

NEURAL NETWORK DERIVED LANDSLIDE RISK ASSESSMENT SYSTEM

Nur Hisyam RAMLI^{1,4✉}, Siti Noor Linda TAIB¹, Norazzlina M. SA'DON¹,
Dayangku Salma AWANG ISMAIL¹, Raudhah AHMADI¹, Imtiyaz Akbar NAJAR¹,
Nazeri ABDUL RAHMAN², Rosmina AHMAD BUSTAMI³, Tarmiji MASRON⁴

¹Department of Civil Engineering, Faculty of Engineering, Universiti Malaysia Sarawak, Kota Samarahan, 94300 Sarawak, Malaysia

²Department of Chemical Engineering and Energy Sustainability, Faculty of Engineering, Universiti Malaysia Sarawak, Kota Samarahan, 94300 Sarawak, Malaysia

³UNIMAS Water Centre (UWC), Faculty of Engineering, Universiti Malaysia Sarawak, Kota Samarahan, 94300 Sarawak, Malaysia

⁴Centre for Spatially Integrated Digital Humanities (CSIDH), Faculty of Social Sciences and Humanities, Universiti Malaysia Sarawak, Kota Samarahan, 94300 Sarawak, Malaysia

Highlights:

- a transparent landslide susceptibility prediction method based on the inner workings of an Artificial Neural Network was derived in the form of equations with accompanying components;
- the method preserves the non-linear capabilities for the model, with a Root Mean Squared Error of 0.0724 and Pearson Correlation of 0.9722 when compared spatially to the model;
- the method removes the model's "black-box" dependency, resulting in an accurate landslide susceptibility predicting without the need for model inferencing.

Article History:

- received 7 March 2025
- accepted 27 February 2026

Abstract. Landslide is a frequent form of natural disaster in a tropical region such as Western Sarawak, Malaysia, and the occurrences are commonly associated with the wet season. Contemporarily, Landslide Susceptibility Map (LSM) produced by machine learning models are used to assess the landslide risk. However, LSMs are known to have spatiotemporal issues, where recent changes in the spatial data used as inputs in the models are not reflected in the LSMs, causing unjustified landslide susceptibility score in certain areas. The same issue was also observed in Western Sarawak's LSM. Thus, this study was focused on transforming the Artificial Neural Network model used developed the region's LSM into a set of equations that can be used as a landslide risk assessment system. The new system was determined to be highly accurate with a score of 1 for both Area Under the Curve score, and recall.

Keywords: environment impact assessment, environment monitoring, environment process modelling, landslide, machine learning, remote sensing.

✉Corresponding author. E-mail: 22020251@siswa.unimas.my

1. Introduction

Landslides are complex natural disasters that are commonly reported during the rainy season for a tropical country such as Malaysia. Its frequency has been observed to increase as a result of climate change which intensifies the weather conditions (Tan et al., 2019). Shallow slides are the dominant form of landslides in Malaysia, a result of weather conditions that facilitate the formation of thick surface soils, that are sensitive to heavy rainfall (Lei et al., 2024).

Landslide occurrences are not only influenced by a single variable, instead they are multivariate with each variable contributing to landslide occurrences at different weights.

Thus, to justifiably assign the landslide factors weightages, contemporary studies have opted for a Machine Learning (ML) based approach. ML provides a purely data-driven approach in determining factors weightages, making it very suitable for landslide risks assessment (Tang et al., 2021).

Commonly, the final product in assessing landslide risk via an ML based approach is a Landslide Susceptibility Map (LSM), which is a graphical representation of landslide susceptibility predicted by the ML model based on the inputs raster files (Chowdhury et al., 2024). However, LSM accuracy is limited to the time where input raster files are obtained, as recent changes may not be tracked, leading to a less accurate LSM.

One of the largely adopted ML models for the purpose of assessing landslide risks is an Artificial Neural Network (ANN). An ANN model offers a highly customizable parameters that are adjustable depending on the need (Roy & Saha, 2022). Furthermore, an ANN model decision making process is designed to mimic interconnected in the human mind, where information is passed from the input layer to the hidden layers to the output layer (Abidoeye et al., 2018). Thus, it is very interpretable to form a landslide risks assessment system.

This study was focused on developing a new Landslide Risk Assessment System (LARAS) from an ANN model for Western Sarawak, Malaysia. Western Sarawak is a region in Malaysia that is currently experiencing rapid development resulting in an increase in population density (Liaw, 2020). Without its region-specific landslide risk assessment system, it is worrying that the general population are exposed to hazards due to the depleting number of flat lands to accommodate the increase in settlements.

2. Methodology

2.1. Study area

Sarawak is the largest state in Malaysia, with an estimated area of 124,449 km², spanning 800 km along the North-west coast of Borneo. Albeit having such a widespread area, most of the state's population with 38% of 2,907,500 pupils, are located in Western Sarawak, a region with an area of 7,122 km² (Phang et al., 2023). Being home to several higher education institutions, industrial parks, and the state's capital city, Western Sarawak is a region that is currently undergoing rapid expansion. However, development does not come without its trade-off in Malaysia, as it also has been identified to be a significant contributor to landslide in the country (Rahman & Mapjabil, 2017). This was especially true as eight landslide cases were reported from

28th to 29th January of 2025, as compared to a total of 49 recorded cases in the last five years as seen in Figure 1 (Mohd Noor et al., 2025).

2.2. Data collection and preparation

Data is the basis for any ML model. The ANN model in this study was developed with a supervised learning approach to predict landslide susceptibility, thus, both features (landslide factors) and targets (landslide or non-landslide occurrence) are required. The target of 52 historical landslides was obtained from the local news station, and the 1,168 non-landslide instances were manually selected at points with slope angle at a maximum of 5°.

As the main goal of the ANN model development is to formulate the LARAS, the features must be measurable on-site and related to landslides occurrences in Western Sarawak, considering the aspect of geological, human activity, and topographical. Although rainfall is the major trigger for landslides in Western Sarawak, it was not included in the features as it does not always occur (Rosly et al., 2022). Thus, LARAS's landslide susceptibility levels can be interpreted as slopes susceptibility to failure in the presence of heavy rainfall.

Geological aspect was represented by dominant soil type with data provided by the Food and Agriculture Organization of United Nations and International Institute for Applied Systems Analysis (2023). Dominant soil type is a great representation of geological features for LARAS, as most landslide cases in the region are reported to be shallow slides at thick surface soils. Dominant soil type in Western Sarawak consists of Albic Arenosols, Dystric Gleysols, Dystric Histosols, Ferralic Cambisols, and Orthic Acrisols.

Human activities were represented through Distance from Built Areas (DFBA). DFBA was developed by extracting the built areas data from the Land Use and Land Cover

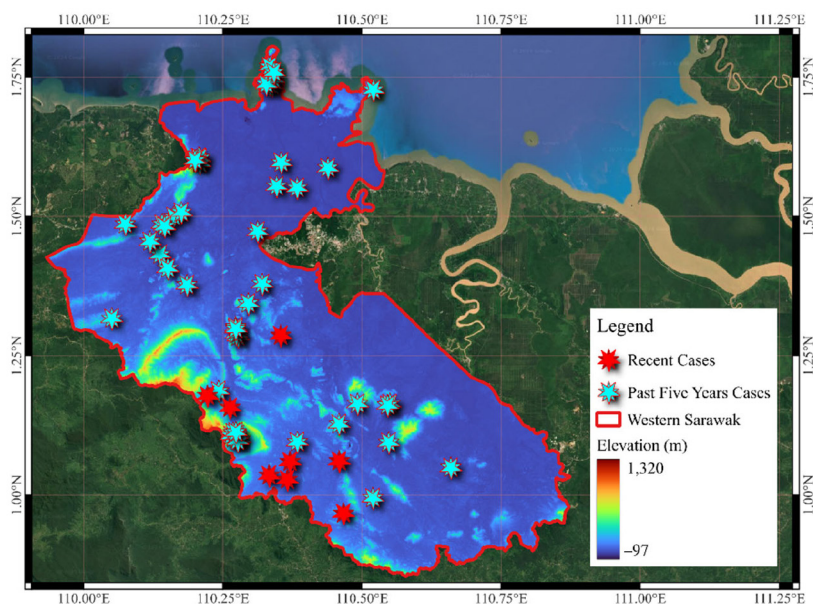


Figure 1. Western Sarawak landslide cases

(LULC) data provided by the Environmental Systems Research Institute for the year 2023, before building a proximity raster from said data (Karra et al., 2021). DFBA was a better representation in comparison to LULC for human activities, as only two out of eight LULC type in Western Sarawak have experienced landslides.

Topographical features were represented through aspect, curvature, elevation, and slope angle. The elevation raster file was provided by the National Space Agency for the year 2000, and through the System for Automated Geoscientific Analyses, the other topographical features were derived from the elevation (Farr et al., 2007).

Aspect refers to the orientation of a slope face, starting from 0° North, circling clockwise. Each orientation differs in landslide susceptibility due to the localised weather conditions (Guo & Ma, 2023). Curvature refers to the protrusion of a slope, ranging from concave to flat to convex. Conventionally, concave and flats are more susceptible to landslides due to a higher ponding tendency (Ahmed et al., 2024).

Elevation refers to the vertical distance of a given point from a datum, while slope angle refers to the steepness of slope. As elevation, and slope angle increases, the slope's stability decreases due to a higher potential energy, making it more susceptible to landslides (Moragues et al., 2024).

To create the primary dataset for the ANN model development, the values of features at each target point were extracted through the raster sampling tool in the Quantum Geographical Information System (QGIS) (Janzandr, 2020). The primary dataset was split into training data and testing data with a ratio of 80:20.

2.3. Data pre-processing

There are two types of features available in the datasets, which are categorical and numerical. The categorical variables consisted of aspects, curvature and soil type. To provide analysis ready data to the ANN model, one-hot encoding is conducted on the categorical variables, where aspect was categorized into groups with a class range of 45°, curvature was categorized into concave, flat or convex, and soil type were categorized based on the dominant soil type (UI Haq et al., 2019). The rest of the features are numerical variables that are processed through min-max scaling, a method that scales the values to be from 0 to 1, by finding the ratio between unscaled value deducted by minimum value for variable to the total range. This avoid any overpowering issues from a single variable (Lin & Billa, 2021).

2.4. ANN model development

The ANN model was developed in R using the “neuralnet” package. The hyperparameters of number of neurons in the hidden layer, learning rate, and maximum steps were adjusted to obtain the lowest Root Mean Squared Error (RMSE), while the sigmoid activation function, and back-

propagation learning algorithm were kept constant (Khajehzadeh et al., 2022).

2.5. ANN model evaluation and deployment

The ANN model was evaluated through Area Under the Curve (AUC), and recall. AUC provides an overall evaluation of accuracy, and recall is more focused on the minority landslide instances as it is a measure of true positive rate (Iban & Sekertekin, 2022).

In this study, the deployment of the ANN model was to form LARAS, and LSM of Western Sarawak, to display the importance of a spatiotemporal free system for landslide risks assessment. LARAS is a set of equations that is designed based on the ANN model to determine landslide risks in Western Sarawak without the need to access a computing device. Thus, biases and weightages of each feature assigned by the ANN model are required. The LSM on the other hand was plotted by converting the features raster files into a numerical data frame using the “terra” package in R (Waongo et al., 2025). The numerical data frame first underwent data pre-processing before the ANN model can be deployed to plot the LSM.

3. Results and discussions

3.1. ANN model hyperparameter and evaluation

RMSE is a measure of potential error in form of unit deviation from the actual values for the prediction (Ramayanti et al., 2022). In an ML model development process, the goal is to minimize the RMSE by tuning the hyperparameters. Three trials were conducted as seen in Table 1, where all trials resulted in the same RMSE of 0.025. An RMSE of 0.025 indicates that there is a potential prediction error of 2.5%, which is very acceptable in this study. Hyperparameters of trial 2 was selected for the ANN model, as it has the least number of neurons in the hidden layer with two, the largest learning rate of 0.010, and a maximum steps of 1.00E+06. This is due to the hyperparameter complexity is responsible for the LARAS complexity (Abidoye et al., 2018).

Table 1. RMSE based on ANN model hyperparameters

Trial	n	Learning rate	Stepmax	RMSE
1	4	0.010	1.00E+06	0.025
2	2	0.010	1.00E+06	0.025
3	2	0.001	1.00E+08	0.025

The ANN model was evaluated for both training phase and testing phase through AUC and recall. The training phase of the ANN model was deemed to be highly successful, an indicator that the model was able to learn comprehensively from the training data with an AUC of 0.999, and a recall of 1 with all 45 landslide cases were predicted correctly as seen in Figure 2a. During the testing phase,

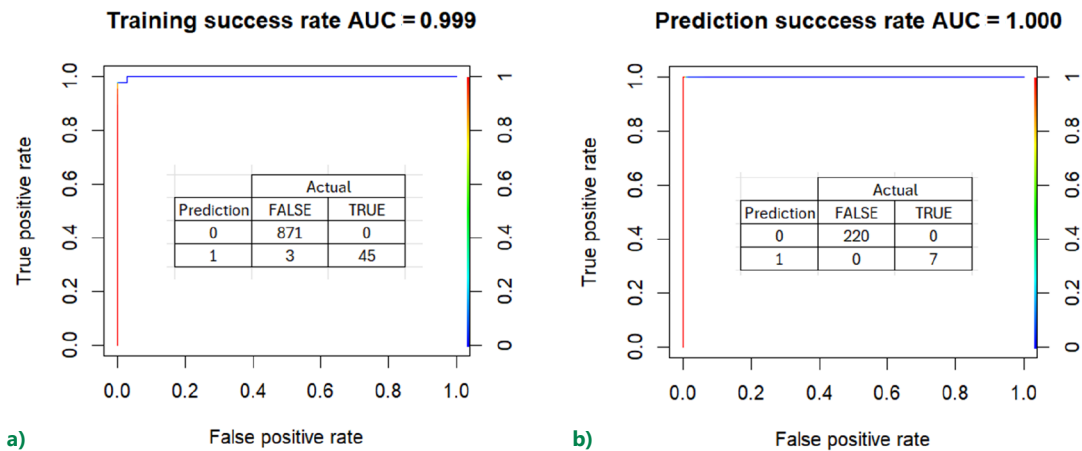


Figure 2. Performance metrics for ANN model: a) Training phase; b) testing phase

the ANN model was tested for its predictive performance based on the never-before-seen testing data. The ANN model testing phase yielded a score of 1 for both AUC and recall, making it highly successful as seen in Figure 2b.

3.2. Landslide susceptibility map and landslide risk assessment system

An LSM is a graphical representation of landslide susceptibility plotted by the ANN model with the corresponding landslide factors raster files as the inputs. Spatiotemporal issues exist in an LSM if major changes had occurred in the landslide factors raster files, post data retrieval time (Gómez et al., 2023). The LSM developed in this study as seen in Figure 3 was able to predict landslide susceptibility accurately in regions that did not undergo major changes such as the village of Tangga Mawang, where a minor landslide was observed on a point with a susceptibility score of 0.81. As for areas that have undergone major changes

such as Nirvana cemetery with its slope remediation, the LSM was less accurate as the point was predicted to have a susceptibility score of 0.81, even though the slopes have been remediated.

LARAS on the other hand was developed based on the ANN model. Thus, its application starts with data pre-processing. The min and range required for numerical variables scaling can be seen in Table 2.

Table 2. Min and range for numerical variable scaling

DFBA (m)		Elevation (m)		Slope angle (°)	
Min	Range	Min	Range	Min	Range
0	805.492	-97	1,417	0	86.850

Figure 4 shows the weights and biases that are used to determine the output at each neuron in LARAS.

To determine the final landslide susceptibility risk, the output at each neuron in the hidden layer was required as

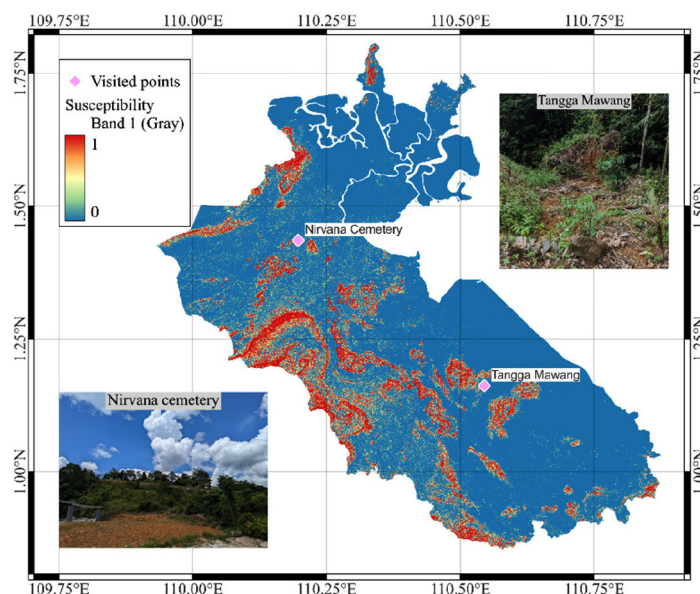


Figure 3. Western Sarawak LSM

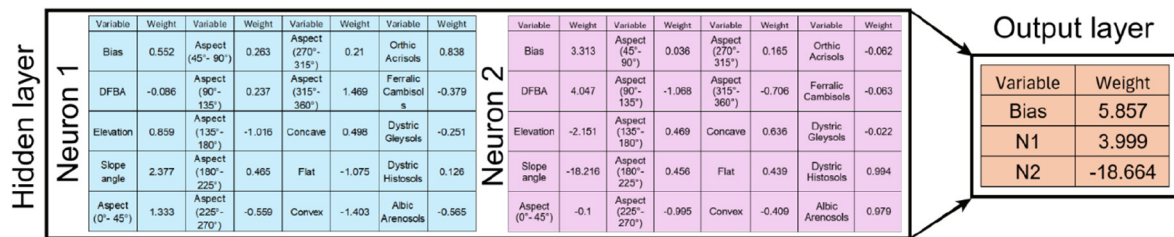


Figure 4. LARAS weights and biases

the output layer input. Each neuron's output was calculated using the same modified sigmoid activation function (Eq. 1). Evaluating LARAS's performance using the testing data has shown that the conversion was a success with an AUC and recall of 1 as seen in Figure 5.

$$O_{ut} = \frac{1}{1 + e^{-(b + \sum(w \times x_{post}))}}, \quad (1)$$

where O_{ut} is the output, b is the bias assigned by the ANN at the respective neuron, w is weightage assigned by the ANN model for an input variable, and x_{post} is the pre-processed variable value.

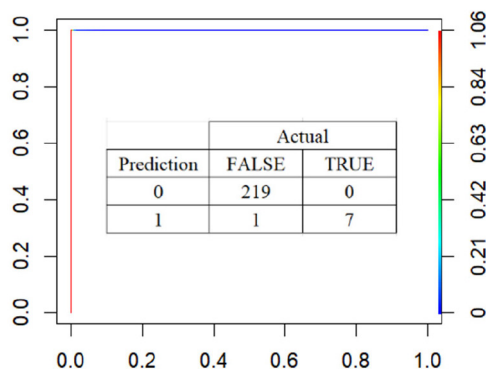


Figure 5. LARAS performance metrics

4. Conclusions

Landslide is a complex natural disaster that is contemporarily analyzed through an ML based approach. One of the main outputs from the analysis is an LSM. However, LSM accuracy is limited based on the time of landslide factors data retrieval. In this study, LARAS, a computing device-free and non-spatiotemporal landslide risk assessment system for Western Sarawak was developed from an ANN model. It was evaluated to be highly successful with predictive performance AUC of 1, and a recall 1.

Acknowledgements

The authors acknowledge Ministry of Higher Education Malaysia, Fundamental Research Grant Scheme, FRGS/1/2022/TK06/UNIMAS/01/1 and Universiti Malaysia Sarawak for supporting this project.

Author contributions

The authors confirm contribution to the paper as follows: study conception and design: Nur Hisyam Ramli, Siti Noor Linda Taib; data collection: Nur Hisyam Ramli, Imtiaz Akbar Najar; analysis and interpretation of results: Nur Hisyam Ramli, Rosmina Ahmad Bustami, Raudhah Ahmadi, Norazzlina M. Sa'don, Dayangku Salma Awang Ismail, Tarmiji Masron, Nazeri Abdul Rahman; draft manuscript preparation: All authors reviewed the results and approved the final version of the manuscript.

Disclosure statement

Authors declare that there is no conflict of interest regarding the publication of the paper.

References

- Abidoye, L. K., Mahdi, F. M., Idris, M. O., Alabi, O. O., & Wahab, A. A. (2018). ANN-derived equation and ITS application in the prediction of dielectric properties of pure and impure CO₂. *Journal of Cleaner Production*, 175, 123–132. <https://doi.org/10.1016/j.jclepro.2017.12.013>
- Ahmed, I., Das (Pan), N., Debnath, J., Bhowmik, M., & Bhattacharjee, S. (2024). Flood hazard zonation using GIS-based multi-parametric Analytical Hierarchy Process. *Geosystems and Geoenvironment*, 3(2), Article 100250. <https://doi.org/10.1016/j.geogeo.2023.100250>
- Chowdhury, S., Naimur, R., Sujon, S., Abu, S., Mahmud, K. H., & Hafsa, B. (2024). GIS-based landslide susceptibility mapping using logistic regression, random forest and decision and regression tree models in Chattogram District, Bangladesh. *Heliyon*, 10(1), Article e23424. <https://doi.org/10.1016/j.heliyon.2023.e23424>
- Food and Agriculture Organization of the United Nations & International Institute for Applied Systems Analysis. (2023). *Harmonized World Soil Database version 2.0*. <https://doi.org/10.4060/cc3823en>
- Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., & Alsdorf, D. E. (2007). The shuttle radar topography mission. *Reviews of Geophysics*, 45(2), 1–33. <https://doi.org/10.1029/2005RG000183>
- Gómez, D., García, E. F., & Aristizábal, E. (2023). Spatial and temporal landslide distributions using global and open landslide databases. *Natural Hazards*, 117, 25–55. <https://doi.org/10.1007/s11069-023-05848-8>

- Guo, Y., & Ma, C. (2023). Elucidating the role of soil hydraulic properties on aspect-dependent landslide initiation. *Hydrology and Earth System Sciences*, 27(8), 1667–1682. <https://doi.org/10.5194/hess-27-1667-2023>
- Iban, M. C., & Sekertekin, A. (2022). Machine learning based wild-fire susceptibility mapping using remotely sensed fire data and GIS: A case study of Adana and Mersin provinces, Turkey. *Ecological Informatics*, 69, Article 101647. <https://doi.org/10.1016/j.ecoinf.2022.101647>
- Janzandr. (2020). Version: [2196] One Click Raster Stacking 1.1.0. QGIS. <https://plugins.qgis.org/plugins/oneclickrasterstacking/version/1.1.0/>
- Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J. C., Mather, M., & Brumby, S. P. (2021, July 11–16). Global land use / land cover with Sentinel 2 and deep learning. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS* (pp. 4704–4707). Brussels, Belgium. IEEE. <https://doi.org/10.1109/IGARSS47720.2021.9553499>
- Khajehzadeh, M., Taha, M. R., Keawsawasvong, S., Mirzaei, H., & Jebeli, M. (2022). An effective artificial intelligence approach for slope stability evaluation. *IEEE Access*, 10, 5660–5671. <https://doi.org/10.1109/ACCESS.2022.3141432>
- Lei, D., Zhang, Y., Lu, Z., Lin, H., & Jiang, Z. (2024). Predicting factor of safety of slope using an improved support vector machine regression model. *Mathematics*, 12(20), Article 3254. <https://doi.org/10.3390/math12203254>
- Liaw, S. E. (2020). *Business and investment opportunities in Sarawak at MIDA Invest Series: Unfolding states' business potential* [PowerPoint slides]. Ministry of Industrial and Entrepreneur Development Sarawak. https://www.mida.gov.my/wp-content/uploads/2020/07/20190131110151_190128-MIED-Slides-for-MIDA-INVEST-SERIES.pdf
- Lin, J. M., & Billa, L. (2021). Spatial prediction of flood-prone areas using geographically weighted regression. *Environmental Advances*, 6, Article 100118. <https://doi.org/10.1016/j.envadv.2021.100118>
- Mohd Noor, S. N. A., Che Lat, D., Razali, R., & Zainuddin, A. N. (2025). Historical analysis of landslides events in Malaysia: Impact on physical and socio-economic. *Malaysian Journal of Civil Engineering*, 37(1), 11–21. <https://doi.org/10.11113/mjce.v37.23280>
- Moragues, S., Lenzano, M. G., Jeanneret, P., Gil, V., & Lannutti, E. (2024). Landslide susceptibility mapping in the Northern part of Los Glaciares National Park, Southern Patagonia, Argentina using remote sensing, GIS and frequency ratio model. *Quaternary Science Advances*, 13, Article 100146. <https://doi.org/10.1016/j.qsa.2023.100146>
- Phang, P., Labadin, J., Suhaila, J., Aslam, S., & Hazmi, H. (2023). Exploration of spatiotemporal heterogeneity and socio-demographic determinants on COVID-19 incidence rates in Sarawak, Malaysia. *BMC Public Health*, Article 1396. <https://doi.org/10.1186/s12889-023-16300-8>
- Rahman, H. A., & Mapjabil, J. (2017). Landslides disaster in Malaysia: An overview. *Health & the Environment Journal*, 8(1), 58–71.
- Ramayanti, S., Nur, A. S., Syifa, M., Panahi, M., Achmad, A. R., Park, S., & Lee, C. W. (2022). Performance comparison of two deep learning models for flood susceptibility map in Beira area, Mozambique. *Egyptian Journal of Remote Sensing and Space Science*, 25(4), 1025–1036. <https://doi.org/10.1016/j.ejrs.2022.11.003>
- Rosly, M. H., Mohamad, H. M., Bolong, N., & Harith, N. S. H. (2022). An overview: Relationship of geological condition and rainfall with landslide events at East Malaysia. *Trends in Sciences*, 19(8), Article 3464. <https://doi.org/10.48048/tis.2022.3464>
- Roy, J., & Saha, S. (2022). Ensemble hybrid machine learning methods for gully erosion susceptibility mapping: K-fold cross validation approach. *Artificial Intelligence in Geosciences*, 3, 28–45. <https://doi.org/10.1016/j.aiig.2022.07.001>
- Tan, M. L., Juneng, L., Tangang, Chan, W., N., & Ngai, S. T. (2019). Future hydro-meteorological drought of the Johor River Basin, Malaysia, based on CORDEX-SEA projections. *Hydrological Sciences Journal*, 64(8), 921–933. <https://doi.org/10.1080/02626667.2019.1612901>
- Tang, X., Machimura, T., Liu, W., Li, J., & Hong, H. (2021). A novel index to evaluate discretization methods: A case study of flood susceptibility assessment based on random forest. *Geoscience Frontiers*, 12(6), Article 101253. <https://doi.org/10.1016/j.gsf.2021.101253>
- Ul Haq, I., Gondal, I., Vamplew, P., & Brown, S. (2019). Categorical features transformation with compact one-hot encoder for fraud detection in distributed environment. In R. Islam, Y. S. Koh, Y. Zhao, G. Warwick, D. Stirling, C. H. Li, & Z. Islam (Eds.), *Communications in Computer and Information Science* (Vol. 996, pp. 69–80). Springer. https://doi.org/10.1007/978-981-13-6661-1_6
- Waongo, M., Ouedraogo, O. A. Y., Sodoré, A. A., & Atta, S. (2025). An agroecological zoning approach for sustainable agriculture in Burkina Faso, West Africa. *Atmospheric and Climate Sciences*, 15(2), 289–313. <https://doi.org/10.4236/acs.2025.152014>