Jurnal Kejuruteraan 37(3) 2025: 1307-1325
https://doi.org/10.17576/jkukm-2025-37(3)-17

Prediction of Soil Erodibility Factor, ROM Scale Erodibility Index (EI_{ROM}) and ROM Scale Category using Multiple Linear Regression (MLR) and Artificial Neural Network (ANN)

Adnan Derahman[a], Rohaya Alias[b*], Farah Wahida Mohd Latib[b], Muhamad Fuad Shukor[c] & Mohd Fairuz Bachok[c]

[a] *Faculty of Civil Engineering, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor, Malaysia.*

[b] *Faculty of Civil Engineering, Universiti Teknologi MARA Pahang Branch, 26400 Jengka, Pahang, Malaysia.*

[c] *Faculty of Civil Engineering, Universiti Teknologi MARA Johor Branch, 81750 Masai, Johor, Malaysia.*

Corresponding author: rohaya_alias@uitm.edu.my

Received 2 June 2024, Received in revised form 24 September 2024
Accepted 24 October 2024 , Available online 30 May 2025

ABSTRACT

Soil erosion is one of the environmental problems, which often leads to land degradation worldwide. Determination of factors that cause soil erosion involves an experimental approach that is not only highly cost, time-consuming and needs manpower work at the site but also requires appropriate equipment to perform the test. This study aims to predict soil erodibility factor, ROM scale erodibility index (EI_{ROM}) and ROM scale category based on slope features, erosion features and rainfall data using multiple linear regression (MLR) analysis and Artificial Neural Network (ANN). This study involves activities such as identifying the studied slope and rainfall stations, determination of soil erodibility factor, EI_{ROM} and ROM Scale category, physical assessment of slope and erosion features, rainfall data analysis, identification of significant slope and/or erosion features and/or rainfall data, establishment and validation of the prediction model. The input variables for the prediction model were slope features, erosion features, and rainfall data. Meanwhile, the soil erodibility factor, EI_{ROM} and ROM scale category were used as the output variables. Determination coefficient (R) has been used to evaluate prediction accuracy for both models. The results revealed that the ANN model successfully predicted the soil erodibility factor, EI_{ROM} and the ROM scale category with good accuracy and reliability compared to the MLR. Therefore, the ANN model can be used as an alternative tool in soil investigation parameters especially the soil erodibility factor with minimal field work and without laboratory work.

Keywords: Artificial neural network; multiple linear regression; prediction; soil erodibility; erosion features

INTRODUCTION

The escalating concerns in Malaysia revolve around slope stability and soil erodibility, which a range of issues including deforestation, land conversion for roadways, logging activities, and industrial or urbanisation objectives have influenced. Runoff erosivity is a prominent erosion issue in Malaysia, primarily impacted by characteristics such as high mean annual rainfall, storm frequency, and density. Malaysia receives an average of 2,250 mm of rainfall annually throughout the country (Mohammad et al. 2023). An average annual rainfall exceeding 2000 mm is susceptible to land degradation due to soil erosion (Roslan et al. 2017). The aforementioned actions have resulted in the occurrence of soil erosion, a phenomenon that has adverse implications for the natural environment and has the potential to worsen additional problems, including water contamination and the loss of habitats (Paramananthan et al. 2021). Soil erosion poses a heightened concern in regions, especially areas that are prone to experiencing severe soil erosion, such as hilly or inclined terrain, since the stability of the soil assumes a critical role in mitigating the landslide risk resulting from human activities. Currently, predicting slope failure induced by soil erosion has become a highly complex task for engineers working on-site, particularly in the absence of soil sampling techniques. The process of laboratory analysis and simulation necessitates a greater amount of time. The issues may be categorised into three primary aspects; 1. the intricate nature of soil behaviour, 2. the constraints of existing predictive models, and 3. the impact of external influences. Soil is a heterogeneous substance characterised by qualities that can exhibit substantial variation within short distances. The inherent variety of soil poses challenges in accurately forecasting its behaviour under various circumstances. Several parameters, including soil composition, moisture content, density, and the existence of organic substances, exert an influence on soil stability. The geochemical properties of the minerals comprising the soils are significant elements in assessing stability and soil erodibility in the advanced study (Lee et al. 2022). Furthermore, the behaviour of soil is non-linear and can change over time as a result of phenomena such as weathering, erosion, and biological activity (Fukuhara et al. 2024). The intricate nature of these problems necessitates the use of advanced models capable of considering a broad spectrum of factors. However, even the most sophisticated models have difficulty capturing all the subtleties of soil behaviour. Existing prediction models, encompassing both empirical and computational methodologies, possess intrinsic constraints. Historical data and observations are the foundation of empirical models, which may not always be relevant to novel or distinctive circumstances. High-performance computational models, such as finite element analysis (FEA) and discrete element modelling (DEM), necessitate substantial computational resources and precise input data (Bharti & Samui 2024). These models may exhibit sensitivity to initial conditions and assumptions, resulting in substantial fluctuations in forecasts. Furthermore, the precision of these models is frequently constrained by the quantity and quality of the accessible data, which may be limited or incomplete.Soil stability can be greatly influenced by external variables such as precipitation, seismic events, and human activity. Predicting these phenomena necessitates precise meteorological prediction and knowledge of water infiltration and its impact on soil characteristics (Torres & Dungca 2024). Industrial operations, such as building, clearing of forests, and mining, can disrupt the inherent equilibrium of soil and heighten the likelihood of failure (Robson et al. 2024). The inclusion of these external variables introduces an additional level of intricacy to the process of prediction as they are frequently unpredictable and can interact in unforeseen manners. An accurate prediction of soil erosion needs to be carried out for the development and management of an area. The limited number of parameters and a lack of evaluation criteria are major disadvantages to the use of conventional soil erosion models such as USLE and RUSLE (Avand et al. 2023). Although the determination of slope stability has been simplified with the establishment of the soil erodibility factor approach (K) (Wischmeier & Smith 1978), the conventional models still have constraints of applicability to regions and ecological conditions other than from which data were used in their development (Merritt et al. 2003). Therefore, a new approach and advanced method are required to predict soil erodibility better than the conventional methods.

Soil erosion can be estimated using machine learning methods that measure the linear or nonlinear correlation between soil erosion and its influencing variables, including slope, vegetation cover, and soil moisture (Sahour et al. 2021). Machine learning (ML) algorithms have been extensively evaluated and are gaining popularity due to their enhanced accuracy, performance, and other advantageous features, as stated by Mousavi et al. (2017). Benmakhlouf et al. (2023) applied a machine-learning model for landslide susceptibility mapping and the results showed that this model has excellent precision. Gholami et al. (2018) employed artificial neural networks (ANN) to do soil erosion estimation and erosion hazard mapping. They discovered that the ANN approach can effectively reduce expenses and study time. Furthermore, ANN can efficiently predict soil erosion at any location and time. Therefore, the objective of this study is to predict soil erodibility factor, ROM scale erodibility index (EI_{ROM}) and ROM scale category based on physical characteristics of slope features, erosion features and rainfall data using multiple linear regression (MLR) and artificial neural network (ANN).

MATERIALS AND METHODS

The framework of the study is to determine the soil erodibility factor at any location (slope) in Malaysia according to slope and erosion features. The step-by-step of the study is shown in Figure 1.

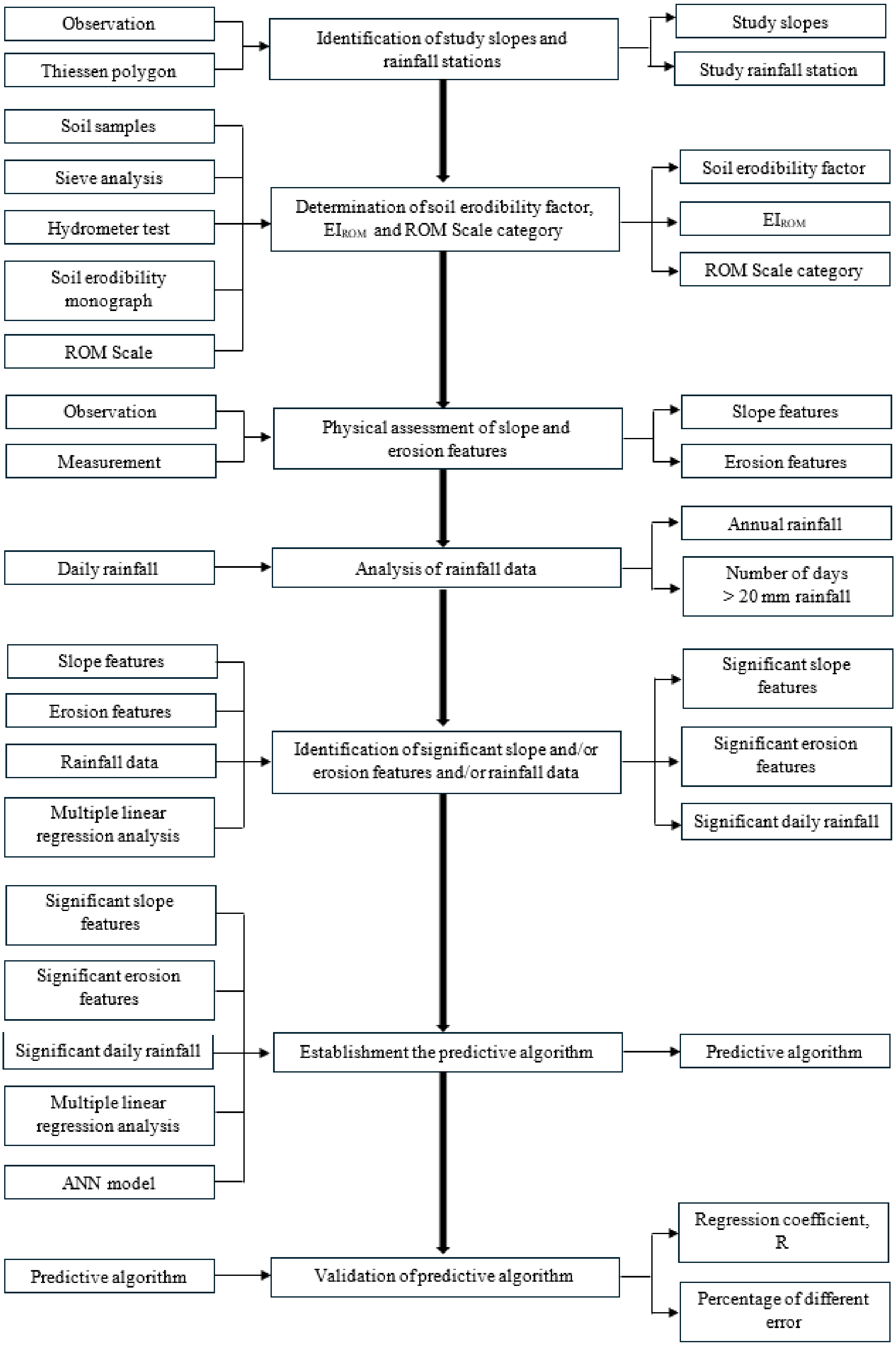

FIGURE 1. Step-by-step study methodology

IDENTIFICATION OF STUDY SLOPES AND RAINFALL STATIONS

Identification of study slopes is conducted through observation of their physical surface. Slopes with erosion features such as sheet, rill, and gully only will be considered. There are 30 slopes as study spots. Rainfall stations for each slope have also been identified through the Thiessen Polygon method using ArcGIS Pro software (Figure 2). The location of the study spots is shown in Figure 3.

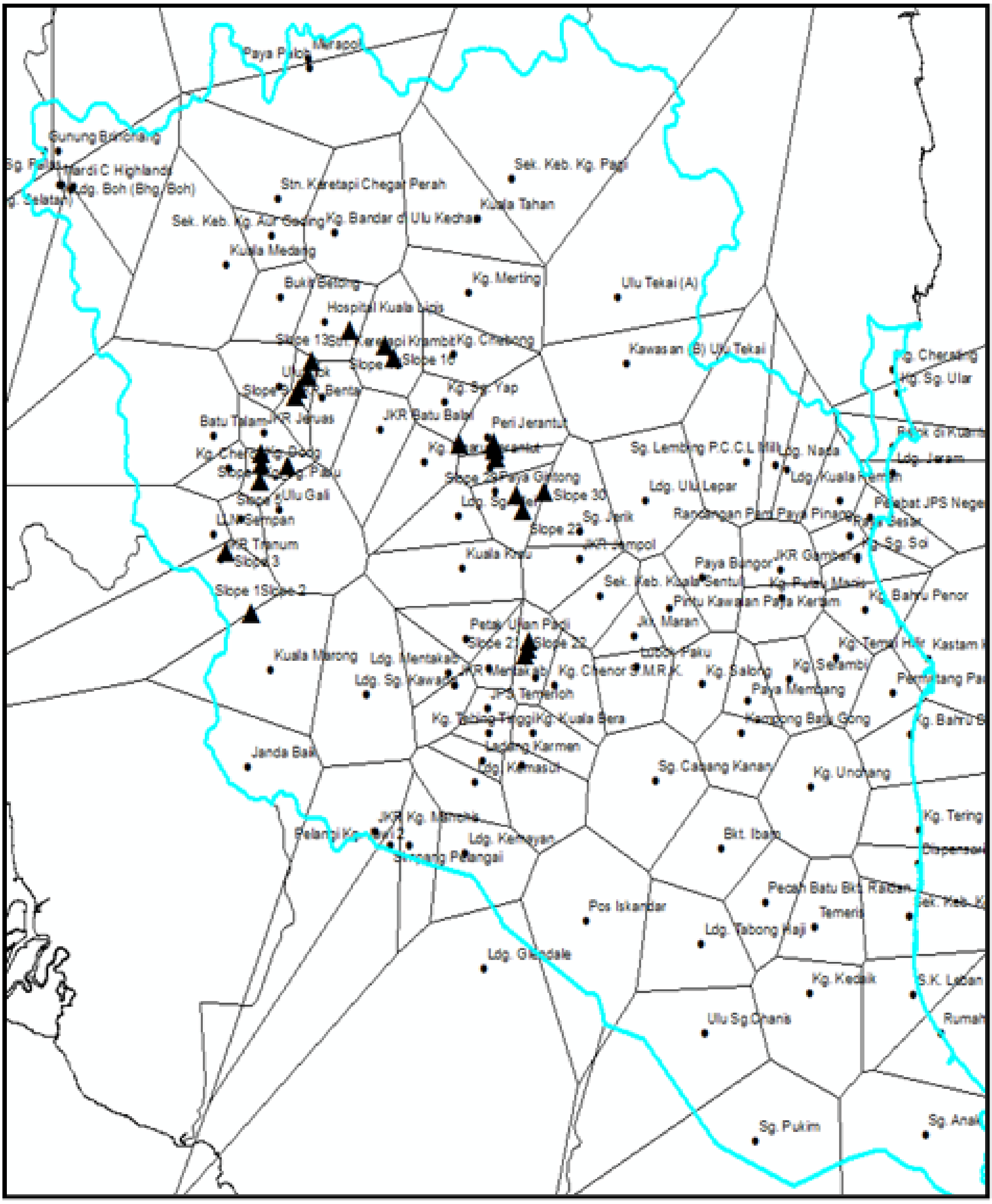

FIGURE 2. Area coverage of rainfall station

FIGURE 3. Location of study spots

DETERMINATION OF SOIL ERODIBILITY FACTOR, EI_{ROM} AND ROM SCALE CATEGORY

Two soil samples for each slope at a depth of 0.3 meters from the slope surface were collected using a hand auger. Then, the soil samples' raw data were analyzed by conducting the particle size distribution which comprises a sieve and hydrometer experiment. Sieve analysis and hydrometer test were carried out based on engineering standards BS 1377-2: 1990. According to Adriana et al. (2023), the laser diffraction method is suitable for determining the particle size distribution characterization of materials with small particles of ≤ 50 mm, in size. From the particle size distribution data, the soil erodibility factor was determined by using the Malaysian soil erodibility monograph (Tew 1999). Erodibility Index (EI_{ROM}) for each sample was determined by Equation (1) and the ROM Scale category is according to the value of EI_{ROM} in Table 1. The ROM Scale categorizes the erodibility index into five distinct categories, with values below 1.5 representing the lowest danger and values beyond 12 indicating a serious risk (Roslan & Mazidah 2002). Table 2 shows the soil erodibility factor, EI_{ROM} and ROM Scale category for each sample.

$$EI_{ROM} = \frac{(\%\,Sand + \%\,Silt)}{2\,(\%\,Clay)} \quad (1)$$

TABLE 1. ROM Scale

Erodibility Index (EI_{ROM})	ROM Scale category
< 1.5	Low
1.5 – 4.0	Moderate
4.0 – 8.0	High
8.0 – 12.0	Very high
> 12.0	Critical

Source: Roslan & Mazidah (2002)

TABLE 2. Soil erodibility, EI_{ROM} and ROM Scale category for each sample

Study spot	Soil sample	Soil erodibility factor (ton. ha.hr.MJ^{-1} $ha^{-1}.mm^{-1}$)	EI_{ROM}	ROM Scale category
Slope 1	1	0.10	5.6	High
	2	0.08	4.8	High
Slope 2	1	0.04	8.4	Very high
	2	0.04	8.4	Very high
Slope 3	1	0.05	8.8	Very high
	2	0.05	6.9	High
Slope 4	1	0.20	9.0	Very high
	2	0.20	9.0	Very high
Slope 5	1	0.20	11.8	Very high
	2	0.22	24.0	Critical
Slope 6	1	0.14	7.3	High
	2	0.14	6.1	High
Slope 7	1	0.19	11.5	Very high
	2	0.18	9.1	Very high
Slope 8	1	0.17	8.7	Very high
	2	0.16	5.3	High
Slope 9	1	0.19	6.4	High
	2	0.19	5.6	High
Slope 10	1	0.23	9.4	Very high
	2	0.22	6.6	High
Slope 11	1	0.07	7.5	High
	2	0.05	6.2	High
Slope 12	1	0.07	4.7	High
	2	0.07	3.5	Moderate
Slope 13	1	0.21	7.8	High
	2	0.21	6.6	High
Slope 14	1	0.10	6.5	High
	2	0.10	6.6	High
Slope 15	1	0.29	16.0	Critical
	2	0.29	24.3	Critical
Slope 16	1	0.15	5.9	High
	2	0.15	5.1	High
Slope 17	1	0.01	3.4	Moderate
	2	0.01	4.1	High
Slope 18	1	0.23	12.0	Critical
	2	0.24	12.0	Critical
Slope 19	1	0.25	11.4	Very high
	2	0.25	11.4	Very high
Slope 20	1	0.20	5.8	High
	2	0.20	5.8	High
Slope 21	1	0.19	6.4	High
	2	0.18	5.5	High
Slope 22	1	0.20	9.3	Very high
	2	0.20	9.3	Very high
Slope 23	1	0.22	5.8	High
	2	0.22	5.8	High
Slope 24	1	0.16	6.4	High
	2	0.15	4.8	High
Slope 25	1	0.20	8.5	Very high
	2	0.20	10.8	Very high
Slope 26	1	0.15	20.8	Critical
	2	0.14	13.7	Critical
Slope 27	1	0.25	9.3	Very high
	2	0.24	9.3	Very high
Slope 28	1	0.15	8.9	Very high
	2	0.15	8.9	Very high
Slope 29	1	0.26	11.3	Very high
	2	0.28	11.3	Very high
Slope 30	1	0.12	3.4	Moderate
	2	0.12	3.0	Moderate

contine ...

... cont.

PHYSICAL ASSESSMENT OF SLOPE AND EROSION FEATURES

Besides slope and erosion features assessed through observation, the other assessment is to measure its dimensions using a measurement apparatus such as a measuring tape, digital laser measure tool, digital vernier caliper and inclinometer. The physical assessment of slope features via observation includes shape, cutting topography, structure, main cover type and slope cover, meanwhile, the dimension of slope height, length and steepness are measured by measurement apparatus. For the erosion features, erosion type via observation and other features like erosion channel width, erosion channel depth and erosion channel direction through measurement. The dimensions of the erosion channel have been measured because of on slope land surface, the physical appearance of the channel dimensions describes the type of erosion. Rill erosion is more localized and occurs when water creates small channels in the soil typically on sloped surfaces. Gully erosion is more severe and leads to deep channels in the landscape. In the context of soil erodibility and EI_{ROM}, Rohaya et al. (2011) found that soil samples collected from slopes with gully erosion features had higher soil erodibility factor and EI_{ROM} than rill erosion. Indirectly, it indicates that there is a relationship between channel dimensions with soil erodibility factor and EI_{ROM}. Identifying the significant features through multiple linear regression analysis should be in a numerical form due to numerical output, thus some of the features need to be converted as a rating. These assessments and ratings are regarded as the input of the Slope Management and Risk Tracking (SMART) system (Public Works Department 2004) as shown in Table 3. SMART is a slope assessment system (SAS) to predict the potential of landslide occurrence based on the slope conditions. Even though there are other SAS used in the country, however, the study adopted the input in SMART since the system is satisfactory in predicting landslides with accuracy greater than 70% (Jamaludin et al. 2006). Slope and erosion features assessment for each study spot are shown in Table 4 and Table 5 respectively.

TABLE 3. Rating of slope and erosion features regarding the input of the SMART system

Slope and erosion features	Variable	Rating
Slope shape	Simple	1
	Planar	2
	Asymmetrical	3
	Compound	4
Plan profile	Convex	1
	Concave	2
	Straight	3
Structure	None	1
	Crib wall	2
	Piled wall	3
	Surface netting	4
	Soil nailing	5
	Gabion wall	6
	Rock bolts / stitching	7
	Concrete wall	8
	Masonry wall	9
	Others	10
Main cover type	Grass	1
	Shrub	2
	Fern	3
	Jungle	4
	Plantation	5
	Agricultural	6
	Others	7
Slope cover	Good (100 %)	1
	Average (80 to 100 %)	2
	Poor (< 80 %)	3
Erosion type	Sheet	1
	Rill	2
	Gully	3

ANALYSIS OF RAINFALL DATA

The main rainfall data analyzed from daily rainfall are an average annual rainfall and several days received > 20 mm. These two variables have been considered because annual rain is one of the contributing climate factors that affect the volume of soil loss (Taha & Kaniraj 2013), meanwhile, in Malaysia, the minimum amount of rainfall could lead the soil erosion process to take place is 20 mm/day (Roslan et al. 2017). These daily data were collected from the Department of Irrigation and Drainage (DID) between 2011 and 2020 (10 years). Table 6 shows the analysis of rainfall data for each study spot.

TABLE 4. Slope features for each study spot

Study spot	Soil sample	Slope height (m)	Slope length (m)	Slope steepness (degree)	Slope shape	Plan profile	Structure	Main cover type	Slope cover
Slope 1	1	8.6	8.6	45	Simple	Concave	None	Shrub	Poor
	2	6.0	6.0	45	Simple	Concave	None	Shrub	Poor
Slope 2	1	2.9	5.0	30	Asymmetrical	Concave	None	Shrub	Poor
	2	3.5	6.0	30	Asymmetrical	Concave	None	Shrub	Poor
Slope 3	1	3.0	5.2	30	Asymmetrical	Straight	None	Jungle	Average
	2	2.9	5.0	30	Asymmetrical	Straight	None	Jungle	Average
Slope 4	1	5.5	2.0	70	Asymmetrical	Straight	None	Shrub	Average
	2	6.0	2.2	70	Planar	Straight	None	Shrub	Average
Slope 5	1	22.0	8.0	70	Asymmetrical	Straight	None	Shrub	Poor
	2	11.0	4.0	70	Asymmetrical	Concave	None	Shrub	Poor
Slope 6	1	11.0	4.0	70	Simple	Concave	None	Shrub	Poor
	2	12.4	4.5	70	Planar	Concave	None	Shrub	Poor
Slope 7	1	9.9	3.6	70	Asymmetrical	Straight	None	Shrub	Poor
	2	10.2	3.7	70	Asymmetrical	Concave	None	Shrub	Poor
Slope 8	1	16.5	6.0	70	Asymmetrical	Concave	None	Shrub	Poor
	2	16.5	6.0	70	Planar	Concave	None	Shrub	Poor
Slope 9	1	19.2	7.0	70	Asymmetrical	Concave	None	Shrub	Poor
	2	19.5	7.1	70	Asymmetrical	Concave	None	Shrub	Poor
Slope 10	1	20.6	7.5	70	Asymmetrical	Straight	None	Shrub	Poor
	2	20.6	7.5	70	Asymmetrical	Straight	None	Shrub	Poor
Slope 11	1	13.7	5.0	70	Asymmetrical	Straight	None	Fern	Average
	2	13.7	5.0	70	Asymmetrical	Straight	None	Fern	Average
Slope 12	1	8.2	3.0	70	Asymmetrical	Straight	None	Shrub	Poor
	2	8.2	3.0	70	Planar	Straight	None	Shrub	Poor
Slope 13	1	19.2	7.0	70	Asymmetrical	Straight	None	Fern	Average
	2	19.2	7.0	70	Asymmetrical	Straight	None	Fern	Average
Slope 14	1	19.2	7.0	70	Asymmetrical	Concave	None	Fern	Poor
	2	19.2	7.0	70	Asymmetrical	Concave	None	Fern	Poor
Slope 15	1	22.0	8.0	70	Asymmetrical	Concave	None	Fern	Poor
	2	22.0	8.0	70	Asymmetrical	Concave	None	Fern	Poor
Slope 16	1	6.0	6.0	45	Asymmetrical	Straight	None	Shrub	Poor
	2	6.0	6.0	45	Asymmetrical	Straight	None	Shrub	Poor
Slope 17	1	3.8	3.8	45	Asymmetrical	Convex	None	Shrub	Poor
	2	3.0	3.0	45	Asymmetrical	Convex	None	Shrub	Poor
Slope 18	1	28.4	5.0	80	Asymmetrical	Straight	None	Shrub	Poor
	2	34.0	6.0	80	Asymmetrical	Straight	None	Shrub	Poor
Slope 19	1	39.7	7.0	80	Asymmetrical	Concave	None	Shrub	Poor
	2	39.7	7.0	80	Asymmetrical	Concave	None	Shrub	Poor
Slope 20	1	11.3	4.1	70	Planar	Convex	None	Grass	Poor
	2	12.1	4.4	70	Planar	Convex	None	Grass	Poor
Slope 21	1	23.1	8.4	70	Planar	Convex	None	Grass	Poor
	2	23.1	8.4	70	Planar	Straight	None	Grass	Poor
Slope 22	1	14.3	5.2	70	Planar	Convex	None	Grass	Poor
	2	14.3	5.2	70	Planar	Convex	None	Grass	Poor

contine ...

... cont.

Slope 23	1	12.4	4.5	70	Planar	Straight	None	Grass	Average
	2	12.4	4.5	70	Planar	Straight	None	Grass	Average
Slope 24	1	8.8	3.2	70	Asymmetrical	Straight	None	Shrub	Poor
	2	8.8	3.2	70	Asymmetrical	Straight	None	Shrub	Poor
Slope 25	1	17.9	6.5	70	Planar	Convex	None	Fern	Poor
	2	17.9	6.5	70	Planar	Convex	None	Fern	Poor
Slope 26	1	13.7	5.0	70	Asymmetrical	Convex	None	Shrub	Poor
	2	13.7	5.0	70	Asymmetrical	Convex	None	Shrub	Poor
Slope 27	1	6.5	6.5	45	Asymmetrical	Straight	None	Shrub	Poor
	2	6.5	6.5	45	Asymmetrical	Straight	None	Shrub	Poor
Slope 28	1	16.5	6.0	70	Asymmetrical	Straight	None	Fern	Poor
	2	16.5	6.0	70	Asymmetrical	Straight	None	Fern	Poor
Slope 29	1	22.0	8.0	70	Asymmetrical	Straight	None	Shrub	Poor
	2	22.5	8.2	70	Asymmetrical	Straight	None	Shrub	Poor
Slope 30	1	14.2	2.5	80	Asymmetrical	Straight	None	Grass	Poor
	2	14.2	2.5	80	Asymmetrical	Straight	None	Grass	Poor

TABLE 5. Erosion features for each study spot

Study spot	Sample	Erosion type	Erosion channel width (m)	Erosion channel depth (m)	Erosion channel direction (degree)
Slope 1	1	Sheet	0.40	0.20	90
	2	Rill	0.11	0.45	35
Slope 2	1	Sheet	0.25	0.15	90
	2	Rill	0.35	0.15	90
Slope 3	1	Rill	0.50	0.20	35
	2	Rill	0.40	0.30	35
Slope 4	1	Rill	0.20	0.18	90
	2	Gully	0.25	0.20	90
Slope 5	1	Rill	0.20	0.30	90
	2	Rill	0.30	0.20	135
Slope 6	1	Rill	0.30	0.13	90
	2	Rill	0.30	0.13	135
Slope 7	1	Rill	0.19	0.17	90
	2	Rill	0.14	0.12	90
Slope 8	1	Rill	0.16	0.16	90
	2	Rill	0.20	0.50	90
Slope 9	1	Sheet	0.10	0.09	90
	2	Sheet	0.12	0.08	90
Slope 10	1	Sheet	0.16	0.08	90
	2	Sheet	0.12	0.09	90
Slope 11	1	Sheet	0.30	0.08	90
	2	Sheet	0.30	0.08	90
Slope 12	1	Sheet	0.20	0.06	90
	2	Sheet	0.16	0.06	90
Slope 13	1	Sheet	0.17	0.02	90
	2	Sheet	0.20	0.05	90
Slope 14	1	Rill	0.15	0.50	135
	2	Rill	0.20	0.70	135

contine ...

... cont.

Slope 15	1	Rill	0.15	0.07	90
	2	Rill	0.20	0.13	90
Slope 16	1	Rill	0.26	0.20	90
	2	Rill	0.20	0.20	90
Slope 17	1	Rill	0.80	0.25	90
	2	Rill	0.90	0.15	90
Slope 18	1	Rill	0.30	0.13	90
	2	Rill	0.35	0.14	90
Slope 19	1	Gully	0.40	0.15	90
	2	Gully	0.40	0.18	90
Slope 20	1	Rill	0.30	0.10	90
	2	Rill	0.12	0.05	90
Slope 21	1	Rill	0.20	0.20	90
	2	Rill	0.20	0.15	90
Slope 22	1	Rill	0.22	0.12	90
	2	Rill	0.20	0.15	90
Slope 23	1	Rill	0.25	0.20	90
	2	Rill	0.23	0.15	90
Slope 24	1	Gully	0.12	0.80	90
	2	Gully	0.10	0.50	90
Slope 25	1	Gully	0.70	0.40	90
	2	Gully	0.50	0.30	90
Slope 26	1	Rill	0.80	0.18	90
	2	Rill	0.80	0.20	90
Slope 27	1	Rill	0.40	0.30	90
	2	Rill	0.50	0.15	135
Slope 28	1	Gully	0.50	0.30	90
	2	Gully	0.60	0.15	90
Slope 29	1	Gully	0.70	0.40	90
	2	Gully	0.70	0.15	90
Slope 30	1	Rill	0.30	0.10	90
	2	Rill	0.20	0.10	90

TABLE 6. Annual rainfall and number of days received > 20 mm rainfall for each study spot

Study slope	Average annual rainfall (mm)	Day received > 20 mm rainfall (no.)
Slope 1	1993	33
Slope 2	1993	33
Slope 3	1582	24
Slope 4	1548	23
Slope 5	1548	23
Slope 6	1856	28
Slope 7	1856	28
Slope 8	1856	28
Slope 9	1965	33
Slope 10	1965	33
Slope 11	1965	33
Slope 12	1965	33
Slope 13	1965	33

contine ...

... cont.

Slope 14	2207	39
Slope 15	1982	32
Slope 16	1982	32
Slope 17	2020	31
Slope 18	2020	31
Slope 19	2020	31
Slope 20	1766	27
Slope 21	1766	27
Slope 22	1766	27
Slope 23	1940	30
Slope 24	1940	30
Slope 25	2020	31
Slope 26	2020	31
Slope 27	2020	31
Slope 28	2020	31
Slope 29	1875	29
Slope 30	1892	29

IDENTIFICATION OF SIGNIFICANT SLOPE AND/OR EROSION FEATURES AND/OR RAINFALL DATA

Multiple Linear Regression (MLR) and Artificial Neural Networks (ANN) are widely used in various fields to predict and analyse complex relationships between variables. Both MLR and ANN models are prevalent in numerous applications, as they offer the ability to model and forecast intricate relationships among variables. MLR is suitable for simpler problems where relationships are linear, and interpretability is important. Its performance can be limited if the data exhibits non-linear patterns or complex interactions whereas ANN excels in handling complex, non-linear relationships and can achieve higher accuracy in many scenarios. However, it requires more computational resources, careful tuning, and may offer less interpretability. The choice between MLR and ANN depends on the nature of the problem, the data, and the specific requirements for performance and interpretability.

The MLR analysis has been used to identify the significant slope and/or erosion features and/or rainfall data through SPSS version 26 software. This inferential statistics method has been selected since the method can estimate the relationship among variables by formulating the linear relation equation between dependent and independent variables as a result relation (Uyanik & Guler 2013). These significant variables have been applied to establish a predictive algorithm using multiple linear regression analysis and ANN model. On the other hand, although the ANN model is capable to establish a predictive algorithm from any inputs without identifying the significant inputs beforehand, it could create this predictive algorithm might utilize also the unrelated inputs to simulate the output. This causes the predictive algorithm to be more complicated and waste time in determining these unrelated inputs. Therefore, that is why MLR is adopted in the study to determine the significant inputs.

ESTABLISHMENT OF THE PREDICTIVE ALGORITHM

The ability of multiple linear regression analysis to create a relationship between independent variables to a single dependent variable by summarizing the relationship of a set of predictors to the observed criterion (Equation 2) made this learning tool able to establish the predictive algorithm (Aiken et al. 2003). On the other hand, regression models where one dependent variable had a relationship with more than independent variables (Uyanik & Guler 2013).

Criterion variable score,

$$\hat{Y} = b_1X_1 + b_2X_2 + \cdots + bkXk + a \quad (2)$$

where X = predictor variable
b = regression coefficient for each predictor variable
a = regression constant

ANN model has been selected as another learning tool to establish the predictive algorithm due to its ability to combine linear and non-linear for complex relationships between variables and is widely used for prediction in areas related to soil engineering such as Shahin et al. (2001); Zhiming et al. (2018) and Lendo-Siwicka et al. (2023). ANN model requires inputs and targets (outputs) so that it can establish the algorithm to simulate the outputs. As a result, Figure 4 shows the units in establishing the ANN model back-propagation predictive algorithms.

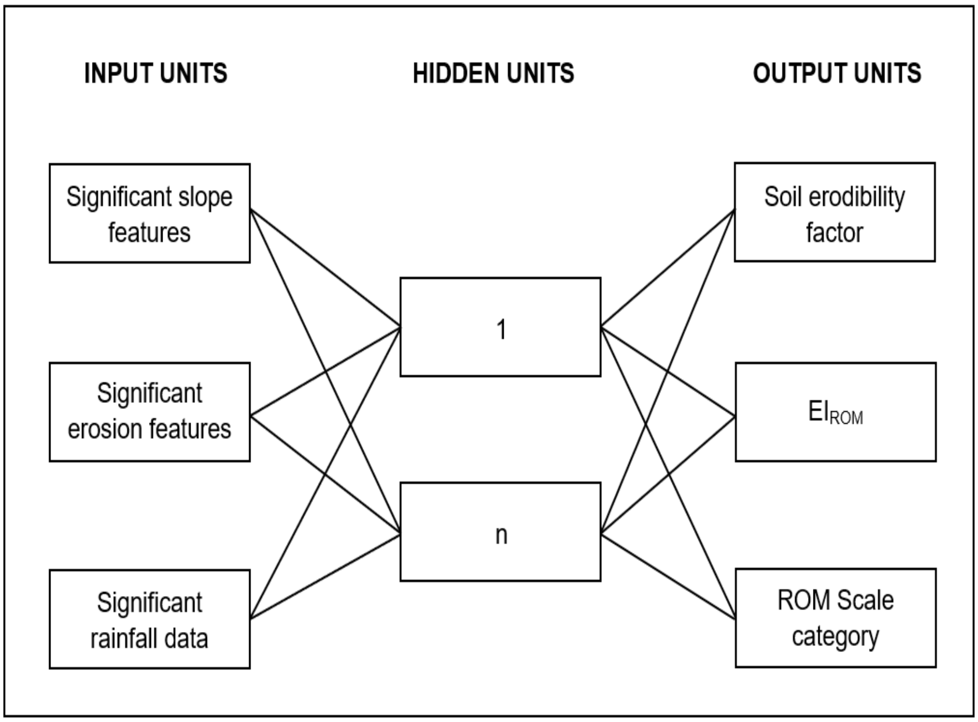

FIGURE 4. A network architecture of ANN model back-propagation predictive algorithm

VALIDATION OF A PREDICTIVE ALGORITHM

In order to measure the precision of output likely to be predicted by the predictive algorithms established by multiple linear regression analysis and ANN Model, the R-value is to be the indicator (Equation 3) and the different error in percentage (Equation 4). R-value is a number between 0 and +1 which measures the degree of association and regression strength between two variables, the observed data, and predicted data (Table 7). On the other hand, it measures the performance of the trained model by performing a linear regression analysis between model output and target output. The different error method is to validate the predictive algorithm If the different errors are low in percentage, therefore it indicates that the predictive algorithm has the potential to predict at high precision. If otherwise, the improvement needs to be worked out in the algorithm.

Regression coefficient,

$$R = (\beta_1 ry_1 + \beta_2 ry_2 + \cdots + \beta_k ry_k)^{1/2} \tag{3}$$

where β = beta coefficient
r = correlation between the criterion variable

$$Error\ (\%) = \frac{(Estimated - Observed)}{Observed}\ x\ 100 \tag{4}$$

TABLE 7. Strength of the regression coefficients

Regression coefficient size	Regression strength
.91 – 1.00 or -.91 – -1.00	Very strong
.71 – .90 or -.71 – -.90	Strong
.51 – .70 or -.51 – -.70	Average/medium
.31 – .50 or -.31 – -.50	Weak
.01 – .30 or -.01 – -.30	Very weak
.00	No correlation

Source: (Piaw 2013)

RESULTS, ANALYSIS AND DISCUSSION

All these study slopes have been assessed its slope and erosion features where slope features consist of slope height, slope length, slope steepness, cutting topography, structure, main cover type and slope cover whereas erosion features consist of erosion type, erosion channel width and erosion channel direction. These slope and erosion feature assessments for each study slope will be an input in the derivation of predictive algorithm either using multiple linear regression analysis or ANN model.

The performance of the predictive algorithm established by multiple linear regression analysis in predicting soil erodibility factor and EI_{ROM} are shown in Tables 8 and 9 respectively. It can be concluded that the slope height and number of days with rainfall > 20 mm variables given the highest R-values. Meanwhile, the performance of the predictive algorithm established by multiple linear regression analysis in predicting ROM Scale category and established by the ANN model in predicting soil erodibility factor are shown in Table 10 and Table 11 respectively. As a result, slope height, erosion channel width and number of days with rainfall > 20 mm variables were given the highest R-values. Nevertheless, the performance of the predictive algorithm established by the ANN model in predicting the EI_{ROM} and ROM Scale category is shown in Tables 12 and Table 13. Slope height, erosion channel width and number of days with rainfall > 20 mm variables given the highest R-values.

TABLE 8. Performance of predictive algorithm established by multiple linear regression analysis in predicting soil erodibility factor

Predictive algorithm	Variables	R-value	Correlation strength	Verification (Different error method)		
				Minimum error (%)	Maximum error (%)	Average error (%)
$Y = 0.005X_1 + 0.09$	Y = Soil erodibility X_1 = Slope height	0.602	Medium	1.3	33.0	15.4
$Y = 0.016X_1 + 0.076$	Y = Soil erodibility X_1 = Slope length	0.399	Weak	0.2	113.6	29.9
$Y = 0.003X_1 - 0.019$	Y = Soil erodibility X_1 = Slope steepness	0.532	Medium	0.5	34.1	13.2
$Y = 0.033X_1 + 0.099$	Y = Soil erodibility X_1 = Erosion type	0.292	Very weak	1.0	43.1	21.7
$Y = 0.073X_1 – 0.201X_2 - 0.131X_3 + 0.001X_4 + 0.022$	Y = Soil erodibility X_1 = Erosion type X_2 = Erosion channel width X_3 = Erosion channel depth X_4 = Erosion channel direction	0.589	Medium	4.4	24.6	16.9
$Y = 0.006X_1 – 0.006X_2 + 0.271$	Y = Soil erodibility X_1 = Slope height X_2 = Number of day with rainfall > 20 mm	0.668	Medium	0.0	29.3	17.3

TABLE 9. Performance of predictive algorithm established by multiple linear regression analysis in predicting EI_{ROM}

Predictive algorithm	Variables	R-value	Correlation strength	Verification (Different error method)		
				Minimum error (%)	Maximum error (%)	Average error (%)
$Y = 0.16X_1 + 6.2$	Y = EIROM X_1 = Slope height	0.306	Weak	10.3	178.4	57.6
$Y = 1.981X_1 + 3.247$	Y = EIROM X_1 = Slope shape	0.262	Very weak	7.4	202.1	61.6

TABLE 10. Performance of predictive algorithm established by multiple linear regression analysis in predicting ROM Scale category

Predictive algorithm	Variables	R-value	Correlation strength	Verification (hit the category) (%)
$Y = 0.029X_1 + 3.115$	Y = ROM Scale category X_1 = Slope height	0.301	Very weak	66.7
$Y = 0.371X_1 + 2.797$	Y = ROM Scale category X_1 = Erosion type	0.293	Very weak	83.3
$Y = 0.995X_1 + 3.218$	Y = ROM Scale category X_1 = Erosion channel width	0.255	Very weak	50.0
$Y = 0.03X_1 + 1.063X_2 + 2.758$	Y = ROM Scale category X_1 = Slope height X_2 = Erosion channel width	0.405	Weak	66.7

TABLE 11. Performance of predictive algorithm established by ANN model in predicting soil erodibility factor

Main variables	Sub-main variables	Predictive algorithm	R-value	Correlation strength	Verification (Different error method)		
					Minimum error (%)	Maximum error (%)	Average error (%)
Slope features	Slope height	Net100_erodi_slop_heig	0.9057	Very strong	0.0	7.8	1.3
Slope features	Slope length	Net100_erodi_slop_leng	0.7492	Strong	0.0	33.7	10.5
Slope features	Slope steepness	Net100_erodi_slop_stee	0.5420	Medium	3.2	36.6	15.4
Erosion features	Erosion type	Net100_erodi_eros_type	0.2967	Very weak	2.0	42.5	21.7
Erosion features	Erosion type Erosion channel width Erosion channel depth Erosion channel direction	Net100_erodi_eros_feat	0.7177	Strong	0.9	26.4	12.5
Slope features + rainfall	Slope height Number of day with rainfall > 20 mm	Net100_erodi_slhe_ra20	0.9801	Very strong	0.0	32.7	9.1

TABLE 12. Performance of predictive algorithm established by ANN model in predicting EI_{ROM}

Main variables	Sub-main variables	Predictive algorithm	R-value	Correlation strength	Verification (Different error method)		
					Minimum error (%)	Maximum error (%)	Average error (%)
Slope features	Slope height	Net100_eirom_slop_heig	0.6864	Medium	0.2	14.7	3.0
Slope features	Slope shape	Net100_eirom_slop_shap	0.2657	Very weak	7.8	236.8	68.1

TABLE 13. Performance of predictive algorithm established by ANN model in predicting ROM Scale category

Main variables	Sub-main variables	Predictive algorithm	R-value	Correlation strength	Verification (hit the category) (%)
Slope features	Slope height	Net100_romsc_slop_heig	0.7797	Strong	100.0
Erosion features	Erosion type	Net100_romsc_eros_type	0.3059	Weak	83.3
Erosion features	Erosion channel width	Net100_romsc_eros_chwi	0.4388	Weak	50.0
Slope features + erosion features	Slope height Erosion channel width	Net100_romsc_slhe_cnwi	0.7860	Strong	100.0

Potential and suggestion of predictive algorithms by utilizing these two machine learning tools in predicting soil erodibility, EI_{ROM} and ROM Scale category are as shown in Tables 14 and Table 15 respectively. All the predictive algorithms derived from multiple regression analysis were not selected since predictive algorithms derived from the ANN model showed better performance and accuracy of prediction value. However, based on the comparison between these three predictive algorithms derived from the ANN model, only the predictive algorithm for predicting soil erodibility and ROM Scale category can predict with high accuracy by using slope features, erosion features and rainfall data as a predictor. Soil erodibility predictive algorithm has advantages due to not only this factor can be predicted using erosion features as predictors but also slope features and rainfall data as well. Therefore, this predictive algorithm that uses both slope features and rainfall data as predictand can be used as a validation algorithm. Although the EI_{ROM} was unable to be predicted, this matter does not contribute such a significant effect since the EI_{ROM} is only a threshold to classify the ROM Scale Category. Besides, this category can be predicted. Moreover, the classification of the ROM Scale Category is more important than EI_{ROM} because the selection of mitigation measures for the stability of a slope is based on the category. Prediction potential erosion-induced landslide locations by determining the soil susceptibility for failure in terms of its soil erodibility index value with regards to the ROM Scale category, in purpose is to identify the high-risk locations, so that these locations should be prioritized in taking proper mitigation measures to minimize the risk of erosion induced landslide (Roslan & Zulkifli 2005). Therefore, soil erodibility and ROM Scale category are two parameters that can be determined through this approach by utilizing the ANN model predictive algorithm and slope features, erosion features and rainfall data as a predictor.

Figure 5 below shows a comparison between the current and proposed practices. The current practice takes significantly longer and is much more costly compared to the proposed practice. Starting with the soil sample collection process, sieve analysis and hydrometer tests for each sample can take several weeks to complete. Therefore, in most cases involving this scope of work, the duration required for soil characterization, analysis, and investigation is very critical, especially when dealing with a large number of soil samples.

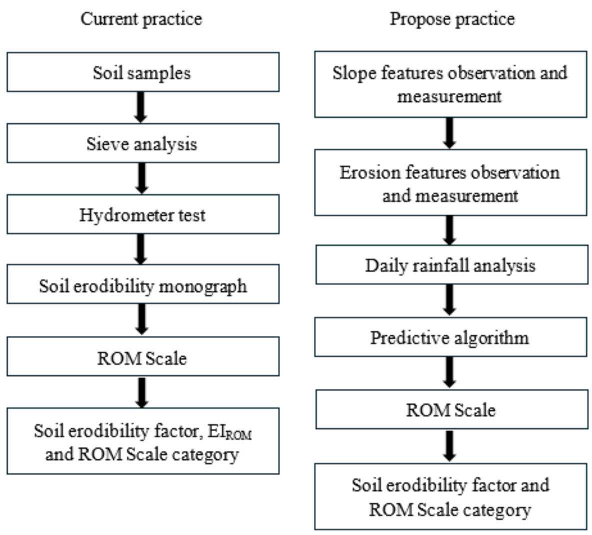

FIGURE 5. Comparison between current and propose practice in determining soil erodibility factor, EI_{ROM} and ROM Scale category

TABLE 14. The potential of predictive algorithms in predicting soil erodibility, EI_{ROM} and ROM Scale category

Predictor	Predictive algorithm	As predictive algorithm	Remarks
Soil erodibility	$Y = 0.005X_1 + 0.09$	No	All the predictive algorithms derived from multiple regression analysis were not selected since there are predictive algorithms derived from the ANN model showed better performance and high accuracy of prediction value
	$Y = 0.016X_1 + 0.076$	No	
	$Y = 0.003X_1 - 0.019$	No	
	$Y = 0.033X_1 + 0.099$	No	
	$Y = 0.073X_1 – 0.201X_2 - 0.131X_3 + 0.001X_4 + 0.022$	No	
	$Y = 0.006X_1 – 0.006X_2 + 0.271$	No	
	Net100_erodi_slop_heig	Yes	Good performance and one predictand
	Net100_erodi_slop_leng	Yes	Good performance and one predictand
	Net100_erodi_slop_stee	No	Low performance and accuracy of prediction value
	Net100_erodi_eros_type	No	Low performance and accuracy of prediction value
	Net100_erodi_eros_feat	Yes	Good performance and the algorithm also used all the erosion feature variables as predictand
	Net100_erodi_slhe_ra20	Yes	Good performance and the algorithm used between slope feature and triggering factor variables as predictand
EI_{ROM}	$Y = 0.16X_1 + 6.2$	No	Low performance and accuracy of prediction value
	$Y = 1.981X_1 + 3.247$	No	Low performance and accuracy of prediction value
	Net100_eirom_slop_heig	No	Low performance and accuracy of prediction value
	Net100_eirom_slop_shap	No	Low performance and accuracy of prediction value
ROM Scale category	$Y = 0.029X_1 + 3.115$	No	All the predictive algorithms derived from multiple regression analysis were not selected since there are predictive algorithms derived from the ANN model showed better performance and highly accurate prediction value
	$Y = 0.371X_1 + 2.797$	No	
	$Y = 0.995X_1 + 3.218$	No	
	$Y = 0.03X_1 + 1.063X_2 + 2.758$	No	
	Net100_romsc_slop_heig	Yes	Good performance and one predictand
	Net100_romsc_eros_type	No	Low performance and accuracy of prediction value
	Net100_romsc_eros_chwi	No	Low performance and accuracy of prediction value
	Net100_romsc_slhe_cnwi	Yes	Good performance and the algorithm used between slope and erosion feature variables as predictand

TABLE 15. Suggestion predictive algorithms for prediction of soil erodibility, EI_{ROM} and ROM Scale category

Predictor	Predictive algorithm	Sub-main variables	Remarks
Soil erodibility	Net100_erodi_eros_feat and Net100_erodi_slhe_ra20	Erosion type Erosion channel width Erosion channel depth Erosion channel direction Slope height Number of days with rainfall > 20 mm	Main predictive algorithms Average value between these two predictive algorithms as a predictive value
	Net100_erodi_slop_heig	Slope height	Validation of a predictive algorithm to main predictive algorithms
	Net100_erodi_slop_leng	Slope length	Validation of a predictive algorithm to main predictive algorithms

contine ...

... cont.

EI_{ROM}	-	-	None
ROM Scale category	Net100_romsc_slhe_cnwi	Slope height Erosion channel width	Main predictive algorithms
	Net100_romsc_slop_heig	Slope height	Validation of a predictive algorithm to main predictive algorithms

CONCLUSION

The objective of the study is to examine the correlation between soil erodibility factor, EI_{ROM} and ROM Scale category with physical characteristics of slope features, erosion features and rainfall data. The correlation validity was subsequently verified through the algorithm which was derived by utilizing multiple linear regression analysis or ANN model. As is widely recognized, these three parameters as a crucial metric to mitigate the exacerbation of phenomena such as slope stability. Conventionally, the acquisition of these parameters has posed significant challenges due to field and laboratory works which led to a lack of emphasis on its inclusion as a primary input for soil investigation by many agencies. These fields and laboratory work require more time, energy, and cost. Thus, by considering the advancement of artificial intelligence technology in facilitating and speeding up results, this study was conducted to find out the potential ability of this machine learning tool to determine the value of these three parameters without effectively being associated with the conventional method.

According to the study, findings showed that the ANN model had more potential than multiple linear regression analysis as a machine tool to be utilized in producing predictive algorithms for these three parameters. Nevertheless, the ANN model predictive algorithm also used fewer predictors. However, in comparison between these three parameters, predictive algorithms for soil erodibility factor and ROM Scale category only be able to be predicted by ANN model predictive algorithm with physical characteristics of slope features, erosion features and rainfall data as predictors. Meanwhile, prediction of EI_{ROM} needs other than study variables of physical characteristics of slope features, erosion features and rainfall data as predictors or utilizing another machine learning tool such as Fuzzy Logic. ANN model predictive algorithm for soil erodibility factor more likely predicts at high precision since not only used variables from physical characteristics of slope features, erosion features and rainfall data as predictors but there are also predictive algorithms that can validate the prediction value.

Hence, it can be concluded that the utilization of machine learning tools such as ANN models had the potential to assess the soil investigation parameters especially the soil erodibility factor with minimal field work and without laboratory work. This assessment was previously regarded as challenging and burdensome by certain stakeholders.

ACKNOWLEDGEMENT

The authors would like to thank the Research Management Centre (RMC), Universiti Teknologi MARA (UiTM) for the financial support through Special Research Grant/Geran Penyelidikan Khas UiTM 2020 RMI file number: 600-RMC/GPK 5/3 (215/2020).

DECLARATION OF COMPETING INTEREST

None.

REFERENCES

Adriana, E.A., Hidayati, A., Habib, M.M., Hassanel, Z.A. & Nazrein, A.A. 2023. Physicochemical and microstructural characterization of klias peat, lumadan POFA, and GGBFS for geopolymer based soil stabilization. *HighTech and Innovation Journal* 4(2): 327-348.

Aiken, L.H., Clarke, S.P., Cheung, R.B., Sloane, D.M. & Silber, J.H. 2003. Educational levels of hospital nurses and surgical patient mortality. *Journal of American Medical Association* 290(12): 1617-1623.

Avand, M., Mohammadi, M., Mirchooli, F., Kavian, A. & Tiefenbacher, J. 2023. A new approach for smart soil erosion modeling: Integration of empirical and machine-learning models. *Environ Model Assess* 28: 145-160.

Benmakhlouf, M., El Kharim, Y., Galindo-Zaldivar, J. & Sahrane, R. 2023. Landslide susceptibility assessment in western external rif chain using machine learning methods. *Civil Engineering Journal* 9(12): 3218-3232.

Bharti, J.P. & Samui, P. 2024. Predicting slope failure with intelligent hybrid modeling of ANFIS with GA and PSO. *Multiscale and Multidiscip. Model. Exp. and Des.* 7, 4539-4555.

Fukuhara, M., Uchimura, T., Wang, L., Tao, S. & Tang, J. 2024. Study on the prediction of slope failure and early warning thresholds based on model tests. *Geotechnics* 4: 1-17.

Gholami, V., Booij, M.J., Nikzad Tehrani, E. & Hadian, M.A. 2018. Spatial soil erosion estimation using an artificial neural network (ANN) and field plot data. *CATENA* 163: 210-218.

Jamaludin, S., Huat, B. B. K. & Omar, H. 2006. Evaluation of slope assessment systems for predicting landslides of cut slopes in granitic and meta-sediment formations. *American Journal of Environmental Sciences* 2(4): 135-141.

Lee, S., Chu, M.L., Guzman, J.A. & Flanagan, D.C. 2022. Modelling soil erodibility and critical shear stress parameters for soil loss estimation. *Soil and Tillage Research* 218: 105292.

Lendo-Siwicka, M., Zabłocka, K. Sobol, E., Markiewicz, A. & Wrzesinski, G. 2023. Application of an artificial neural network (ANN) model to determine the value of the damping ratio (D) of clay soils. *Applied Sciences* 13: 1-14.

Merritt, W.S., Letcher, R.A. & Jakeman, A.J. 2003. A review of erosion and sediment transport model. *Environmental Modelling and Software* 18: 761-799.

Mohammad, H.R., Habib, M.M., Nurmin, B., & Noor, S.H.H. 2023. Relationship of rainfall intensity with slope stability. *Civil Engineering Journal* 9: 75-82.

Mousavi, S.M., Golkarian, A., Naghibi, S.A., Kalantar, B. & Pradhan, B. 2017. GIS-based groundwater spring potential mapping using data mining boosted regression tree and probabilistic frequency ratio models in Iran. *AIMS Geosciences* 3(1): 91-115.

Paramananthan, S., Nurfashareena, M. & Pereira, J. J. 2021. Soil related factors controlling erosion and landslides in Malaysia. *Bulletin of the Geological Society of Malaysia* 72: 165-175.

Piaw, C.Y. 2013. Mastering research statistics. McGraw-Hill Education (Malaysia) Sdn. Bhd. Kuala Lumpur, Malaysia.

Public Works Department 2004. Slope protection study for Federal Route 22, Tamparuli: Sandakan, Sabah. Public Works Department, Kuala Lumpur, Malaysia.

Robson, E., Milledge, D., Utili, S. & Dattola, G. 2024. A computationally efficient method to determine the probability of rainfall-triggered cut slope failure accounting for upslope hydrological conditions. *Rock Mech Rock Eng* 57: 2421–2443.

Rohaya, A., Mohd Fairuz, B., Noraida, M.S. & Wan Zukri, W.A. 2011. Soil erodibility assessment at Ibu Pejabat Polis Kontijen (IPPK) Shah Alam, Selangor. *Proceedings of the Konferensi Akademik 2011, Jengka, Pahang, Malaysia*. 260-265.

Roslan, Z.A. & Mazidah, M. 2002. Establishment of soil erosion scale with regards to soil grading characteristic. *Proceedings of the 2nd World Engineering Congress, Sarawak, Malaysia*. 235–239.

Roslan, Z.A. & Zulkifli, A.H. 2005. 'ROM' scale for forecasting erosion induced landslide risk on hilly terrain. In: Sassa, K., Fukuoka, H., Wang, F., Wang, G. (Eds.) Landslides (pp. 197-202). Springer.

Roslan, Z.A., Mohd Sofiyan, S. & Yusoff, N. 2017. Erosion risk assessment: A case study of the Langat riverbank in Malaysia. *International Soil and Water Conservation Research* 5(1): 26-35.

Sahour, H., Gholami, V., Vazifedan, M. & Saeedi, S. 2021. Machine learning applications for water-induced soil erosion modeling and mapping. *Soil and Tillage Research* 211: 105032.

Shahin, M.A., Jaksa, M.B. & Maier, H.R. 2001. Artificial neural network applications in geotechnical engineering. *Australian Geomechanics Journal* 36(1): 49–62.

Taha, F. & Kaniraj, R.S. 2013. Study of soil erosian at a site near chemical engineering laboratory in UNIMAS. *Journal of Civil Engineering, Science and Technology* 4(2): 1-6.

Tew, K.H. 1999. Production of Malaysian soil erodibility nomograph in relation to soil erosion issues. VT Soil Erosion Research and Consultancy, Selangor, Malaysia.

Torres, E. & Dungca, J. 2024. Prediction of soil liquefaction triggering using rule-based interpretable machine learning. *Geosciences* 14: 156.

Uyanik, G.K. & Guler, N. 2013. A study on multiple linear regression analysis. *Procedia – Social and Behavioral Sciences* 106: 234-240.

Wischmeier, W.H. & Smith, D.D. 1978. Predicting rainfall erosion losses - A guide to conservation planning. USDA Agricultural Handbook 537.

Zhiming, C., Guotao, M., Ye, Z., Yanjie, Z. & Hengyang, H. 2018. The application of neural network in geotechnical engineering. *IOP Conference Series: Earth and Environmental Science* 189: 1–6.