International Journal on Advanced Science Engineering Information Technology

Vol.12 (2022) No. 4
ISSN: 2088-5334

GIS-Based Binary Logistic Regression for Landslide Susceptibility Mapping in the Central Part of Pacitan Regency, East Java, Indonesia

Muhammad Dimyati [a,*], Iqbal Putut Ash Shidiq [a], Dimas Bayu Ichsandya [a]

[a] *Department of Geography, Faculty of Mathematics and Natural Science, Universitas Indonesia, Depok, 16424, Indonesia*
*Corresponding author: *m.dimyati@sci.ui.ac.id*

Abstract*— Landslides are among the most hazardous phenomena in the Pacitan Regency, especially in the Sub-Districts of Pacitan, Kebonagung, Tulakan, and Arjosari, where the landslide mainly occurs. Strategic planning through GIS analysis can be applied to minimize potential losses and strengthen resilience to natural disasters. This study combined the binary logistic regression method and GIS to map the landslide susceptibility in the Sub-Districts of Pacitan, Kebonagung, Tulakan, and Arjosari, Pacitan Regency, East Java, Indonesia. An inventory map of 293 landslides was randomly divided into 80%-20% basis for model training and testing. Fourteen landslide conditioning factors including elevation, slope, aspect, plan curvature, profile curvature, topographic wetness index (TWI), land use, proximity to roads, proximity to rivers, proximity to faults, soil types, lithology, normalized difference vegetation index (NDVI) and rainfall was used. Analysis shows that fourteen landslide conditioning factors are contributed to 22.7%. The analysis shows that 36.59% or 17,734.95 Ha of the study area has high-very high susceptibility. The area of high-very high susceptibility is mainly located in the western part of the study area. It is related to high slope value and volcanic and sediment-volcanic rock from the formation of Arjosari and Mandalika. The validation using AUROC showed an excellent fit of 0.806. Validation of susceptibility map using testing data showed 0.711 accuracy value and 0.694 precision value, which meant that the susceptibility model was quite sensible. This information could be helpful to support the local government for hazard mitigation efforts.

***Keywords*— Landslide susceptibility mapping; GIS; binary logistic regression.**

Manuscript received 4 Sep. 2021; revised 15 Dec. 2021; accepted 9 Apr. 2022. Date of publication 31 Aug. 2022.

I. INTRODUCTION

Ground deformation is a dynamic process on the earth's surface [1]. This process can occur due to human intervention or naturally [2], [3]. One form of surface deformation is a landslide. Landslides are events of mass movement of soil or rock falling down the slope due to disruption of the stability of the soil or rock [4], [5]. Landslides are the most detrimental event in Pacitan Regency. In 2017, more than 150 landslides killed 19 people and damaged 615 houses, with a total loss of > 600 billion rupiahs [6]. Strategic planning through GIS analysis can be applied to minimize potential losses and strengthen resilience to natural disasters [7]–[11].

This study utilizes GIS and logistic regression in landslide susceptibility mapping (LSM) in the Sub-Districts of Pacitan, Kebonagung, Tulakan, and Arjosari. The four sub-districts have the highest landslide occurrence in Pacitan Regency. Some types of LSM methods include inventory, deterministic, and probabilistic [10], [12]. For regional planning and mitigation, the probabilistic method can be used on a medium scale (1:25,000 to 1:50,000) [10] with several advantages: (1) Probabilistic methods with statistical approaches have higher objectivity than deterministic methods [13], (2) the probabilistic method can assess the contribution of each factor that is considered to affect the occurrence of landslides [13], and (3) the probabilistic method does not require any field survey data such as the inventory method, so it is considered more efficient to map landslide susceptibility on a medium scale [10].

One type of probabilistic method widely used is logistic regression [4], [10], [14]–[17]. The logistic regression method was applied to predict the probability of landslide occurrence. The relationship between the occurrence of landslides in an area and its influence on several variables (causative factors) can be traced, both those that have a full or partial effect.

Logistic regression consists of two types of variables, namely independent and dependent variables. The independent variables, also called the physical aspect, are the factors that influence or control the occurrence of landslides[18]–[20]. Meanwhile, the dependent variable is a description of the landslide itself, in the form of numbers 0 and 1, which reflect the presence or absence of landslide events [16]. Physical aspects can be continuous, interval, dichotomous, categorical, or combined, while the dependent variable is only dichotomous [21]. By utilizing logistic regression and GIS, this study aims to identify the relationship of each physical aspect with the occurrence of landslides and identify the level of landslide susceptibility in the Sub-Districts of Pacitan, Kebonagung, Tulakan, and Arjosari.

II. Materials and Method

A. Study Area

The Sub-Districts of Pacitan, Kebonagung, Tulakan, and Arjosari are located in the central part of Pacitan Regency, with about 484.69 km² and 250,217 people in 2020 [22]. The Southern Mountains formed a hilly topography in most of the study area with an elevation of 0 – 950 m above sea level (Fig. 1) and dominated with a slope value of 15 - 25°. Sedimentary rock types and volcanic sediments form hills in some research areas, especially in Mandalika and Arjosari Formations [23]. In addition, several high and steep hills result from magma intrusions, such as Sepang hill in Tulakan Sub-District and Limo hill, Kukusan hill, and Lanang hill in Kebonagung Sub-District. A large flat area is only found downstream of the Grindulu River in Pacitan Sub-District, which serves as the economic and activity center for Pacitan Regency.

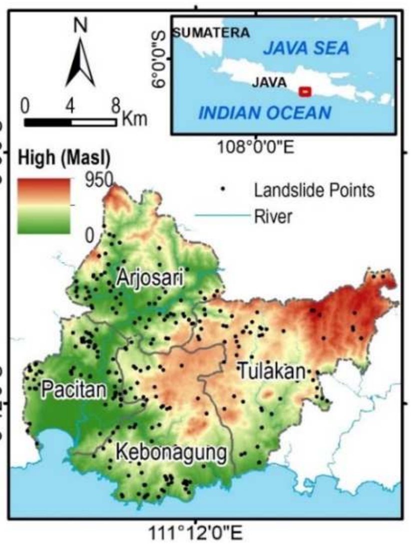

Fig. 1 The administrative area of Pacitan, Kebonagung, Tulakan, and Arjosari Sub-Districts

B. Landslide Inventory and Spatial Database

The methodology is described in Fig. 2, divided into (1) data collection, (2) landslide susceptibility modeling, and (3) validation. In the independent variable with categorical data types, the assessment of each class is obtained by comparing the landslide density of each particular physical aspect [9], [24] with the following equation:

$$Density = \frac{\left(\frac{B_i}{A_i}\right)}{\sum_{i=1}^{N}\left(\frac{B_i}{A_i}\right)} \quad (1)$$

where A_i is the area in the i^{th} sub-parameter of specific physical aspect parameters, B_i is the landslide area in the i^{th} sub-parameter of specific physical aspect parameters [24].

Landslides mainly occur in the rainy season from December to March, with an average rainfall of 240 - 543 mm/month. Landslide points data came from a combination of the landslide inventory of the BPBD of Pacitan Regency for 2017 - 2020, field observations in November 2020, and identification of satellite images from 2016 - 2020. As many as 293 landslide points were obtained across the four sub-districts. The number of landslide points is then separated; as much as 80 percent of the data is used to build the model (training), and the remaining 20 percent is used for validation purposes [8], [25], respectively of 234 and 59 points.

The selection of 14 physical aspects predicted to influence landslides is carried out based on [20], [21], [26], [27], and a review by [28] on 220 papers. Fourteen physical aspects are most often used: slope, lithology, aspect or direction towards the slope, land use, proximity to rivers, altitude, proximity to faults, plan curvature, profile curvature, soil type, proximity to roads, TWI, rainfall, and NDVI. Raster data of 14 physical aspects have a pixel size of 15 x 15 meters [16], as shown in Fig. 3.

C. Binary Logistic Regression

Logistic regression is a multivariate statistical model. All physical aspects that are thought to contribute to landslides are processed together. The resulting relationship will predict (probability) the presence or absence of landslides in the study area [16], [27], [29]. The logistic regression model was used because some data on physical aspects were not normally distributed [16]. Both the physical aspect and the dependent variable must be in the form of raster or grid data [29]. A simple model of logistic regression can be expressed as follow:

$$\mathrm{P}(\mathrm{y}=1)=\frac{(\mathrm{e}^{\mathrm{Z}})}{1+\mathrm{e}^{\mathrm{Z}}} \quad (2)$$

$P(y = 1)$ is the probability of a successful event or the occurrence of a landslide whose value ranges from 0 to 1 on an S-shaped curve [21]. If the prediction result is close to 1, the landslide is more likely to occur [17]. Z is defined as the equation below, and the value varies between $-\infty$ to ∞. β_0 is a constant, β_i is the i^{th} regression coefficient, and x_i is the i^{th} physical aspect.

$$Z = \beta_0 + \beta_1 x_1 + \ldots + \beta_i x_i \quad (3)$$

Binary logistic regression calculations were performed using the R Studio software. Meanwhile, the formation of raster data for each physical aspect and the dependent variable was processed using ArcGIS.

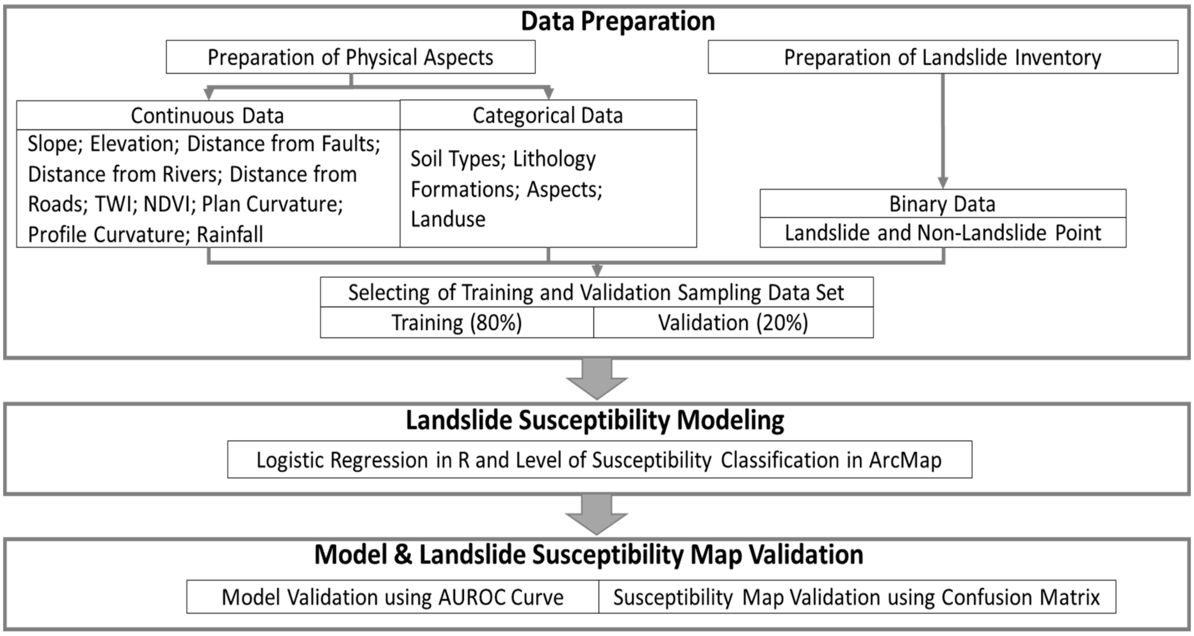

Fig. 2 Methodology of the study

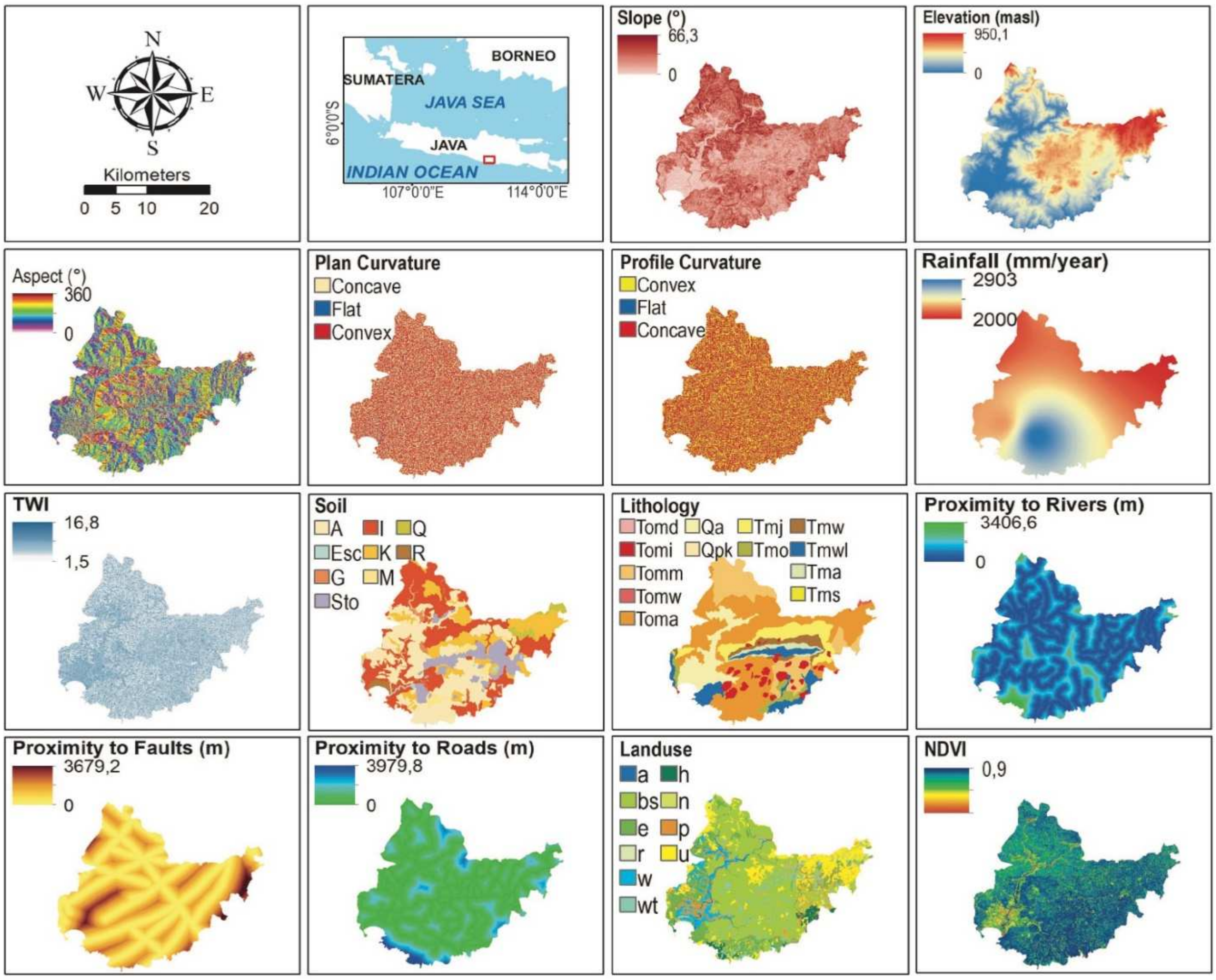

Fig. 3 Raster data of fourteen physical aspects

Based on a 1:25,000 scale topographic map sourced from [30] and further reclassified, land use in the study area is dominated by shrub (bs) of 22,248.02 Ha, dryland agriculture (u) of 7,371.9 Ha, plantations (e) of 6,318.72 Ha, settlements or build-up area (p) of 4,525.88 Ha, and others including rainfed paddy field (wt), irrigated paddy field (w), forest (h), water body (a), grassland (r), other non-cultivated vegetation (n). The topographic map from [30] is further analyzed to create proximity to rivers and proximity to faults maps. Lithological formation in the study area is dominated by Arjosari Formation (Toma) of 20,222.92 Ha, Mandalika Formation (Tomm) of 7,631.55 Ha, Alluvial (Qa) of 5,056.53 Ha, Jaten Formation (Tmj) of 4,346.92 Ha, and others including Wonosari Formation (Tmw), Intrusive rocks (Tomi), Wuni Formation (Tmw), Oyo Formation (Tmo), Nampol Formation (Tmn), and Watupatok Formation (Tomw) [23], [31].

The soil type in the study area is dominated by lithosols (I) of 15,775.51 Ha, alluvial soil (A) of 15,697,27 Ha, cambisols (K) of 8,174.63 Ha, rock outcrop (Sto), and others, including Mediterranean (M), Arenosols (Q), gleisols (G), escarpments (Esc), and regosols (R) [32]. The digital elevation model or DEMNAS product sourced from *[30]* is further analyzed to create the physical aspects of slope, elevation, aspect, plan and profile curvature, and TWI. The yearly rainfall maps are sourced from [33] and NDVI map NDVI maps are processed from Copernicus Sentinel-2 satellite imagery.

D. Validation

In this study, validation was carried out by constructing a ROC or Receiver Operating Characteristic curve. The ROC curve analysis method is widely used to assess the quality of logistic regression models by comparing the area under curves or AUC [26], [34]–[36]. For example, the logistic regression model is acceptable if the area curve under the ROC or AUC is between 0.7 and 0.8. Likewise, the model is considered excellent if the AUC is between 0.8 and 0.9 [16], [37]. Moreover, if the AUC is above 0.9, the model is considered extraordinary [37].

Validation is also done by forming a confusion matrix that compares the final prediction results with actual data on landslide and non-landslide points from observations [19], [38]. Data were classified into true positive (TP), true negative (TN), false-positive (FP), and false-negative (FN) to assess accuracy and precision (positive predictive value) [14], [27] using the following equation:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (4)$$

$$Pr\,e\,cision = \frac{(TP)}{(TP+FP)} \quad (5)$$

Accuracy is defined as the model's ability to choose which classes to select and which classes to reject. In other words, accuracy describes the model's ability to classify data correctly. *Precision* is defined as comparing the amount of data relevant to the classification results [39].

III. Result and Discussion

A. Relationship of each Physical Aspect

Based on the results of logistic regression calculations in R, 14 physical aspects have different significance. The significance column is the result of the Wald test. At =0.05, there are two hypotheses, namely:

$$H_0 : \beta_i = 0$$

(logit coefficient is not significant to the model) (6)

$$H_1 : \beta_i \neq 0$$

(logit coefficient is significant to the model) (7)

The model has six significant physical aspects: proximity to roads, NDVI, rainfall, slope, lithology, and land use. Meanwhile, the other eight physical aspects: proximity to rivers, profile curvature, plan curvature, elevation, TWI, aspect, and soil types, were considered insignificant to the model. Model testing using the McFadden R^2 test resulted in a value of 0.227. In other words, the 14 physical aspects could explain the dependent variable as far as 22.7 percent. The physical aspects that greatly influence the test data on landslide events in the study area include slopes of 8.5 percent, lithology of 4.9 percent, and land use of 3.3 percent, as shown in the importance column in table 1.

TABLE I
LOGISTIC REGRESSION COEFFICIENTS, SIGNIFICANCE, IMPORTANCE, AND VIF TEST VALUE OF FOURTEEN PHYSICAL ASPECTS

Physical Aspect	**Coeffi-cients**	**Signifi-cance**	**Impor-tance**	**VIF**
Intercept	-5.851	*P*<0.001	-	-
Slope	0.109	*P*<0.001	0.085	1.356
Slope Aspect	0.338	P=0.196	0.002	1.049
Elevation	-0.001	*P*=0.412	0.009	1.346
TWI	0.076	*P*=0.457	0.007	1.622
Profile Curvature	0.011	*P*=0.928	0.000	1.180
Plan Curvature	0.354	*P*=0.141	0.003	1.607
Proximity to faults	0.0001	*P*=0.418	0.000	1.068
Proximity to roads	-0.0004	*P*=0.046	0.003	1.102
Proximity to rivers	0.0005	*P*=0.051	0.007	1.117
NDVI	-3.464	*P*<0.001	0.017	1.281
Rainfall	0.001	*P*=0.043	0.009	1.316
Land use	1.167	*P*<0.001	0.003	1.199
Soil type	0.264	*P*=0.483	0.006	1.089
Lithology	1.229	*P*<0.001	0.049	1.109

The Hosmer and Lemeshow GOF test on the model have a P-value of 0.235. In other words, the model is suitable and can be used because the P-value > 0.05. The VIF test is used to assess the relationship between physical aspects in the model [14]. If the VIF value is > 5, it means that there is multicollinearity. Based on the VIF column in table 1 above, the VIF value in 14 physical aspects is <5, meaning that there is no multicollinearity problem, and 14 physical aspects can be used in the model.

The regression coefficient table above has positive (+) and negative (-) numbers. Physical aspects with positive coefficient values include slope, aspect (direction towards slope), rainfall, land use, soil types, lithology, plan curvature, profile curvature, proximity to faults, rivers, and TWI. Meanwhile, physical aspects with negative coefficient values include elevation, proximity to the road, and NDVI. A positive coefficient value indicates that landslides are more likely to occur as the physical aspect value increases [10]. Meanwhile, a negative coefficient value indicates that landslide events are less likely to occur when the value of the physical aspect increases. Physical aspect value is the value of each pixel in each physical aspect.

B. Level of Landslide Susceptibility

Jenks Natural Break is used to divide landslide probability into five-level, namely very low, low, medium, high, and very high [8], [14], [17]. Based on the calculation results, the study area is dominated by a very low to medium landslide susceptibility level of 63.41 percent or 30,735.38 Ha. Meanwhile, the high to very high landslide susceptibility level is about 36.59 percent or 17,734.95 Ha, as shown in Table 2.

TABLE II
PROBABILITY CLASSIFICATION AND THE AREA OF EACH LANDSLIDE SUSCEPTIBILITY LEVEL

Probability Class	Susceptibility Level	Area (Ha)	Percentage (%)
0.00006 – 0.21181	Very Low	10,275.07	21.20
0.21181 – 0.39218	Low	10,457.59	21.58
0.39218 – 0.57647	Medium	10,002.19	20.64
0.57647 – 0.76861	High	9,422.28	19.44
0.76861 – 0.99995	Very High	8,312.67	17.15

Areas with a very high level of landslide susceptibility are mainly located in the western part of the study area in the Sub-District of Kebonagung, the northern part of Pacitan Sub-District, and the southern part Arjosari Sub-District, and small parts of Tulakan Sub-District as shown in Fig. 4.

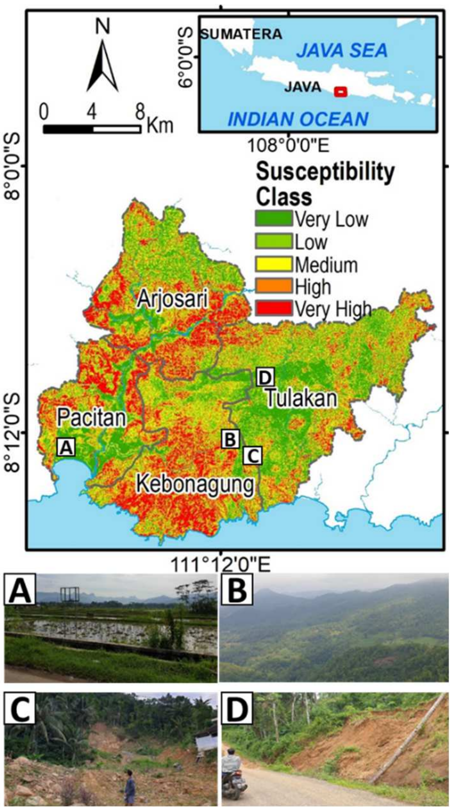

Fig. 4 Landslide Susceptibility Level in the study area; (a) Very Low Level of Landslide Susceptibility Found in Flat Area in Pacitan Sub-District, (b) Very High Level of Landslide Susceptibility Found in Steep Hilly Area in Kebonagung Sub-District, (c) Huge landslide area at Gembuk Village, Kebonagung Sub-districts, (d) Small landslide besides the road at Ngile Village, Tulakan Sub-districts

Sedimentary and volcanic sedimentary rocks from Arjosari Formation dominate this area. The average rainfall of 2,300 mm/year makes the alluvial soil in this area more easily eroded. Steep hills also dominate this area with shrub cover that is less able to bind the soil, as shown in Fig. 4a to 4c. Usually, alluvial soil is a source of landslide mass because it has high water absorption and is quickly saturated [15], [40]. If the saturation reaches the slip plane, it can cause a landslide [16], [36].

Meanwhile, areas with a very low landslide susceptibility dominate the eastern part of the study area, especially in the central part of the Tulakan Sub-District. This area has a hilly topography with low to moderate slopes and is used for rice fields and other dryland agriculture. Areas with a very low level of landslide susceptibility also can be found along and downstream of the Grindulu River in Pacitan and Arjosari Sub-Districts, as shown in Fig. 4a. The flow of the Grindulu River forms a flat sedimentary plain on its side and is used as rice fields and settlements.

C. Validation

The results of model validation using the AUROC curve produce a value of 0.806, as shown in Fig. 5. If, for example, 100 decisions are made, then the model can distinguish between landslide and non-landslide conditions as much as 80.6 percent is correct. These results are in the range of 0.8 to 0.9 according to the class made by [37], indicating an excellent value.

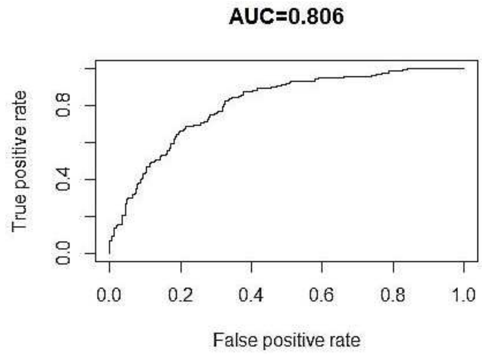

Fig. 5 Plot of the area under the curve

A total of 59 validation points were used to assess the accuracy and precision of the landslide susceptibility level. Susceptible areas are considered high and very high susceptibility levels, while non-susceptible areas are considered very low, low, and medium susceptibility levels [19]. Table 3 below is a confusion matrix that shows the number of landslide points that fall into the level of susceptible and non-susceptible.

TABLE III
CONFUSION MATRIX TO VALIDATE THE LANDSLIDE SUSCEPTIBILITY LEVEL

	Target Class (Observation)	
Validation Sample	Susceptible Area (High – Very High Level of Landslide Susceptibility)	Non-Susceptible Area (Very Low – Medium Level of Landslide Susceptibility)
Landslide Points (1)	41	18
Non - Landslide Points (0)	16	43

$$Accuracy = \frac{(41+43)}{(41+18+16+43)} = 0.711$$

$$Precision = \frac{(41)}{(41+18)} = 0.694$$

Accuracy has a value of 0.711 or 71.1 percent. As a result, the model can correctly predict the occurrence of landslides and non-landslides as much as 71.1 percent. Meanwhile, the precision has a value of 0.694 or 69.4 percent. Thus, the model can correctly predict landslide events in as many as 69.4 percent of all landslide events.

IV. Conclusion

The combination of the logistic regression method and GIS has succeeded in mapping the level of landslide susceptibility in the study area, as evidenced by the results of model validation using AUROC and accuracy and precision assessment using a confusion matrix. However, the model must be continuously improved. The physical aspect of the slope and lithological formation greatly influence landslide occurrence in the study area. A total of 17,739.95 ha of the area is included in the high to a very high level of landslide susceptibility, especially in the western part of the study area, which is dominated by steep hills composed of sedimentary rocks and volcanic sediments from the Arjosari Formation, covered by shrub.

Acknowledgment

We are grateful to the Ministry of Research and Technology/National Research and Innovation Agency, Ministry of Education, Culture, Research and Technology, and Department of Geography, Universitas Indonesia for their support and encouragement for this research.

References

[1] Z. Zeng, Y. Wang, Y. Yan, N. Xiao, and D. Chen, 'Analyzing Landslide-Prone Loess Area of Heifangtai, Gansu, China Using SBAS-InSAR Technique', in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 2018, pp. 4889–4892, DOI: 10.1109/IGARSS.2018.8518328.

[2] B. H. Martins, M. Suzuki, P. E. Yastika, and N. Shimizu, 'Ground surface deformation detection in complex landslide area—bobonaro, Timor-Leste—using SBAS DinSAR, UAV photogrammetry, and field observations', *Geosci.*, vol. 10, no. 6, pp. 1–26, 2020, DOI: 10.3390/geosciences10060245.

[3] H. Shu *et al.*, 'Relation between land cover and landslide susceptibility in Val d'Aran, Pyrenees (Spain): Historical aspects, present situation, and forward prediction', *Sci. Total Environ.*, vol. 693, pp. 1–14, 2019, DOI: 10.1016/j.scitotenv.2019.07.363.

[4] H. Hong *et al.*, 'Landslide susceptibility assessment at the Wuning area, China: a comparison between multi-criteria decision making, bivariate statistical and machine learning methods', vol. 96, no. 1. Springer Netherlands, 2018.

[5] S. T. Piralilou *et al.*, 'Landslide detection using multi-scale image segmentation and different machine learning models in the higher Himalayas', *Remote Sens.*, vol. 11, no. 21, 2019, DOI: 10.3390/rs11212575.

[6] Regency Office of Pacitan, 'Bencana Pacitan Rusak 615 Rumah Warga dan 23 Ribu Meter Jalan', *pacitankab.go.id*, 2017. [Online]. Available: https://pacitankab.go.id/bencana-pacitan-rusak-615-rumah-warga-dan-23-ribu-meter-jalan/. [Accessed: 18-Nov-2020].

[7] S. Ali, P. Biermanns, R. Haider, and K. Reicherter, 'Landslide susceptibility mapping by using GIS along the China–Pakistan economic corridor (Karakoram Highway), Pakistan', *Nat. Hazards Earth Syst. Sci.*, pp. 1–28, 2018, DOI: 10.5194/nhess-2018-39.

[8] Y. W. Rabby and Y. Li, 'Landslide susceptibility mapping using integrated methods: A case study in the Chittagong hilly areas, Bangladesh', *Geosci.*, vol. 10, no. 12, pp. 1–26, 2020, DOI: 10.3390/geosciences10120483.

[9] P. Reichenbach, M. Rossi, B. D. Malamud, M. Mihir, and F. Guzzetti, 'A review of statistically-based landslide susceptibility models', *Earth-Science Rev.*, vol. 180, no. March, pp. 60–91, 2018, doi: 10.1016/j.earscirev.2018.03.001.

[10] L. Shano, T. K. Raghuvanshi, and M. Meten, 'Landslide susceptibility evaluation and hazard zonation techniques – a review', *Geoenvironmental Disasters*, vol. 7, no. 1, 2020, DOI: 10.1186/s40677-020-00152-0.

[11] T. Mersha and M. Meten, 'GIS-based landslide susceptibility mapping and assessment using bivariate statistical methods in Simada area, northwestern Ethiopia', *Geoenvironmental Disasters*, vol. 7, no. 1, 2020, DOI: 10.1186/s40677-020-00155-x.

[12] S. Fu *et al.*, 'Landslide hazard probability and risk assessment at the community level: A case of western Hubei, China', *Nat. Hazards Earth Syst. Sci.*, vol. 20, no. 2, pp. 581–601, 2020, DOI: 10.5194/nhess-20-581-2020.

[13] B. Komac and M. Zorn, 'Statistical landslide susceptibility modeling on a national scale: the example of Slovenia', *Rev. Roum. géographie*, vol. 53, no. 2, pp. 179–195, 2009.

[14] A. Arabameri, S. Saha, J. Roy, W. Chen, T. Blaschke, and D. T. Bui, 'Landslide Susceptibility Evaluation and Management Using Different Machine Learning Methods in The Gallicash River Watershed, Iran', *Remote Sens.*, vol. 12, p. 475, 2020.

[15] F. Huang *et al.*, 'Landslide susceptibility prediction considering regional soil erosion based on machine-learning models', *ISPRS Int. J. Geo-Information*, vol. 9, no. 6, 2020, DOI: 10.3390/ijgi9060377.

[16] Y. Zhao, R. Wang, Y. Jiang, H. Liu, and Z. Wei, 'GIS-based logistic regression for rainfall-induced landslide susceptibility mapping under different grid sizes in Yueqing, Southeastern China', *Eng. Geol.*, vol. 259, 2019, DOI: 10.1016/j.enggeo.2019.105147.

[17] A. Aditian, T. Kubota, and Y. Shinohara, 'Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia', *Geomorphology*, vol. 318, pp. 101–111, 2018, DOI: 10.1016/j.geomorph.2018.06.006.

[18] J. Roy and S. Saha, 'Landslide susceptibility mapping using knowledge-driven statistical models in Darjeeling District, West Bengal, India', *Geoenvironmental Disasters*, vol. 6, no. 1, 2019, DOI: 10.1186/s40677-019-0126-8.

[19] E. Psomiadis, A. Papazachariou, K. X. Soulis, D. S. Alexiou, and I. Charalampopoulos, 'Landslide mapping and susceptibility assessment using geospatial analysis and earth observation data', *Land*, vol. 9, no. 5, 2020, DOI: 10.3390/LAND9050133.

[20] V. Vakhshoori, H. R. Pourghasemi, M. Zare, and T. Blaschke, 'Landslide susceptibility mapping using GIS-based data mining algorithms', *Water (Switzerland)*, vol. 11, no. 11, pp. 7–13, 2019, DOI: 10.3390/w11112292.

[21] X. Sun, J. Chen, Y. Bao, X. Han, J. Zhan, and W. Peng, 'Landslide susceptibility mapping using logistic regression analysis along the Jinsha river and its tributaries close to Derong and Deqin County, southwestern China', *ISPRS Int. J. Geo-Information*, vol. 7, no. 11, pp. 1–29, 2018, DOI: 10.3390/ijgi7110438.

[22] BPS-Statistics of Pacitan Regency, 'Pacitan Regency in Figures 2020', *BPS-Statistics of Pacitan Regency*, 2020. [Online]. Available: https://pacitankab.bps.go.id/publication/2020/04/27/59b932e91b04de081ce4c48f/kabupaten-pacitan-dalam-angka-2020.html. [Accessed: 28-Dec-2020].

[23] H. Samodra, S. Gafoer, and Tjokrosapoetro, 'Geological Map of the Pacitan Quadrangle, Java', Bandung, 1992.

[24] L. Zhu and J. F. Huang, 'GIS-based logistic regression method for landslide susceptibility mapping in regional scale', *J. Zhejiang Univ. Sci.*, vol. 7, no. 12, pp. 2007–2017, 2006, DOI: 10.1631/jzus.2006.A2007.

[25] Q. H. Nguyen *et al.*, 'Influence of data splitting on performance of machine learning models in prediction of shear strength of soil', *Math. Probl. Eng.*, vol. 2021, no. February 2021, DOI: 10.1155/2021/4832864.

[26] M. A. Machichi, A. Saadane, and P. L. Guth, 'On the viability of Neural Networks for landslide susceptibility mapping in the Rif, North of Morocco', *Proc. - 2020 IEEE Int. Conf. Moroccan Geomatics, MORGEO 2020*, pp. 0–5, 2020, DOI: 10.1109/Morgeo49228.2020.9121904.

[27] K. Taalab, T. Cheng, and Y. Zhang, 'Mapping landslide susceptibility and types using Random Forest', *Big Earth Data*, vol. 2, no. 2, pp. 159–178, 2018, DOI: 10.1080/20964471.2018.1472392.

[28] H. R. Pourghasemi and M. Rossi, 'Landslide susceptibility modeling in a landslide prone area in Mazandarn Province, north of Iran: a comparison between GLM, GAM, MARS, and M-AHP methods', *Theor. Appl. Climatol.*, vol. 130, no. 1–2, pp. 609–633, 2016, DOI: 10.1007/s00704-016-1919-2.

[29] C. Xu, W. Zhang, Y. Yi, and Q. Xu, 'Landslide Susceptibility Mapping Using Logistic Regression Model Based On Information Value for the Region Along China-Thailand Railway from Saraburi To Sikhio, Thailand', in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 2019, vol. 2, pp. 9650–9653, DOI: 10.1109/igarss.2019.8900041.

[30] Geospatial Information Agency, 'Peta Rupabumi Digital Indonesia', *Peta RBI Indonesia*, 2016. [Online]. Available: http://tanahair.indonesia.go.id/portal-web/. [Accessed: 25-Dec-2020].

[31] Sampurno and H. Samodra, 'Geological Map of the Ponorogo Quadrangle, Java', Bandung, 1997.

[32] BBSDLP Ministry of Agriculture, 'Peta Jenis Tanah Kabupaten Pacitan 2016', Bogor, Jawa-Barat, 2016.

[33] BBWS Bengawan Solo, 'Curah Hujan Harian Stasiun di DAS Bengawan Solo', Sukoharjo, Jawa-Tengah, 2020.

[34] L. Lombardo and P. M. Mai, 'Presenting logistic regression-based landslide susceptibility results', *Eng. Geol.*, vol. 244, no. January, pp. 14–24, 2018, doi: 10.1016/j.enggeo.2018.07.019.

[35] A. S. Soma, T. Kubota, and H. Mizuno, 'Optimization of causative factors using logistic regression and artificial neural network models for landslide susceptibility assessment in Ujung Loe Watershed, South Sulawesi Indonesia', *J. Mt. Sci.*, vol. 16, no. 2, pp. 383–401, 2019, doi: 10.1007/s11629-018-4884-7.

[36] B. Feby, A. L. Achu, K. Jimnisha, V. A. Ayisha, and R. Reghunath, 'Landslide susceptibility modelling using integrated evidential belief function based logistic regression method: A study from Southern Western Ghats, India', *Remote Sens. Appl. Soc. Environ.*, vol. 20, no. August, 2020, DOI: 10.1016/j.rsase.2020.100411.

[37] D. W. Hosmer and S. Lemeshow, 'Applied Logistic Regression'. New York: John Wiley and Sons, 2000.

[38] A. Arabameri and H. R. Pourghasemi, 'Spatial Modeling of Gully Erosion Using Linear and Quadratic Discriminant Analyses in GIS and R'. Elsevier Inc., 2019.

[39] V. Kotu and B. Deshpande, 'Chapter 8 - Model Evaluation', in *Data Science (Second Edition)*, San Francisco: Morgan Kaufmann, 2019, pp. 263–279.

[40] F. E. S. Silalahi, Pamela, Y. Arifianti, and F. Hidayat, 'Landslide susceptibility assessment using frequency ratio model in Bogor, West Java, Indonesia', *Geosci. Lett.*, vol. 6, no. 1, 2019, DOI: 10.1186/s40562-019-0140-4.