28], [29], [30]. Accordingly, the water flow accumulation will increase as the TWI score is also higher—vice versa. The distribution of the TWI class in the research area is mostly classified as moderate to low, which can be seen in Figure 4a.

Topographic Position Index (Figure 4b) displayed the altitude difference between a certain area with the surrounding area [31]. Henceforth, the valley, slope and ridge of a terrain can be seen from the TPI value [24]. Gradually, the valley of terrain has the smallest TPI value and the ridge has the highest TPI value.

Stream Power Index (Figure 5a) indicated the potential erosion as an illustration of a geomorphologic process [24]. SPI visualized the water flow energy which affected by the potential flow erosion in a certain location [32]. As the SPI value increased, the slope erosion risk was also increased [32]. In the research area, the highest SPI value was located in the upper to lower slope of Anjasmoro and Welirang Mountain areas.
Distance to drainage (Figure 5b) can affect the infiltration, runoff and soil moisture rate. As the distance to drainage is closer, the infiltration, runoff and soil moisture rate can increase. Most of the research area has a distance to drainage of 0—400 meters.

There are 14 geological units in the research area which located in the Arjuna-Welirang and Anjasmaru Tua Volcanic quarter complex based on 1:50,000 Geological Map of Pandaan Page [33]. The Anjasmaru Tua Volcanic complex is symbolized as Qvok and the rest of the symbol is part of Arjuno Welirang complex. The Anjasmaru Tua Volcanic complex is older than Arjuno Welirang, which formed in the early Pleistocene era.

Rainfall is one of the triggering factors of landslides [21]. The high intensity of rainfall can fasten the soil saturation process; thus, the slope has a greater soil mass. Consequently, it will increase the landslide hazard in the area [21]. In the research area, rain falls between 2,000 to 2,800 mm/year with the highest rainfall in the slope area of Anjasmaru and Arjuna-Welirang Mountain.
The profile curvature in this research was classified into three classes. Profile curvature showed the gradient value of the slope, such as convex, flat and concave. The negative value represented concave, zero (0) represented flat and positive value represented convex. The extreme positive or negative value of a slope indicated an unstable slope [34].

Slope is the most important factor for slope stability analysis. A steeper slope has a higher stream number and water transport energy. It is because of the linear correlation between gravitational acceleration and slope steepness [35]. A slope can be classified into 7 classes using Van Zuidam (1973) classification, namely 0-2° (Flat), 2-4° (Gentle), 4-8° (More Gentle), 8-16° (Slightly Steep), 16-35° (Steep), 35-55° (Very Steep), dan >55° (Extreme Steep) [36]. In the research area, the lower and middle parts of Welirang and Anjasmoro Mountain are classified as Flat—Gentle and Steep—Very Steep slope respectively.

Elevation in the research area is varied between <250 masl (mainly in the northern part) to >2,500 masl (mainly in the southern part, with the range of 850 masl to 2,500 masl). Generally, landslide tends

Figure 6. Landslide Triggering Factors (a. Profile Curvature; b. Slope; c. Elevasi; d. Aspect)
Source: Authors (2020)
to occur in a higher elevation area [37], [38]. Such a condition happens because of the higher soil moisture and greater rainfall distribution, as well as its mountainous topographic feature.

To identify the characteristic of the erosion, runoff and sediment disposition of a slope, the aspect parameter is used in this research. Aspect value could illustrate a higher solar intensity and rainfall, which fasten the material decomposition process due to erosion and weathering. In addition, it could also affect the stability of a slope regarding its soil moisture condition.

![Figure 7. Landslide Triggering Factors (a. Land Use; b. Distance to Road)](source: Authors (2020))

Land use can also affect the slope stability of an area [39]. Furthermore, it also reflects the human activities and the environmental conditions in the research area. In particular, soil stability can be improved mechanically and hydrologically by the existing vegetation characteristics. Some of the land uses in the research area are forest, agriculture, bare lands, settlements and grasslands. The highest proportions are forest and agriculture.

Distance to road is one of the most important factors to assess a landslide susceptibility of an area. It is because the distance to road factor can see the human-related morphological changes. Some of the landslide events in the research area are triggered by cut slopes for roads.

2.3 Data Normalization

Frequency ratio (FR) computation was done using Equation 5 as shown below. From the equation, score coefficients for each class in each factor could show the correlation between landslide probability events [40].

$$FR_{vi} = \frac{LVC_i/L_s}{AC_i/A_s}$$

Where $LVC_i$ is the pixel number of landslide events in each $i$ class factors; $L_s$ is the total pixel number landslide events in each class factors; $AC_i$ is the total pixel of the research area.

Data normalization was done after computation using Equation 5 had done. This process resulted in a value ranging from 0.01 to 0.99 after applying a min-max method [11,27,41].

$$Value_{new} = \frac{value - min(v)}{max(v) - min(v)} \times (upper - lower) + lower$$

Where $Value_{new}$ is the data normalization result, $value$ is the true value of a class, $upper$ is a 0.99 constant and $lower$ is a 0.01 constant.
2.4 Landslide Modeling

This research used a logistic regression method for dataset training. The analysis was done using Equation 3. A landslide probability is ranging from 0 (0% landslide probability) to 1 (100% landslide probability) [42]. The correlation between landslide events with the triggering factors can be seen from the computation using Equation 4.

\[
P = \frac{1}{1 + e^{-z}} = \frac{e^z}{1 + e^z}
\]  

(3)

Where \( P \) is landslide probability; \( e \) is exponential; \( z \) is combination score from independent variable which calculated using Equation 7.

\[
Z = b_0 + b_1X_1 + b_2X_2 + \ldots + b_nX_n
\]  

(4)

Where \( b_0 \) is a constant from the model; \( b_n \) is score from landslide triggering factor of \( X_n \).

2.5 Model Validation

The trend to use an ROC curve algorithm to evaluate machine learning's performance has been increasing for the last decade [43], [44]. The curve can evaluate the quality of a model based on the AUC value. In order to generate the ROC curve, true positive rate (sensitivity) and false positive rate (1-sensitivity) values need to be plotted. The rates were computed using Equation 5 and 6 as follows:

\[
\text{sensitivity} = \frac{TP}{TP + FN}
\]

\[
\text{specificity} = \frac{TN}{TN + FP}
\]

(5)

(6)

Where \( TP \) is true positive value; \( FN \) is false negative value; \( TN \) is the total of true negative; \( FP \) is the total of false positive.

3. Results and Discussion

3.1 Frequency Analysis of Landslide Events on Each Factor Class

Frequency analysis of landslide events on each factor class can be seen in Table 1. The table includes the calculation of FR value until the normalization value (\( Nv \)). The total pixel of the all research area is 634,629 with the 383 pixels of them are the landslide events. As shown in the normalization value (\( Nv \)) column, the value of each factor varied from 0 to 1. Those numbers indicated the influence significancy of each triggering factor to the landslide events—if the numbers are close to 0, the influence is insignificant. On the other hand, if the numbers are close to 1, the factor has a significant influence on landslide.

| Conditioning Factors | Classes       | Lsci | Aci       | FR      | \( Nv \)   |
|----------------------|---------------|------|-----------|---------|------------|
| Elevasi              | < 500         | 0    | 162085    | 0       | 0.00       |
|                      | 500 - 800     | 105  | 171390    | 1.015138| 0.45       |
|                      | 800 - 1100    | 138  | 102166    | 2.238174| 1.00       |
|                      | 1100 - 1400   | 95   | 78722     | 1.999625| 0.89       |
|                      | 1400 - 1600   | 20   | 38999     | 0.849763| 0.38       |
|                      | > 1600        | 25   | 81267     | 0.509738| 0.23       |
| Slope                | < 2           | 0    | 17048     | 0       | 0.00       |
|                      | 2-4           | 0    | 51529     | 0       | 0.00       |
|                      | 4-8           | 15   | 150953    | 0.164653| 0.05       |
|                      | 8-16          | 90   | 170472    | 0.874804| 0.29       |
|                      | 16-35         | 214  | 209487    | 1.692692| 0.55       |
|                      | 35-55         | 63   | 34183     | 3.053877| 1.00       |
|                      | >55           | 1    | 957       | 1.731447| 0.57       |
| Aspect     | Flat | North  | Northeast | East  |
|-----------|------|--------|-----------|-------|
|            | 0    | 126    | 89        | 54    |
|            | 471  | 144026 | 100420    | 55717 |
|            | 0    | 1.449609 | 1.468557  | 1.605932 |
|            | 0.00 | 0.86   | 0.87      | 0.95  |
| Aspect     | Shoutheast | South   | Southwest | West  |
|            | 16   | 1      | 6         | 19    |
|            | 15729 | 6256   | 34049     | 105927|
|            | 1.685544 | 0.264865 | 0.29199   | 0.297213|
|            | 1.00 | 0.16   | 0.17      | 0.18  |
| Aspect     | Northwest |                 |
|            | 72   | 172034 |
|            | 0.693489 | 0.693489 |
|            | 0.41  | 0.41   |
| Topographic Wetness Index | < -0.617 | -0.617 - 4.977 | 4.977 - 10.573 | 10.573 - 16.168 |
|            | 0    | 170    | 202       | 10    |
|            | 17   | 173139 | 412158    | 44089 |
|            | 1.429539 | 1.626954 | 0.812099  | 0.37583 |
|            | 0.49  | 0.49   | 0.28      | 0.23  |
| Topographic Position Index | < -3.399 | (-3.399) - (-1.504) | (-1.504) - 0.390 | 0.390 - 2.285 |
|            | 0    | 27     | 207       | 131   |
|            | 2281 | 31296  | 424473    | 165909|
|            | 0.789411 | 1.429539 | 0.812099  | 0.37583 |
|            | 0.49  | 0.49   | 0.28      | 0.23  |
| Topographic Position Index | < 100 | 100 - 500 | 500 - 1200 | > 1200 |
|            | 12   | 223    | 123       | 25    |
|            | 248978 | 235112 | 118412    | 32127 |
|            | 0.079862 | 1.571633 | 1.721197  | 1.28941 |
|            | 0.00  | 0.91   | 1.00      | 0.62  |
| Stream Power Index | -0.07 - (-0.001) | (-0.001) - 0.0 | 0.0 - 0.13 |
|            | 165  | 233    | 195       |
|            | 275081 | 235112 | 271842    |
|            | 0.993904 | 1.571633 | 1.88609   |
|            | 0.74  | 0.74   | 1.00      | 0.74  |
| Profile Curvature | < 50 | 50 - 100 | 100 - 250 | > 1000 |
|            | 44   | 28     | 98        | 52    |
|            | 107901 | 77803  | 157622    | 128143|
|            | 0.675691 | 0.596325 | 1.272008  | 0.672403|
|            | 0.10  | 0.00   | 0.84      | 0.09  |
| Distance to Road | < 100 | 100 - 200 | 200 - 400 | > 400 |
|            | 253  | 98     | 31        | 0     |
|            | 336892 | 195019 | 95873     | 1877  |
|            | 1.244374 | 0.832665 | 0.53578   | 0.00  |
|            | 1.00  | 0.67  | 0.43     | 0.00  |
| Distance to Drainage | < 2300 | 2300 - 2500 | 2500 - 2700 | > 2700 |
|            | 0    | 0      | 81        | 302   |
|            | 50945 | 66077  | 197164    | 320443|
|            | 0.00  | 0.00   | 0.44      | 1.00  |
| Rainfall   | Qlw4 | Qlw3   | Qvok      | Qvw   |
|            | 4    | 0      | 104       | 40    |
|            | 6777 | 2068   | 158469    | 225406|
|            | 0.978011 | 0.978011 | 1.087452  | 0.295225|
|            | 0.22  | 0.00   | 0.24      | 0.07  |
| Geological Unit | Qvlp1 | Qvlp2  |
|            | 26   | 34     |
|            | 17395 | 15646  |
|            | 2.476681 | 3.600781 |
|            | 0.56  | 0.81   |
3.2. LR-FR Model Construction
Landslide probability (susceptibility) using LR-FR model were built from Equation 3 using Python’s programming language. The computations resulted a value of -12.644 for the constant model ($b_0$). After that, a Z value (combination score of independent variables) were generated using Equation 7 as the base to estimate the landslide probability. The value of each factors ($b_n$) indicated the influence significance on landslide events. As the $b_n$ value higher, the contribution to landslide is also increased. Based on the result, slope has the highest $b_n$ value (9.240) and distance to road has the lowest (-2.125). Accordingly, slope has the highest and distance to road has the lowest contribution to landslide events in the research area.

$$Z = -12.644 + 0.638[TWI] + 1.419[TPI] + 2.711[SPI] + 9.240[Slope] - 0.436[Rainfall] - 0.493[Profile Curvature] + 4.285[Land Use] + 0.817[Geology] + 3.218[Elevation] - 2.125[Distance to Road] + 5.261[Distance to Drainage] + 2.170[Aspect]$$

3.3. Model Validation
Model validation was done in order to assure the quality of the prediction. The AUC value is ranging from 0.5 to 1.0. The accuracy of the prediction model is higher if the value approaches 1.0. On the contrary, the prediction model is less accurate as it approaches 0.5.

The result showed that the landslide susceptibility model has an AUC value of 0.91 (Figure 8). It indicated that the model has high accuracy. From the ROC curve, the true positive and false negative plotting is 0.93 and 0.1 respectively.
3.4. Result Discussion
There are 5 classes of landslide susceptibility using a natural brake method, namely very low, low, moderate, high and very high. This method is highly recommended to classify a certain spatial distribution because of its capability to optimized the differences between groups which has the same value [45], [46], [47]. The result shows that the area with very low, low, moderate, high and very high landslide susceptibility has a total percentage of 37.8%, 4.0%, 5.8%, 8.4% and 44.0% respectively (Figure 9).

The very low—low landslide susceptibility class distribution is located in the flat to More Gentle areas, such as in; Desa Kuripansari, Desa Pandanarum, Desa Sumberkembar, Desa Tanjungankongo, Desa Warugunung, Desa Candiwanu, Desa Kesimantengan, Desa Wiyu, Desa Petak, Desa Bendunganjati, some of Desa Mojokembang, Desa Kembangbelor, Desa Nogosari, Desa Cepokolimo, Desa Pacet, Desa Sajen and Desa Kemiri. Areas with moderate landslide susceptibility class are located in undulating to the hilly area with slightly steep slope—covering the foot slope area of Anjasmoro and Welirang Mountain. Areas with high—very high landslide susceptibility class are located in hilly to a mountainous area with steep to extreme steep slope—covering the forest area in the foot slope until the upper slope of Anjasmoro and Welirang Mountain.

In Figure 10, the characteristic of an area with high to very high landslide susceptibility class can be known from the normalization value. The high to very high landslide susceptibility class characteristics are: elevation 800-1400 masl; slope 33-55°; aspect North to Southeast; TWI -0.617-

![Figure 9. Landslide Susceptibility Map of Kecamatan Pacet, Kabupaten Mojokerto](Source: Authors (2020))
4.977; TPI 285-4.181; SPI 100-1200; profile curvature 0.0-0.13; distance to road 100-1000 m; distance
to drainage <100; rainfall >2700; a geological unit of Qvlp2, Qvlp3 and Qvlp4; and land use of bare
land and forest. The very low landslide susceptibility class characteristics are: elevation 500 masl; slope
2-8\degree; aspect flat, south, southwest and west; TWI <-0.617; TPI <-3.399; SPI <100; profile curvature -
0.001 to 0.0; distance to road 50-100 m; distance to drainage >600; rainfall 2300-2500; a geological unit
of Qvlw3, Qvwl, Qvlp4; and land use of building area.

Figure 10. The Result of Normalization Data Graphic
Source: Authors (2020)

The LR-FR model using machine learning techniques resulted a high-quality result in this research,
similar to previous researches. One of the case examples is landslide mapping in Mila Basin, Algeria
which resulted an AUC value of 0.85 [13]. Moreover, it also resulted that slope is the most significant
factor to landslide, similar to this research.

4. Conclusions
Landslide susceptibility model using LR-FR combination model of machine learning analysis has a high
accuracy, which indicated from the AUC value of 0.91 from the ROC curve. There are 5 classes of
landslide susceptibility using a natural brake method, namely very low, low, moderate, high and very
high. The result shows that the area with very low, low, moderate, high and very high landslide
susceptibility has a total percentage of 37.8%, 4.0%, 5.8%, 8.4% and 44.0% respectively. The high to
very high landslide susceptibility class characteristics are: elevation 800-1400 masl; slope 33-55\degree; aspect
North to Southeast; TWI -0.617-4.977; TPI 285-4.181; SPI 100-1200; profile curvature 0.0-0.13;
distance to road 100-1000 m; distance to drainage <100; rainfall >2700; a geological unit of Qvlp2,
Qvlp3 and Qvlp4; and land use of bare land and forest. Slope and distance to drainage have the highest
contribution to landslide events in the research area shown the high value of score from landslide
triggering.
Acknowledgments

Special thanks to Lembaga Penelitian dan Pengabdian (LP2M) Universitas Negeri Malang for funding this research through PNBP UM funding scheme, thus the research, writing and publishing process of this paper can be well-completed.

References*

[1] BNPB 2018 *Data Bencana Indonesia 2017* (Jakarta: Badan Nasional Penanggulangan Bencana)

[2] BNPB 2014 *Informasi Bencana: Informasi Kebencanaan Bulan Teraktual* (Jakarta: Badan Nasional Penanggulangan Bencana)

[3] Ahmed B 2015 Landslide susceptibility mapping using multi-criteria evaluation techniques in Chittagong Metropolitan Area, Bangladesh *Landslides* 12 1077–1095

[4] Wang L-J, Guo M, Sawada K, Lin J, and Zhang J 2015 Landslide susceptibility mapping in Mizunami City, Japan: a comparison between logistic regression, bivariate statistical analysis and multivariate adaptive regression spline models *Catena* 135 271–282

[5] Chen W, Peng J, Hong H, Shahabi H, Pradhan B, Liu J, and Duan Z 2018 Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China *Sci. Total Environ* 626 1121–1135

[6] Chen W, Pourghasemi H R, Kornejady A, and Zhang, N 2017 Landslide Spatial Modeling: Introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques *Geoderma* 305 314-327

[7] Zhu A X, Miao Y, Wang R, Zhu T, Deng Y, Liu J, and Hong H 2018 A comparative study of an expert knowledge-based model and two data-driven models for landslide susceptibility mapping *Catena* 166 317–327

[8] Wang Y, Wu X, Chen Z, Ren F, Feng L., and Du Q 2019 Optimizing the Predictive Ability of Machine Learning Methods for Landslide Susceptibility Mapping Using SMOTE for Lishui City in Zhejiang Province, China *International Journal of Environmental Research and Public Health* 16 368

[9] Wang L J, Guo M, Sawada K, Lin J, and Zhang J, 2016 A comparative study of landslide susceptibility maps using logistic regression, frequency ratio, decision tree, weights of evidence and artificial neural network *Geosci. J.* 20 117–136

[10] Bui D T, Tuan, T A, Klempe H, Pradhan B, and Revhaug I 2016 Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree *Landslides* 13 361–378

[11] Bui D T, Tsangaratos P, Ngo P T, Pham T D, and Pham BT 2019 Flash flood susceptibility modeling using an optimized fuzzy rule based feature selection technique and tree based ensemble methods. *Science of the Total Environment* 668 1038–1054

[12] Zhang K, Wu X, Niu R., Yang, K, and Zhao L 2017 The assessment of landslide susceptibility mapping using random forest and decision tree methods in the Three Gorges Reservoir area, China *Environ. Earth Sci.* 76 405

[13] Merghadi A, Abderrahmane B, and Bui D T 2018 Landslide susceptibility assessment at Mila Basin (Algeria): a comparative assessment of prediction capability of advanced machine learning methods *ISPRS Internasional Journal of Geo-information* 7 268

[14] Pourghasemi H R, Yansari Z T, Panagos P, and Pradhan B 2018 Analysis and evaluation of landslide susceptibility: a review on articles published during 2005–2016 (periods of 2005–2012 and 2013–2016) *Arabian Journal of Geosciences* 11 193

[15] Pamela, Sadisun I A, Kartiko R D, and Arifianti Y 2018 Metode Kombinasi Weight of Evidence (WoE) dan Logistic Regression (LR) untuk Pemetaan Kerentanan Gerakan Tanah di Takengon, Aceh *Jurnal Lingkungan dan Bencana Geologi* 9 77–86

[16] Achour Y and Pourghasemi H R 2020 How do machine learning techniques help in increasing accuracy of landslide susceptibility maps? *Geoscience Frontiers* 11 871–883.

[17] Wang Y, Feng L, Li S, Ren F, and Du Q 2020 A hybrid model considering spatial heterogeneity for landslide susceptibility mapping in Zhejiang Province, China *Catena* 188 104425

[18] Tsangaratos P and Benardos A 2014 Estimating landslide susceptibility through an artificial neural network classifier. *Nat. Hazards* 74 1489–1516
[19] Lombardo L, Cama M, Maerker M, and Rotigliano E 2014 A test of transferability for landslides susceptibility models under extreme climatic events: application to the Messina 2009 disaster Nat. Hazards 74 1951–1989
[20] Cama M, Conoscenti C, Lombardo L, and Rotigliano E 2016 Exploring relationships between grid cell size and accuracy for debris-flow susceptibility models: a test in the Giampilieri catchment (Sicily, Italy). Environ. Earth Sci. 75 1–21
[21] Kornejady A, Ownegh M, and Bahremand A 2017 Landslide susceptibility assessment using maximum entropy model with two different data sampling methods Catena 152 144–162
[22] Pandey V K, Sharma K K, Pourghasemi H R, and Bandoooni S K 2019 Sedimentological characteristics and application of machine learning techniques for landslide susceptibility modelling along the highway corridor Nahan to Rajgarh (Himachal Pradesh), India Catena 182 104150
[23] Nguyen V V, Pham B T, Vu B T, Prakash I, Jha S, Shahabi H, Shirzadi A, Ba D N, Kumar R, Chatterjee J M, Bui D T 2019 Hybrid machine learning approaches for landslide susceptibility modeling Forests 10 157
[24] Bachri S, Sumarmi, Irawan L Y, Utaya S, Nurdiansyah F D, Nurjanah A E, Tyas L W N, Adillah A A, and Purnama D S 2019 Landslide Susceptibility Mapping (LSM) in Kelud Volcano Using Spatial Multi-Criteria Evaluation IOP Conference Series: Earth and Environmental Science 273 1–26
[25] Hong, H., Chen, W., Xu, C., Youssef, A.M., Pradhan, B., Tien Bui, D., 2017. Rainfallinduced landslide susceptibility assessment at the Chongren area (China) using frequency ratio, certainty factor, and index of entropy. Geocarto Int. 32, 139–154.
[26] Costache R 2019 Flash-Flood Potential Assessment In The Upper and Middle Sector of Prahova River Catchment (Romania). A Comparative Approach Between Four Hybrid Models Science of the Total Environment 659 1115–1134
[27] Bui D T, Tsangaratos P, Nguyen V T, Liem N V, and Trinh P T, 2020 Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment Catena 188 104426
[28] Pourghasemi H R, Moradi H R, Fatemi Aghda S M, Gokceoglu C and Pradhan B 2014 GIS-based landslide susceptibility mapping with probabilistic likelihood ratio and spatial multi-criteria evaluation models (North of Tehran, Iran) Arabian Journal of Geosciences 7 1857–1878
[29] Jebur M N, Pradhan B and Tehrany M S 2014 Optimization of landslide conditioning factors using very high-resolution airborne laser scanning (LiDAR) data at catchment scale Remote Sensing of Environment 152 150–165
[30] Pawluszek K and Borkowski A 2017 Impact of DEM-derived factors and analytical hierarchy process on landslide susceptibility mapping in the region of Rożnów Lake, Poland Nat. Hazards 86 919–952
[31] Jiang L, Ling D, Zhao M, Wang C, Liang Q, and Liu K 2018 Effective identification of terrain positions from gridded DEM data using multimodal classification integration ISPRS International Journal of Geo-Information 7 443
[32] Pradhan A M S and Kim Y T 2015 Evaluation of a combined spatial multi-criteria evaluation model and deterministic model for landslide susceptibility mapping Catena 140 125-139
[33] Setijdji L D 2010 Segmented Volcanic Arc and its Association with Geothermal Fields in Java Island, Indonesia Proc. World Geothermal Congress(Bali, Indonesia April 2010) pp. 25-29.
[34] Mandal S and Mondal S 2019 Statistical Approaches fo Landslide Susceptibility Assessment and Prediction (Switzerland : Springer International Publishing AG)
[35] Coanji T 2016 Slope analysis based on SRTM digital elevation model data: study case on Rokan 4 Koto area and surrounding International Journal of Engineering and High-End Technology 1 71-75
[36] Van Zuidam R A 1973 Terrain Analysis and Classification Using Aerial Photographs: A Geomorphological Approach (Enschede : International Institute for Aerial Survey and Earth Sciences (ITC))
[37] Devkota KC, Regmi AD, Yoshida KH, Pourghasemi R, Pradhan B, Althuwaynee OF, Dhital MR 2013 Landslide susceptibility mapping using certainty factor, index of entropy and logistic
regression models in GIS and their comparison at Mugling-Narayanghat road section in Nepal Himalaya *Nat Hazards* **65** 135–165

[38] Umar Z, Pradhan B, and Ahmad A 2014 Earthquake induced landslide susceptibility mapping using an integrated ensemble frequency ratio and logistic regression models in West Sumatera Province, Indonesia *Catena* **118** 124–135

[39] Chikalamo E E, Mavrouli O C, Ettema J, van Westen C J, Muntohar A S, and Mustofa A 2020 Satellite-derived rainfall thresholds for landslide early warning in Bogowonto Catchment, Central Java, Indonesia *International Journal of Applied Earth Observation and Geoinformation* **89** 102093

[40] Cao C, Wang Q, Chen J, Ruan Y, Zheng L, Song S, and Niu C 2016 Landslide susceptibility mapping in vertical distribution law of precipitation area: Case of the Xulong Hydropower station Reservoir Southwestern China *Water(Switzerland)* **8** 270

[41] Hong H, Liu J, Tien Bui D, Pradhan B, Acharya T D, Pham B T, Zhu A.-X, Chen W, and Ahmad B B 2018 Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China) *Catena* **163** 399–413

[42] Zhou S, Wang W, Chen G, Liu B, and Fang L 2016 A Combined Weight of Evidence and Logistic Regression Method for Susceptibility Mapping of Earthquake-induced Landslides: A Case Study of the April 20, 2013 Lushan Earthquake, China. *Acta Geologica Sinica* **90** 511–524

[43] Fawcett T 2006 An introduction to ROC analysis *Pattern Recognition Letters* **27** 861–874

[44] Majnik M and Bosnic Z 2013 ROC analysis of classifiers in machine learning: a survey *Intelligent Data Analysis* **17** 531–558

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

[46] Chalkias C, Ferentinou, Polykretis M, Polykretis C 2014 GIS-based landslide susceptibility mapping on the Peloponnese Peninsula, Greece *Geosciences* **4** 176–190

[47] Fressard M, Thiery Y, Maquaire O 2014 Which data for quantitative landslide susceptibility mapping at operational scale? Case study of the Pays d’Auge plateau hillslopes (Normandy, France) *Nat. Hazards Earth Syst. Sci.* **14** 569–588