www.gi.sanu.ac.rs, www.doiserbia.nb.rs
J. Geogr. Inst. Cvijic. 2022, 72(3), pp. 229–242

Original scientific paper

UDC: 911.2:551.44(910)
https://doi.org/10.2298/IJGI2203229P

Received: April 29, 2022
Reviewed: July 25, 2022
Accepted: September 20, 2022

CAVE ENTRANCE LOCATION MODEL USING BINARY LOGISTIC REGRESSION: THE CASE STUDY OF SOUTH GOMBONG KARST REGION, INDONESIA

Rakhmat Dwi Putra[1], *Wirastuti Widyatmanti*[1]*, *Retnadi Heru Jatmiko*[1], *Tjahyo Nugroho Adji*[2], *Deha Agus Umarhadi*[1]

[1]Universitas Gadjah Mada, Faculty of Geography, Department of Geographic Information Science, Yogyakarta, Indonesia; e-mails: rakhmat.dwi.putra@mail.ugm.ac.id; wwidyatmanti@ugm.ac.id; retnadih@ugm.ac.id; dehaumarhadi@gmail.com
[2]Universitas Gadjah Mada, Faculty of Geography, Department of Environmental Geography, Yogyakarta, Indonesia; e-mail: adji_tjahyo@ugm.ac.id

Abstract: Cave entrance data are crucial as the primary indicators in the underground water inventory of a karst area. The data collection was traditionally conducted by field survey, but it is very costly and not efficient. Remote sensing and Geographic Information System (GIS) can help estimate cave entrance locations more efficiently. In this study, variables for cave entrance identification were determined using remote sensing and GIS. In addition, the accuracy of the Cave Entrance Location Model (CELM) derived from binary logistic regression was examined. Several remote sensing and geological data were used including ALOS PALSAR Digital Elevation Model (DEM), Digital Elevation Model Nasional (DEMNAS), topographic and geological map. Topographic elements were extracted by using Toposhape and Topographic Position Index (TPI). Contours derived from the topographic map showed the highest accuracy for extraction of topographic elements compared to ALOS PALSAR DEM and DEMNAS, hence it was used for further analysis. Binary logistic regression was applied to estimate the probability of cave entrance locations based on the variables used. The result shows that three topographic variables: ravine, stream, and midslope drainage had a significant value for estimating cave entrance location. Using these variables, logit equation was formulated to generate a probability map. The result shows that cave entrances are likely to be located in a dry valley. The accuracy assessment using the field data showed that 52.77% of cave entrances are located in medium to high potential areas. This suggests that the moderate-high potential area can indicate potential water resources in karst area.

Keywords: cave entrance; karst; GIS; binary logistic regression; South Gombong

1. Introduction

Karst area is defined as a surface and subsurface landscape formed by the process of dissolving carbonate rocks (i.e., limestone and dolomite), creating unique characteristics based on its relief and drainage patterns (Gilli, 2015). This landform is characterized by discontinuous surface flow

*Corresponding author, e-mail: wwidyatmanti@ugm.ac.id

patterns, caves, closed depressions, fluted rock outcrops, and large springs (Ford & Williams, 2007). Karst occupies one-fifth of the land area of the Earth, distributed on the surface and near-surface of all dry continents as a home for 1.5 billion people (Ford & Williams, 2007; Gilli, 2015; White et al., 1995). The existence of karst areas in Indonesia covers almost 20% of the total land of the country (Adji et al., 1999). Karst can be distinguished from other landforms by the three main characteristics, i.e., the presence of closed basins, the absence or lack of surface rivers, and the development of subsurface drainage and cave systems (White, 1988).

The secondary porosity in the karst hydrological system causes water to flow into underground rivers, leading to drought on the surface that can commonly be found in the karst area in Indonesia (Cahyadi, 2010). The karst area has abundant water sources in the form of lakes, springs, and underground water, but its use is still dominant in lakes and springs (Cahyadi, 2014). As drought is the main problem in Indonesian karst areas during the dry season, the locations of cave entrances are essential to help the surrounding community find accessibility to the underground water. An inventory of underground river water sources can be done by tracing and mapping cave passages until the flows are found using fluorescent dye and salt (Al-Ghozali et al., 2021; Benischke et al., 2007; Goldscheider et al., 2008; Käss, 1999). A field study should be conducted before tracing, aiming to locate the cave entrance or other features such as ponor and spring (Agniy et al., 2019). Thus, this indicates the importance of cave entrance distribution that helps the inventory of underground river water sources to be more focused on.

The conventional searches of cave entrances are usually carried out by utilizing information from local residents in karst areas, giving information on new cave entrances. Yet, it is only limited to caves that are already known by the local people, whereas cave entrances in unexplored karst areas are mostly unidentified. Field survey for the detection of cave entrances without any initial information requires significant time, energy, and money. Instead, utilization of remote sensing data and geographic information system (GIS) may efficiently provide a preliminary detection of entrance caves. Previous studies have used remote sensing and GIS analysis to determine the location of cave entrances (Hung et al., 2002; Kusumawati & Zuharnen, 2017; Kuswanto, 2016; Muttaqin et al., 2019; Saputra & Jatmiko, 2014). The high probability of cave locations was shown by a fracture zone based on lineament density indices, interpreted from enhanced Landsat image (Hung et al., 2002). Saputra and Jatmiko (2014) and Kusumawati and Zuharnen (2017) identified cave entrances by visual interpretation from several variables, i.e., flow pattern, lineaments, morphology, and vegetation, derived from elevation data and satellite imageries. A machine learning model, maximum entropy, has been explored to predict the probability of caves based on valley and hill classification, the distance from a valley, elevation, slope, aspect, distance from lineament, lineament density, and distance from water source (Muttaqin et al., 2019). A binary logistic regression was also used to detect cave entrances using the variables of archeological cave sites, landforms, lineament, and slope (Kuswanto, 2016). Previous studies reported that topographical conditions are important to describe the characteristics of the environment around the cave entrances, hence more topographic elements are needed to more accurately identify features of landforms (Widyatmanti et al., 2016).

Remote sensing and GIS can help in tracking the cave entrances effectively and efficiently (Putra, 2021). Initial information in the form of areas that have the potential to contain cave entrances will help in finding the location of the cave entrances through terrestrial surveys. Research on the identification of the entrance of the cave through remote sensing and GIS has not explained the significance of the variables used quantitatively. Based on this, a key

interpretation formula is needed to determine areas that have the potential to contain cave entrances. The existence of a cave entrance is correlated with certain environmental conditions. The topographical conditions formed around the entrance of the cave can be used as a basis for predictive modeling to track the existence of the entrance of the cave.

This study aims to (a) determine variables to identify cave entrances using remote sensing and GIS approach, and (b) examine the accuracy of the Cave Entrance Location Model. The research took a study case in the South Gombong Karst Region, Central Java, Indonesia that has not been studied before for modeling cave entrance location. In addition, we also explored more parameters such as Topographic Position Index (TPI) as well as comparing elevation datasets in the modelling process.

2. Study area

South Gombong is a karst region located in the southern zone of Central Java Province, Indonesia (7.756°S–7.645°S, 109.388°E–109.487°E) with a total area of 125.49 km^2 (Figure 1; Asikin et al., 1992). The morphology is generally the typical karst area of a high-intensity precipitation, characterized by cockpit karst with conical residual hills (Haryono et al., 2017). The karst area has a plateau morphology, formed by the uplifted fault blocks during the Late Pliocene (Haryono et al., 2017; van Bemmelen, 1970). The limestone in the area is thin in the southern part and getting thicker toward the north, with the underground drainage system mostly flowing to the north (Haryono et al., 2017).

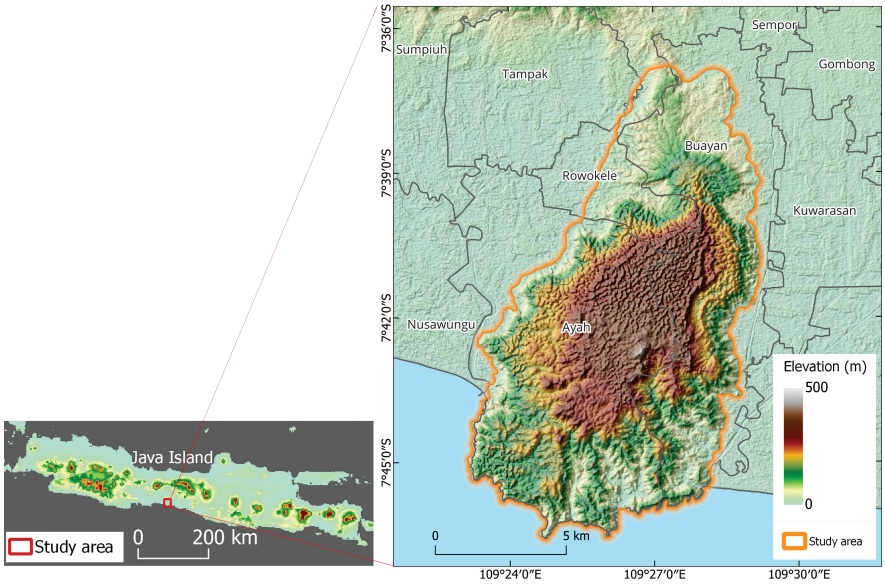

Figure 1. Map of the study area located in South Gombong Karst Region.

Our study area consists of three formations, i.e., Kalipucang, Gabon, and Halang Formations (Asikin et al., 1992). The karst area is formed in the limestone of Kalipucang Formation during the middle Miocene age. Kalipucang Formation overlaps unconformably above Gabon Formation with lower Miocene volcanic breccia rocks. The northern part of the

Kalipucang Formation overlaps with the Halang Formation above it. Halang Formation consists of the alternation of sandstone, claystone, marl, and tuff with breccia intercalations. The limestone in the study area is relatively thin, ranging between 50–350 m (Laksono, 2019).

3. Methods

3.1. Datasets and variable extraction

The variables used in this study are Toposhape, TPI, lineament, and geology. Despite having a similar purpose to extract topographical features, the approaches used by Toposhape and TPI are different and were compared in this study. Topographic elements in the form of Toposhape composition are able to characterize certain landform characteristics (Widyatmanti et al., 2016). TPI classified landforms by calculating the elevation difference between a central grid cell and the average of the surrounding grid cell using the combination of small- and large-scale classifications, so then it defines the relative position of the grid cell along a topographic gradient (Gruber et al., 2017; Guisan et al., 1999; Weiss, 2001).

Two different sources of elevation data were compared to find the most accurate data to extract topographical parameters, consisting of Digital Elevation Model Nasional (DEMNAS) and topographic map at the scale of 1:25,000 (Geospatial Information Agency, 2017). DEMNAS is a national elevation dataset in Indonesia provided at a high spatial resolution (8.25 m). DEMNAS was originally generated from radar satellites including TerraSAR-X, IFSAR, and ALOS PALSAR, as well as stereo-plotting data (Geospatial Information Agency, 2018). The contour feature from topographic map data was interpolated to generate two Digital Elevation Models (DEMs) at different resolutions: 6.25 and 12.5 m, which are hereafter named DEM 6.25 m and DEM 12.5 m.

Toposhape parameter was extracted from the DEM data using Toposhape tool in IDRISI software. The settings of eigenvalue and gradient magnitude were set to 0.00001 (1×10^{-5}) and 0.000001 (1×10^{-6}), respectively. TPI parameter was generated using TPI based landform classification tool in SAGA GIS software (Conrad et al., 2015). The lineament data were obtained from the automatic extraction of DEM using the lineament extraction tool of the PCI Geomatica software. The rule set for the lineament extraction tool refers to Febriani (2019) as listed in Table 1. The distance of lineament buffer was set to 25 m referring to Kuswanto (2016). Geological data are obtained from the Java Geological Map: Sheet Banyumas scale 1:100,000 (Asikin et al., 1992). Data on the distribution of the cave entrances were obtained from Acintyacunyata Speleological Club Yogyakarta.

Table 1. Description of the rule set used within the lineament extraction tool

Parameter	Rule set
Filter radius (RADI)	12
Gradient threshold (GTHR)	90
Length threshold (LTHR)	30
Line fitting error threshold (FTHR)	10
Angular difference threshold (ATHR)	30
Linking distance threshold (DTHR)	20

Note. The rule set is adapted from *Ekstraksi Kelurusan Otomatis DEM Alos Palsar untuk Analisis Hubungan Kelurusan dengan Kemunculan Mataair di Karst Gunungsewu Kabupaten Gunungkidul* [Alos Palsar DEM Automatic Linearity Extraction for Analysis of Relationship between Linearity and Water Spring in Gunungsewu Karst Area, Gunungkidul Regency; Undergraduate Thesis], by K. R. Febriani, 2019, Universitas Gadjah Mada. Copyright 2019 by Universitas Gadjah Mada. Adapted with permission.

The accuracy of the extracted variables was assessed by pinpointing the sampled points with observations conducted in the field. Variables whose accuracy level meets the requirements, calculated using an error matrix, can be further processed for analysis.

3.2. Cave entrance estimation and accuracy assessment

The modeling analysis of the prediction of the cave entrances was carried out by binary logistic regression. Logistic regression is a non-linear regression model that allows the use of either continuous or discrete data as well as their combination, and has no requirement of having a normal distribution of the data (Lee, 2005). Binary logistic regression approach has proven its capability to detect cave entrances as well as other applications such as mapping the groundwater spring potential and landslide susceptibility mapping (Kuswanto, 2016; Ozdemir, 2011; Pourghasemi et al., 2013). In this case, the dependent variable is binary, representing the presence and the absence of cave entrances, whereas the independent variables include Toposhape, TPI, lineament, and geology. Variable datasets were overlaid with the sample point model of the cave entrances, and then the extraction of entrance location pixel values was performed. The results of the extraction of variable pixels at the location of the entrance of a cave were then converted to a binary table to be analyzed using binary logistic regression, with the values of 0 and 1, respectively, corresponding to the absence and the presence of caves. Binary logistic regression analysis was performed based on the input variables. The result of the logistic regression is a formula of estimation model to produce a probability for the existence of a cave entrance.

An accuracy assessment was conducted by extracting the value of cave entrance probability with the testing data. Cave entrance probability was classified into three classes, i.e. low, moderate, and high. The total accuracy value is obtained by summing the total accuracy values at moderate and high classes.

4. Results and discussion

4.1. Topographical features

The topography of the South Gombong Karst Region consists of hills and valleys. The shape of the hills is in the form of a cone with a fairly steep slope. On average, the relative height of the cones is 28.16 m according to our calculation. Based on DEMNAS data, the highest hill in the South Gombong Karst Region has a height of about 400 m a.s.l. while the lowest area has a height of about 6 m a.s.l. The maps of Toposhape and TPI results are illustrated in Figure 2.

There are 12 classes of topographic elements produced from Toposhape: peak, ridge, saddle, flat, ravine, pit, convex hillside, saddle hillside, slope hillside, concave hillside, inflection hillside, and unknown hillside (Widyatmanti et al., 2016). The Toposhape data accuracy assessment was carried out through a field survey in 10 sample points for each class. Not all classes were field checked because there were several classes with too few and insignificant pixel populations in the cave entrance environment. The five Toposhape classes for the accuracy assessment are ridge, ravine, convex hillside, saddle hillside, and concave hillside. The results of the accuracy assessment show that the DEMNAS Toposhape has an accuracy of 79%, while both DEM 6.25 m and 12.5 m produced the same accuracy at 81%. ALOS PALSAR DEM could not be further used since most of the pixels fell to the unknown hillside, thus the topographical features are not clearly identified.

Topographical features in the study area are dominated by saddle hillside (Figure 2), while peak, saddle, flat, pit, slope hillside, inflection hillside, and unknown hillside are not significantly identified. This might be because the Toposhape algorithm used the size of a 3 × 3 moving window (Ercanoglu, 2005). This moving window size is too small for the DEM data used, hence the analysis region is narrow and some of the classes were not detected.

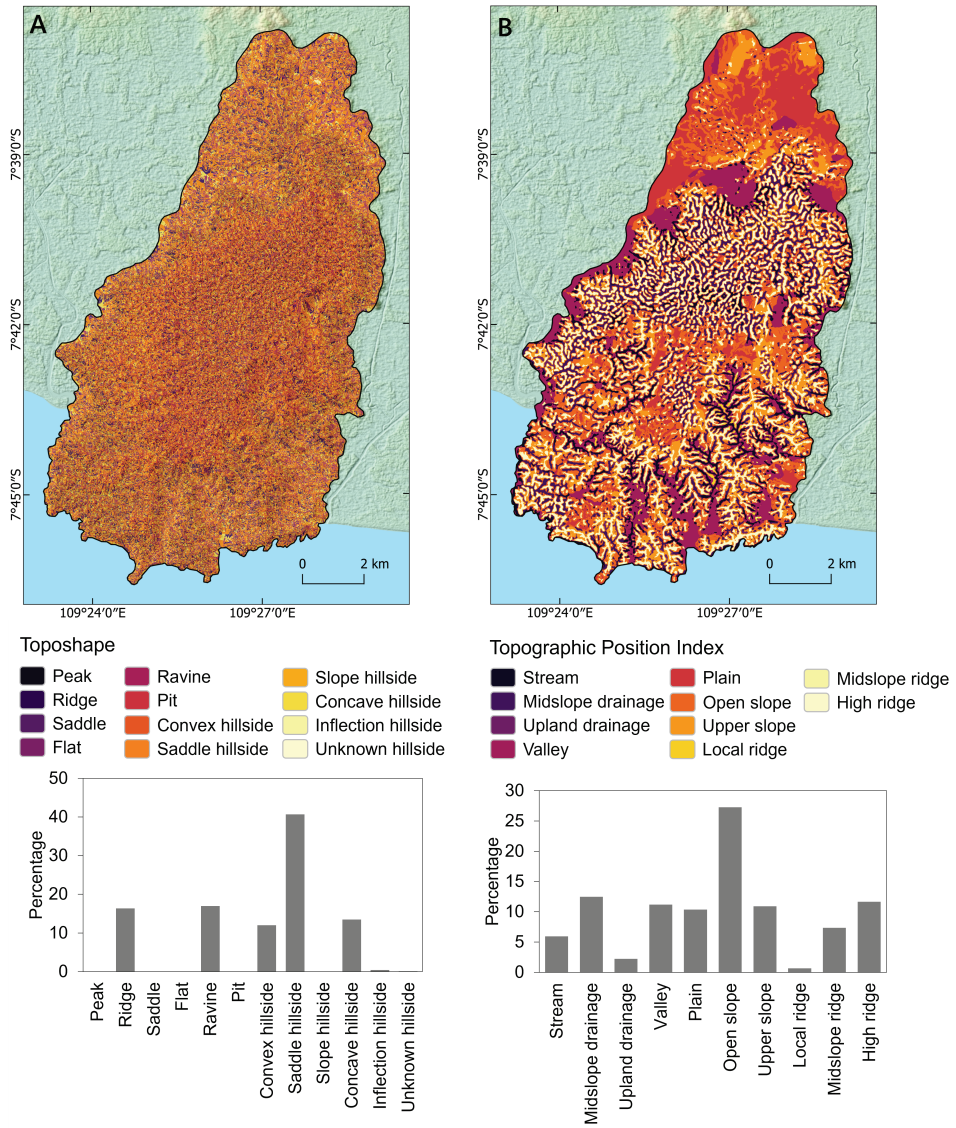

Figure 2. Map and the classes percentage of Toposhape based landform classification (A) and TPI based landform classification (B), both are derived from DEM 6.25 m.

Topographic extraction based on TPI was processed using DEMNAS data, 6.25 m topographic map-based DEM, and 12.5 m topographic map-based DEM. TPI extraction was conducted by setting radius A to 100 m and radius B to 1,000 m on the TPI Based Landform Classification tool of SAGA GIS software. Compared to Toposhape, TPI is visually more representative in characterizing topographical features due to the larger searching radiuses (100 and 1,000 m) than Toposhape which uses a 3 × 3 grid cell.

The DEM data extraction resulted in 10 landform classes, consisting of stream (canyon, deeply incised stream), midslope drainage (shallow valley), upland drainage, valley (U-shape), plain, open slope, upper slope, local ridge, midslope ridge (small hills in plains), and high ridge. The accuracy was assessed by field checking in 10 sample points for each class. The results of the accuracy assessment show that DEMNAS, DEM 6.25 m, and DEM 12.5 m produced an accuracy of 72%, 74%, and 72%, respectively. Open slope dominated the TPI features covering 27.3% of the study area. The percentages of plain, upper slope, valley, high ridge, and midslope drainage range between 10.3–12.5%.

The accuracy assessment results show that the 6.25 m contour DEM data has the highest accuracy rate compared to other DEMs. This could be since DEMNAS and ALOS PALSAR still contain the information of surface elevation, instead of the terrain as the contour represents (Julzarika & Harintaka, 2019; Umarhadi & Danoedoro, 2019). Based on this, the 6.25 m contour DEM data was used further in the analysis for the source of variable data extraction.

4.2. Geological lineament

Lineament extraction was generated using the lineament extraction tool for PCI Geomatica 2015 software. The DEM 6.25 m was converted to a hill-shade image. Afterward, the lineament features were extracted. The resulted lineament characteristics have short lengths and are in the form of minor lineaments. The lineaments that can be extracted are ridge lineaments and valley lineaments. As shown in Figure 3, the lineament in the study area is dominated by the NE-SW direction. The lineament pattern in the Kalipucang Formation is dense and generally parallel pattern toward NE-SW. In the Gabon Formation, the lineaments are not too close with others, yet the direction is random. In the Halang Formation, only several lineaments were extracted (Figure 4).

Identification of lineaments with the location of cave entrances was conducted by creating a lineament buffer of 25 m wide. The buffer distance of 25 m is considered to affect the lineament (Kuswanto, 2016). It can be identified that a total of 15 caves are in the lineament, while 35 caves are not.

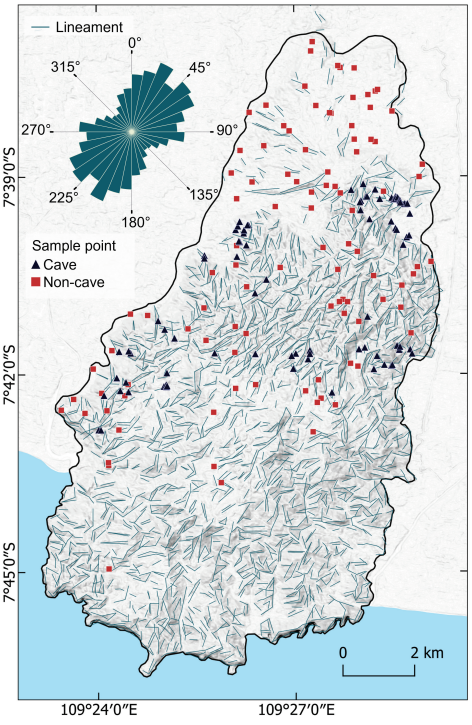

Figure 3. Geological lineaments in South Gombong Karst Region.

4.3. Geology

A total of 50 sample points of caves were overlaid on the geological map to determine the distribution of their locations on the rock characteristics. As shown in Figure 4, the distribution of cave entrances is mostly situated in Kalipucang

Formation (39 points) composed of limestone. A total of 11 cave entrances are in Halang Formation, composed of sandstone, claystone, marl, and tuff with breccia intercalations. Based on the rock characteristics in the Halang Formation, the possibility of cave entrances located in this area is very low since it is not a karst area. Nevertheless, it is due to the use of a coarse geological map (scale at 1:100,000), thus the boundaries of the formation are not detailed. Moreover, as shown in a cross-section of rock formations from south to north, Halang Formation overlaps with the Kalipucang Formation at the border (Figure 4). Based on this map, it is possible that the actual points of cave entrances are in the limestone of Kalipucang Formation. As observed, the points of cave entrances located in Halang Formation are adjacent to Kalipucang Formation. Considering the limestone characteristics, areas other than Kalipucang and Halang formations were excluded from the modeling procedure.

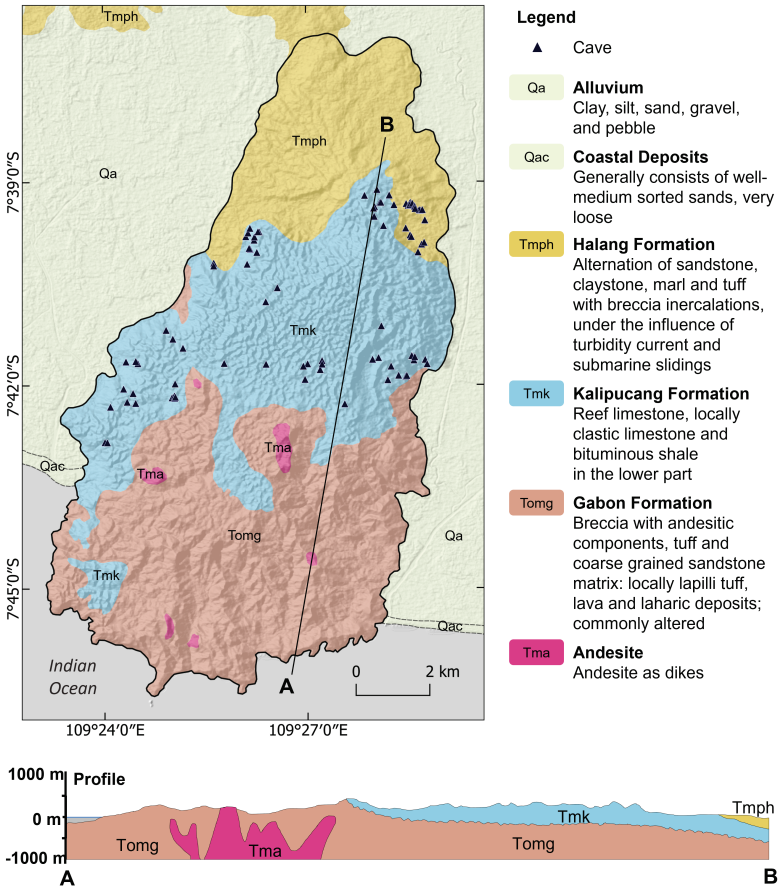

Figure 4. Geological map of South Gombong Karst Region.

Note. Adapted from *Geologic Map of Banyumas Quadrangle, Jawa* (1:100,000) [Map], by S. Asikin, A. Handoyo, B. Prastistho, & S. Gafoer, 1992, GeoMap (https://geologi.esdm.go.id/geomap/pages/preview/peta-geologi-lembar-pangandaran-jawa). In the public domain.

4.4. Modeling and analysis of cave entrance probability

Firstly, we assessed the accuracy of the model using the data utilized in training the model. This is to assess the effectiveness of the predicted classification against the actual data. The assessment shows that the overall accuracy of the data is 72% with an accuracy of 50% for cave samples (Specificity) and 83% for non-cave samples (Sensitivity). This indicates that the data used in the analysis process have sufficient accuracy.

The results of binary logistic regression showed that 14 out of 18 variables could be further processed. These variables can be seen in Table 2. Only three variables managed to show a significance value of less than .05, consisting of ravine, stream, and midslope drainage. Therefore, our model was built based on these parameters. Among these, stream has a higher coefficient (B; regression weight) with a value of 3.421 compared to midslope drainage (B = 2.209) and ravine (B = 1.341).

Table 2. Results of binary logistic regression analysis

Variable	B	*SE*	Wald	*df*	Sig.	Exp(B)
Concave hillside	–.816	.688	1.407	1	.236	.442
Convex hillside	–.222	.792	.078	1	.780	.801
Ravine	1.341	.573	5.486	1	.019	3.824
Ridge	–.038	.670	.003	1	.954	.962
Stream	3.421	1.263	7.334	1	.007	30.604
Midslope drainage	2.209	.967	5.223	1	.022	9.107
Upland drainage	–19.583	27897.947	.000	1	.999	.000
Valley	1.067	1.049	1.033	1	.309	2.906
Plain	–.959	1.359	.498	1	.481	.383
Open slope	1.147	.912	1.580	1	.209	3.147
Upper slope	1.422	.978	2.115	1	.146	4.144
Midslope ridge	–.272	1.406	.037	1	.847	.762
Lineament	.959	.608	2.487	1	.115	2.610
Karst (Kalipucang Formation)	.072	.492	.022	1	.883	1.075
Constant	–2.207	.975	5.123	1	.024	.110

The model formula was calculated by the logit equation. Based on the constant values in Table 2 above, the modeling formula was formed as shown in Equation 1. The probability (*Pi*) formula was also formulated (Equation 2; Fahrmeir et al., 2013; Kuswanto, 2016). The formula was then substituted with the logit equation so that the probability for the existence of a cave entrance was quantified by Equation (3). Within the Equation 1 and 3, *rv*, *st*, and *md* denote ravine, stream, and midslope drainage, respectively. Finally, the probability map for the estimation of the entrance of the cave was generated as illustrated in Figure 5.

$$Logit\ (Y) = -2.207 + (1.341 \cdot rv) + (3.421 \cdot st) + (2.209 \cdot md) \tag{1}$$

$$Pi = \frac{Exp.Vi}{1 + Exp.Vi} = \frac{Odd}{1 + Odd} \tag{2}$$

$$Probability = \frac{Exp\left[(-2.207) + (1.341 \cdot rv) + (3.421 \cdot st) + (2.209 \cdot md)\right]}{1 + Exp\left[(-2.207) + (1.341 \cdot rv) + (3.421 \cdot st) + (2.209 \cdot md)\right]} \tag{3}$$

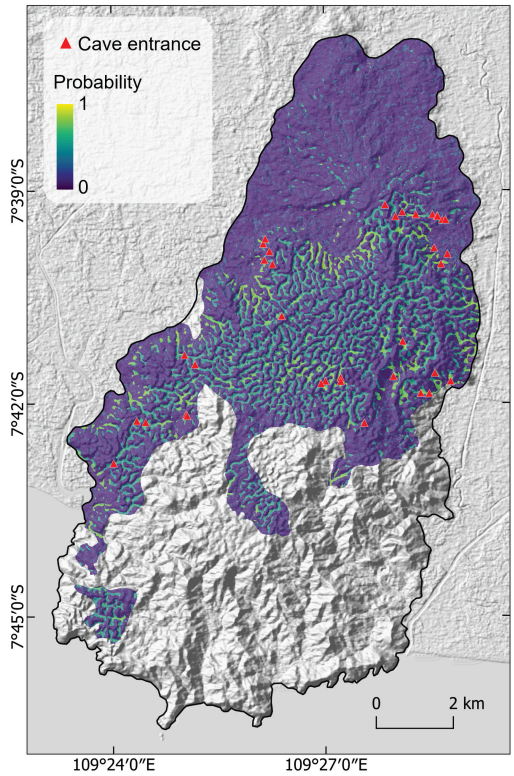

Figure 5. Probability of cave entrance location.

Table 3. Percentage of cave samples in the potential class

Potential class	Number of samples	%
Low	17	47.22
Moderate	12	33.33
High	7	19.44
Moderate + High	19	52.77

The accuracy assessment of the cave entrance estimation model was carried out by comparing the number of cave samples in the study area. A total of 36 cave sample points were used in this accuracy assessment. The distribution of the samples is provided in Figure 5. The probability model shows the lowest value of 0.090 and the highest value of 0.921. These values are categorized into three classes based on equal intervals: (a) *low potential* (0.090–0.367); (b) *moderate* (0.368–0.644), and (c) *high* (0.645–0.921). According to the categorization, 17 testing points are located in the low potential class, while 12 and 7 points are in the moderate and high potential class area. The accuracy was then calculated on each class, as well as the overall accuracy.

The accuracy assessment is shown in Table 3. Based on these calculations, the percentages of the average and high guess classes are added up to determine the total accuracy of the cave entrance estimation model. The result is that our Cave Entrance Estimation Model has an accuracy of 52.78%. These results are included in the medium level of accuracy.

Topographical variables have strong correlations with the presence of cave entrances. Based on the modeling results, the existence of caves has the potential to be found in areas with topography in the form of valleys. The cave estimation model has a high probability in the variables of ravine, stream, and midslope drainage. Previous studies stated that the location of caves is related to lineament (Hung et al., 2002; Saputra & Jatmiko, 2014). However, this study shows that lineament is not significant enough to be involved in the model. This might be due to the automatic extraction using the tool, by which some other lineaments cannot be identified.

Studies about estimating cave entrance locations were also conducted by Hung et al., (2002), Kusumawati and Zuharnen (2017), Kuswanto (2016), and Saputra and Jatmiko (2014). Saputra and Jatmiko (2014) reported that cave entrances that can be identified using remote sensing techniques are located in karst depressions and other negative morphological features. Based on the karst landform, caves are typically found in the collapse features such as dolines (Ensley et al., 2021; Kuswanto, 2016). Our results are in accordance with the previous

studies that the high probability of a cave entrance location is situated in karst valleys. In a tropical karst area, run-off water flows through the valleys, leading to limestone dissolution process, which subsequently creates sinkholes or ponors.

Meanwhile, Hung et al. (2002) and Saputra and Jatmiko (2014) stated that the probability of cave entrance locations is also high in fault or fracture zones. However, this study showed that lineament does not have a significant impact on determining the location of cave entrances, as shown by the lineament parameter significance value greater than .05 (Table 2).

This study demonstrated the use of Toposhape and TPI for extracting topographical features. The results showed that the stream feature generated from TPI is the most important variable in the modeling. However, the tools are not specifically intended for karst landscape that comprises complex features compared to other landscapes. To date, only visual interpretation could precisely identify geomorphological features using multiple remote sensing data (e.g., aerial photographs, DEM, multispectral images, and Light Detection and Ranging [LiDAR]) (Ensley et al., 2021). Thus, the detailed interpretation of karst landforms could increase the accuracy of cave entrance modeling.

Secondly, other learning algorithms that have not been applied for cave entrance identification should be explored for future studies, such as random forests and Bayes-based machine learning algorithms that have been tested to extract karst sinkholes (Taheri et al., 2019; Zhu & Pierskalla, 2016). The use of convolutional neural networks should also be considered to detect features by kernel-to-kernel analysis (Rafique et al., 2022). Nevertheless, the number of sample points should be increased significantly to allow more accurate machine learning results.

5. Conclusion

This study demonstrated the implementation of binary logistic regression to estimate the probability of cave entrance distribution using geological parameters (lithological composition of a formation and lineament) and topographical features derived using Toposhape and TPI methods. Various sources of DEMs, i.e., ALOS PALSAR DEM, DEMNAS, and topographic map, have been tested to find the best data to generate topographical features. DEM based on contour data from a topographic map with a 6.25 m pixel size was used for further analysis considering the best performance compared to the others. Out of 14 variables that can be used for the analysis, only three variables were significant in the binary logistic regression, consisting of stream, ravine, and midslope drainage. Thus, the cave entrance model was developed based on these three variables. The results showed that 52.78% of the cave samples are located in the medium to high potential class. This indicates that the prediction model of cave entrances in South Gombong Karst Region has a moderate level of accuracy. Other machine learning and deep learning algorithms are worth exploring for future studies to more accurately identify cave entrances in karst areas.

Acknowledgements

Authors thank editors and anonymous reviewers for their constructive suggestions.

References

Adji, T. N., Haryono, E., & Woro, S. (1999). *Kawasan Karst dan Prospek Pengembangannya di Indonesia* [Karst Area and its Prospective Development in Indonesia]. Seminar PIT IGI di Universitas Indonesia. https://doi.org/10.31227/osf.io/ykt3f

Agniy, R. F., Adji, T. N., Cahyadi, A., Nurkholis, A., & Haryono, E. (2019). Characterizing the cavities of Anjani Cave in Jonggrangan Karst Area, Purworejo, Central Java, Indonesia. *IOP Conference Series: Earth and Environmental Science*, *256*, Article 012009. https://doi.org/10.1088/1755-1315/256/1/012009

Al-Ghozali, M. Q., Adji, T. N., Haryono, E., Cahyadi, A., Agniy, R. F., Laksono, G. E., Priambada, A. P., Rahmawati, A. I., Mahrizkhal, D. S., Setiawan, A., Fauzi, D. R., Astuti, E. S., Putra, R. D., & Biladi, M. (2021). Identification of Karst Underground River Catchment Areas with Artificial Tracer Tests and Water Balance in Banteng Cave Springs (Karst Gombong Selatan, Central Java). In R. Che Omar, J. T. Sri Sumantyo, B. White, A. Cardenas Tristan, E. Haryono, D. R. Hizbaron, & R. F. Putri (Eds.), *E3S Web of Conferences: Vol. 325. ICST 2021 – The 2nd Geoscience and Environmental Management Symposium* (Article 08007). https://doi.org/10.1051/e3sconf/202132508007

Asikin, S., Handoyo, A., Prastistho, B., & Gafoer, S. (1992). *Geologic Map of Banyumas Quadrangle, Jawa* (1:100,000) [Map]. GeoMap. https://geologi.esdm.go.id/geomap/pages/preview/peta-geologi-lembar-pangandaran-jawa

Benischke, R., Goldscheider, N., & Smart, C. (2007). Tracer techniques. In N. Goldscheider, & D. Drew (Eds.), *Methods in Karst Hydrogeology* (1st ed., Vol. 26, pp. 147–170). Taylor & Francis Group.

Cahyadi, A. (2010, October 13). *Pengelolaan Kawasan Karst dan Peranannya dalam Siklus Karbon di Indonesia* [Karst Area Management and its Functions on Carbon Cycle in Indonesia]. Seminar Nasional Perubahan Iklim di Indonesia, Yogyakarta, Indonesia. https://osf.io/preprints/inarxiv/8gh6d/download

Cahyadi, A. (2014). Keunikan Hidrologi Kawasan Karst: Suatu Tinjauan [The Hydrology of the Karst Region: An Overview]. In A. Cahyadi, B. Argadyanto, T. A. Tivianton, & H. Nugraha (Eds.), *Ekologi Lingkungan Kawasan Karst Indonesia* [Ecological Environment of Indonesian Karst Area] (Vol. 2, pp. 1–13). Deepublish. https://doi.org/10.31227/osf.io/awvqs

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., & Böhner, J. (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, *8*(7), 1991–2007. https://doi.org/10.5194/gmd-8-1991-2015

Ensley, R., Hansen, R. D., Morales-Aguilar, C., & Thompson, J. (2021). Geomorphology of the Mirador-Calakmul Karst Basin: A GIS-based approach to hydrogeologic mapping. *PLOS ONE*, *16*(8), Article e0255496. https://doi.org/10.1371/journal.pone.0255496

Ercanoglu, M. (2005). Landslide susceptibility assessment of SE Bartin (West Black Sea region, Turkey) by artificial neural networks. *Natural Hazards and Earth System Sciences*, *5*(6), 979–992. https://doi.org/10.5194/nhess-5-979-2005

Fahrmeir, L., Kneib, T., Lang, S., & Marx, B. (2013). *Regression*. Springer. https://doi.org/10.1007/978-3-642-34333-9

Febriani, K. R. (2019). *Ekstraksi Kelurusan Otomatis DEM Alos Palsar untuk Analisis Hubungan Kelurusan dengan Kemunculan Mataair di Karst Gunungsewu Kabupaten Gunungkidul* [Alos Palsar DEM Automatic Linearity Extraction for Analysis of Relationship between Linearity and Water Spring in Gunungsewu Karst Area, Gunungkidul Regency; Undergraduate Thesis]. Universitas Gadjah Mada.

Ford, D., & Williams, P. (2007). *Karst Hydrogeology and Geomorphology*. John Wiley & Sons. https://doi.org/10.1002/9781118684986

Geospatial Information Agency. (2017). *Ina-Geoportal* [Data set]. https://tanahair.indonesia.go.id/portal-web

Geospatial Information Agency. (2018). *DEMNAS* [Data set]. https://tanahair.indonesia.go.id/demnas

Gilli, É. (2015). *Karstology: Karsts, Caves and Springs: Elements of Fundamental and Applied Karstology* (C. Fandel, Trans.). CRC Press. (Original work published 2011)

Goldscheider, N., Meiman, J., Pronk, M., & Smart, C. (2008). Tracer tests in karst hydrogeology and speleology. *International Journal of Speleology*, *37*(1), 27–40. https://doi.org/10.5038/1827-806X.37.1.3

Gruber, F. E., Baruck, J., & Geitner, C. (2017). Algorithms vs. surveyors: A comparison of automated landform delineations and surveyed topographic positions from soil mapping in an Alpine environment. *Geoderma*, *308*, 9–25. https://doi.org/10.1016/j.geoderma.2017.08.017

Guisan, A., Weiss, S. B., & Weiss, A. D. (1999). GLM versus CCA spatial modeling of plant species distribution. *Plant Ecology*, *143*(1), 107–122. https://doi.org/10.1023/A:1009841519580

Haryono, E., Putro, S. T., Suratman, S., & Sutikno, S. (2017). Karst Morphology of Karangbolong Area, Java-Indonesia. *Acta Carsologica*, *46*(1), 63–72. https://doi.org/10.3986/ac.v46i1.3589

Hung, L. Q., Dinh, N. Q., Batelaan, O., Tam, V. T., & Lagrou, D. (2002). Remote sensing and GIS-based analysis of cave development in the Suoimuoi Catchment (Son La—NW Vietnam). *Journal of Cave and Karst Studies*, *64*(1), 23–33. https://www.academia.edu/download/43894299/v64n1-Hung.pdf

Julzarika, A., & Harintaka. (2019). Indonesian DEMNAS: DSM or DTM? *2019 IEEE Asia-Pacific Conference on Geoscience, Electronics and Remote Sensing Technology (AGERS)*, 31–36. https://doi.org/10.1109/AGERS48446.2019.9034351

Käss, W. (1999). *Tracing Technique in Geohydrology* (1st ed.). Routledge. https://doi.org/10.1201/9780203735282

Kusumawati, A., & Zuharnen, Z. (2017). Aplikasi Penginderaan Jauh untuk Identifikasi Mulut Goa di Kawasan Karst Kecamatan Tepus Gunungkidul [Remote Sensing Application for Identification of Cave Mouths in Karst Areas, Tepus District, Gunungkidul]. *Jurnal Bumi Indonesia*, *6*(2), 1–9. https://www.neliti.com/publications/228843/aplikasi-penginderaan-jauh-untuk-identifikasi-mulut-goa-di-kawasan-karst-kecamat

Kuswanto, G. D. (2016). *Pemodelan Pendugaan Situs Gua Arkeologis di Kawasan Karst Gunungsewu Berbasis Citra Penginderaan Jauh dan Sistem Informasi Geografis* [Archaeological Predictive Modelling of Cave Sites in Gunungsewu Karst Region Based on Remote Sensing and Geographic Information System; Master's Thesis]. Universitas Gadjah Mada.

Laksono, G. E. (2019). *Kajian Kerusakan Lingkungan Berbasis Indeks Ketergangguan Karst di Kawasan Karst Karangbolong Kabupaten Kebumen* [Study of Environmental Damage Based on Karst Disturbance Index in Karangbolong Karst Area Kebumen Regency; Master's Thesis]. Universitas Gadjah Mada.

Lee, S. (2005). Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. *International Journal of Remote Sensing*, *26*(7), 1477–1491. https://doi.org/10.1080/01431160412331331012

Muttaqin, L. A., Murti, S. H., & Susilo, B. (2019). MaxEnt (Maximum Entropy) model for predicting prehistoric cave sites in Karst area of Gunung Sewu, Gunung Kidul, Yogyakarta. In S. B. Wibowo, A. B. Rimba, S. Phinn, A. A. Aziz, J. T. Sri Sumantyo, H. Widyasamratri, & S. Arjasakusuma (Eds.), *Proceedings Volume 11311*: *Sixth Geoinformation Science Symposium* (Article 113110B). https://doi.org/10.1117/12.2543522

Ozdemir, A. (2011). Using a binary logistic regression method and GIS for evaluating and mapping the groundwater spring potential in the Sultan Mountains (Aksehir, Turkey). *Journal of Hydrology*, *405*(1–2), 123–136. https://doi.org/10.1016/j.jhydrol.2011.05.015

Pourghasemi, H. R., Moradi, H. R., & Fatemi Aghda, S. M. (2013). Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances. *Natural Hazards*, *69*(1), 749–779. https://doi.org/10.1007/s11069-013-0728-5

Putra, R. D. (2021). *Analisis Penginderaan Jauh dan Sistem Informasi Geografis untuk Identifikasi Mulut Gua Melalui Pendekatan Elemen Topografi di Kawasan Karst Gombong Selatan* [Remote Sensing and GIS Analysis of Cave Entrance Identification Based on Topographic Element Approach in Gombong Selatan Karst Area; Undergraduate Thesis]. Universitas Gadjah Mada.

Rafique, M. U., Zhu, J., & Jacobs, N. (2022). Automatic Segmentation of Sinkholes Using a Convolutional Neural Network. *Earth and Space Science, 9*(2), Article e2021EA002195. https://doi.org/10.1029/2021EA002195

Saputra, E., & Jatmiko, R. H. (2014). Aplikasi Teknik Penginderaan Jauh untuk Identifikasi Mulut Gua dan Sebarannya di Kawasan Karst Daerah Semanu, Gunungkidul, Yogyakarta [Application of Remote Sensing Techniques for Identification of Cave Mouths and Their Distribution in the Karst Area of Semanu Region, Gunungkidul, Yogyakarta]. *Jurnal Bumi Indonesia*, *3*(2), 1–7. https://www.neliti.com/publications/228517/aplikasi-teknik-penginderaan-jauh-untuk-identifikasi-mulut-gua-dan-sebarannya-di

Taheri, K., Shahabi, H., Chapi, K., Shirzadi, A., Gutiérrez, F., & Khosravi, K. (2019). Sinkhole susceptibility mapping: A comparison between Bayes-based machine learning algorithms. *Land Degradation & Development*, *30*(7), 730–745. https://doi.org/10.1002/ldr.3255

Umarhadi, D. A., & Danoedoro, P. (2019). Correcting topographic effect on Landsat-8 images: An evaluation of using different DEMs in Indonesia. In S. B. Wibowo, A. B. Rimba, S. Phinn, A. A. Aziz, J. T. Sri Sumantyo, H. Widyasamratri, & S. Arjasakusuma (Eds.), *Proceedings of SPIE: Vol. 11311. Sixth Geoinformation Science Symposium* (Article 113110L). SPIE. https://doi.org/10.1117/12.2549109

van Bemmelen, R. W. (1970). *The geology of Indonesia* (2nd ed., Vol. 1A). Martinus Nijhoff.

Weiss, A. D. (2001, July). *Topographic Positions and Landforms Analysis* [Poster presentation]. ESRI International User Conference, San Diego, CA. http://www.jennessent.com/downloads/tpi-poster-tnc_18x22.pdf

White, W. B. (1988). *Geomorphology and Hydrology of Karst Terrains*. Oxford University Press.

White, W. B., Culver, D. C., Herman, J. S., Kane, T. C., & Mylroie, J. E. (1995). Karst Lands. *American Scientist*, *83*(5), 450–459. https://www.jstor.org/stable/29775522

Widyatmanti, W., Wicaksono, I., & Syam, P. D. R. (2016). Identification of topographic elements composition based on landform boundaries from radar interferometry segmentation (preliminary study on digital landform mapping). *IOP Conference Series: Earth and Environmental Science*, *37*, Article 012008. https://doi.org/10.1088/1755-1315/37/1/012008

Zhu, J., & Pierskalla, W. P. (2016). Applying a weighted random forests method to extract karst sinkholes from LiDAR data. *Journal of Hydrology*, *533*, 343–352. https://doi.org/10.1016/j.jhydrol.2015.12.012