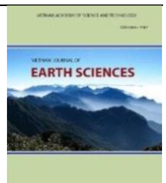
Vietnam Journal of Earth Sciences, 42(3), 247–254, Doi: 10.15625/0866-7187/42/3/15050

(VAST)

Vietnam Academy of Science and Technology

Vietnam Journal of Earth Sciences

http://www.vjs.ac.vn/index.php/jse



Prediction of soil loss due to erosion using support vector machine model

Van Quan Tran[1*], Indra Prakash[2]

[1]*University of Transport Technology, Hanoi 100000, Vietnam*
[2]*Department of Science & Technology, Bhaskarcharya Institute for Space Applications and Geo-Informatics (BISAG), Government of Gujarat, Gandhinagar 382002, India*

Received 10 April 2020; Received in revised form 21 May 2020; Accepted 25 June 2020

ABSTRACT

Soil erosion refers to the detachment and removal of soil particles from land (topsoil), by the natural physical forces (water, glacier and wind). Soil erosion causes soil loss in the catchment or any land areas severely impacting agriculture activity, sedimentation in the dam reservoirs, and hampering developmental activities. Therefore, it is desirable to accurately measure soil loss due to erosion for the development and management of an area. With this objective, a well-known machine learning algorithm Support Vector Machine (SVM) has been used in the development of the soil loss prediction model. Eight erosion affecting variable inputs: ambient temperature Tair, rainfall, Antecedent Moisture Conditions (AMC), rainfall intensity, slope, vegetation cover, soil temperature Tsoil and moisture of the soil. Data on published literature was used in the model study. The accuracy of the proposed SVM was assessed by using three statistical performance evaluation indicators namely Person correlation coefficient (R), Root Mean Squared Error (RMSE), Mean Squared Error (MAE). Partial Dependence Plots (PDP) was used to investigate the dependence of prediction results of eight input variables used in the model study. Model validation results showed that SVM model performed well for the prediction of soil loss for testing ($R = 0.8993$) and also for training ($R = 0.9123$). Rainfall intensity and vegetation cover were found to be the two most important affecting input parameters for the soil loss prediction.

Keywords: soil loss; support vector machine; machine learning; partial dependence plots; soil degradation.

1. Introduction

Soil erosion is a gradual process of movement and transport of topsoil by different agents such as water, wind, and mass movement. Soil erosion is considered a dominant cause of soil degradation (Borrelli et al., 2018). There is a significant need for knowledge of soil erosion assessment of an area to help environment managers for preventing degradation and protecting ground soil. Soil erosion is a complex process that mainly depends on soil properties, ground slope, vegetation, and rainfall amount and intensity. Runoff on the land dislodges and remove soil particles from the ground surface. The erosion can take different forms that combine in time and space: the erosion of the

*Corresponding author, Email: quantv@utt.edu.vn

diffuse slope or in parallel channels (Le Bissonnais et al., 2002).

Evaluation of loss of soil in an area is required to be assessed to know the severity of soil erosion for proper soil management. Soil loss can be determined either by theoretical estimation based on values of watershed parameters or actual direct measurements in the field. Field studies are expensive and time-consuming and need to be collected over many years. These studies also have limitations because of the complexity of parameter interactions and the difficulty of generalizing from the results. On the other hand, soil erosion models can simulate erosion processes in the watershed and may be able to take into account many of the complex interactions that affect rates of erosion.

Processes of soil loss depend on multiple interacting factors thus simulation of the model is a complex process. In this study, eight variable input factors based on the earlier work of (Nearing et al., 1999) and (Gholami et al., 2018) have been used in the model study to predict the soil loss due to erosion. These factors include rainfall, rainfall intensity (Kinnell, 1981; Wischmeier and Smith, 1958), land cover (Peng and Wang, 2012), topography (Martz and de Jong, 1987), and soil moisture (Epstein et al., 1966). On hill slopes, soil loss is depended on the slope rate (Fox and Bryan, 2000) and vegetation cover (Wang et al., 2002). (McDowell and Sharpley, 2002) investigated the effect of Antecedent Moisture Conditions (AMC) on soil loss. Moreover, the surrounding environment temperatures such as ambient temperature and soil temperature also affect the quantity of soil loss (Klik and Eitzinger, 2010).

In view of the complex nature of factors affecting soil erosion, nowadays Machine Learning (ML) approach is being adopted for predicting soil erosion in an area. ML methods have also been used for solving other complex civil engineering problems (Kuo et al., 2009; Pham et al., 2020; Samui, 2008).

In this study, a popular ML method namely Support Vector Machines (SVM) is proposed to predict soil loss based on eight input variables (ambient temperature Tair (°C), rainfall amount (mm), AMC (mm), rainfall intensity (mm/h), slope vegetation cover, slope (°), soil temperature Tsoil (°C) and moisture of soil (%) to simulate the model and to obtain output variable that is soil loss (g/m^2). Data of 41 locations of the Mazandaran Province, Northern Iran from published literature (Gholami et al., 2018) was utilized in the model study. Statistical validation indicators R, RMSE, and MAE were used to verify the model performance. MATLAB software was used for the model study.

2. Support Vector Machine Method

Support Vector Machine (SVM) was firstly introduced by Vapnik (Vapnik, 1999) to solve many real-world problems, including civil engineering. The concept of SVM is to plot the original input space into a high-dimensional feature space with using a hyper plane (Bui et al., 2016). Defining x = x_i as an input variable used in the simulation, and y is the output (forecasted variable). The SVM function is depicted by the equation (1):

$$y = f\left(x\right) = \mathrm{w}\theta\left(x\right) + b \qquad (1)$$

where w is the weight metric, b is the bias of the model and $\theta\left(x\right)$ is considered as the mapped nonlinearly from the input space x. In this study, using SVM was used to forecast the soil loss by taking advantages of ML algorithm, such as capacity of minimization of outliers and noise (Vapnik and Chapelle, 2000). The SVM method was coded in MATLAB, based on the ML toolbox by adopting the problem such as taking into account the random sampling effect and tuning the SVM parameters.

Standard statistical indicators Pearson Correlation Coefficient (R), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were used to evaluate performance of the SVM method. (Le et al., 2019, 2020; Ly et al., 2019c):

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(p_{0,j} - p_{t,j}\right)^2} \quad (2)$$

$$\mathrm{R} = \sqrt{\frac{\sum_{j=1}^{N}\left(p_{0,j} - \overline{p}_0\right)\left(p_{t,j} - \overline{p}_t\right)}{\sqrt{\sum_{j=1}^{N}\left(p_{0,j} - \overline{p}_0\right)^2 \sum_{j=1}^{N}\left(p_{\mathrm{t},j} - \overline{p}_t\right)^2}}} \quad (3)$$

$$\mathrm{MAE} = \frac{1}{\mathrm{N}}\sum_{\mathrm{j}=1}^{\mathrm{N}}\left(\mathrm{p}_{0,\mathrm{j}} - \mathrm{p}_{\mathrm{t,j}}\right) \quad (4)$$

Where: N is the number of data sets, p_0 and $\overline{p}_0$ is the actual experiment value and the average experimental value, p_t and $\overline{p}_t$ is the predicted value and the average predicted value. R measures the predicted and experimental value association, if the R is closer to 1, the model is more accurate. RMSE calculates the square root average difference between the expected values and the experimental values and the difference between the experimental and the predicted values is determined MAE criteria.

3. Database collection

In this study, available published data (Gholami et al., 2018) of 41 locations of the Mazandaran Province, Northern Iran was used in the SVM model application and analysis. Eight input variables used in the model study are: ambient temperature T_{air} (°C), rainfall (mm), AMC (mm), rainfall intensity (mm/h), slope (°), vegetation cover (%), soil temperature T_{soil} (°C), moisture of soil (%). Output parameter is considered to soil loss (g/m^2).

The climate of the study area is semi-humid with annual precipitation 600 mm (approx.) and ambient temperature 10°C. The experimental area is hill slopes (10° to 20°) with 60cm clay-loamy texture soil cover. Data from nine experimental plots were collected. After each rainfall event, runoff and eroded sediments were measured at the end of the downslope of each plot. The mean concentration of the sediment of each plot was estimated for evaluating the mean soil loss (g/m^2) for the hill slopes (Gholami et al., 2018). Rainfall amounts and other climatic variables such as ambient temperature, AMC, rainfall intensity were directly determined by an automatic rain gauge station located in the experimental area. Satellite and Google Earth images were used to develop land cover and vegetation cover maps. The initial statistical analysis of the dataset is summarized in Table 1.

In the present study, dataset of 41 locations was classified in standard 70/30 ratio. 70% of the data (29 samples) used to train the SVM model, whereas remaining 30% (2 samples) used to test or validate the model.

Table 1. Initial statistical analysis of the dataset

Variable	T_{air}	T_{soil}	Soil moisture	Rainfall	AMC	Rainfall intensity	Slope	Vegetation cover	Soil loss
Unit	C°	C°	%	mm	mm	mm/h	°	%	g/m^2
Role	Input	Input	Input	Input	Input	Input	Input	Input	Output
Count	41	41	41	41	41	41	41	41	41
Min	-0.6	3.5	21.0	0.0	0.0	0.0	10.0	10.0	0.0
Q_{25}	2.4	10.01	28.12	1.20	23.20	1.20	10.0	15.0	0.52
Q_{50}	8.4	14.4	30.6	10.5	39.2	4.4	20.0	15.0	1.87
Q_{75}	11.4	17.12	31.32	16.0	53.7	6.61	20.0	15.0	3.29
Max	18.2	18.7	37.0	46.3	109.8	13.5	20.0	70.0	5.18
Mean	7.28	13.43	29.66	13.26	37.55	4.63	15.12	18.78	2.02
STD[a]	5.43	4.47	3.23	13.66	24.21	3.71	5.06	13.45	1.64

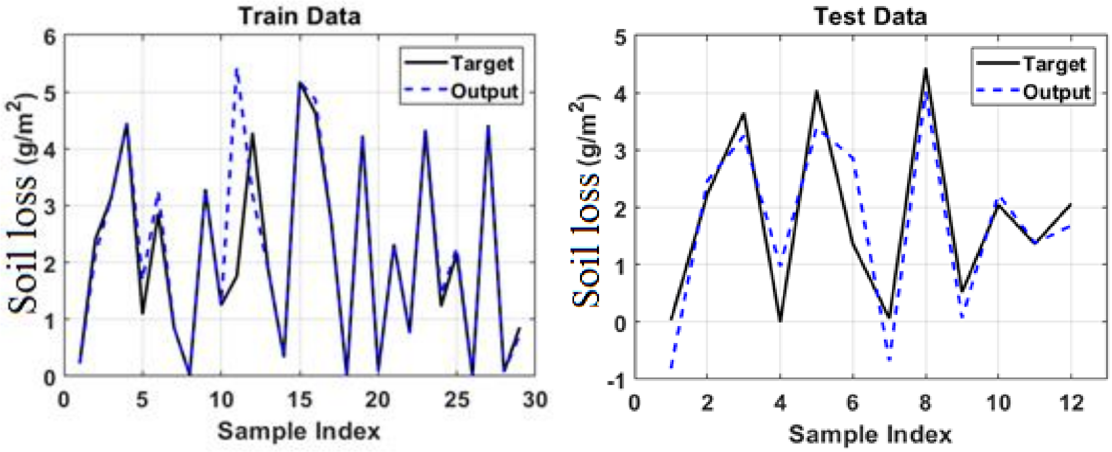
[a]Standard deviation

4. Results and discussions

4.1. Prediction performance of SVM

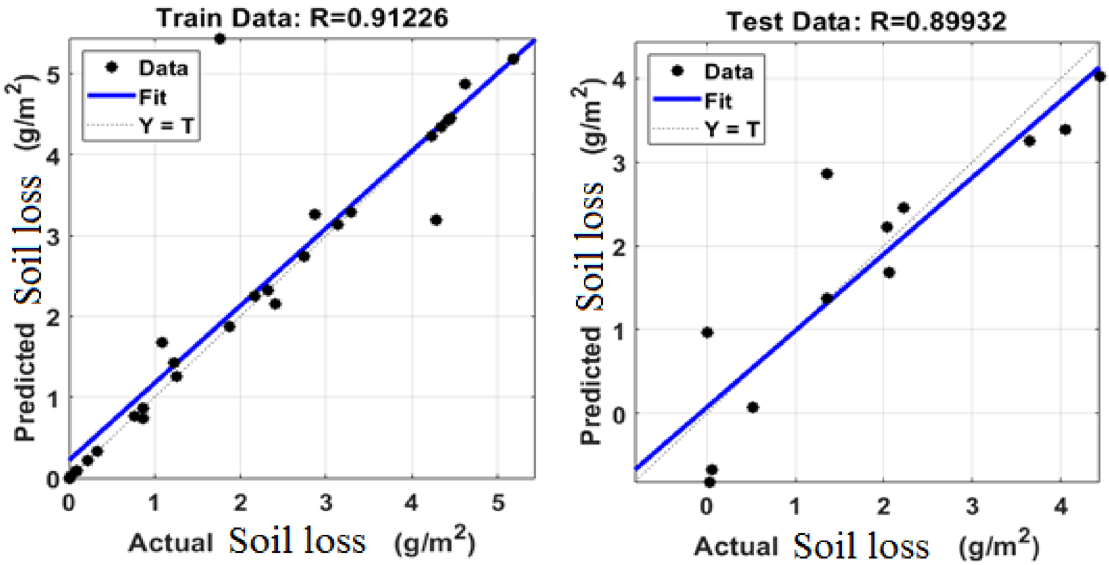
Using MATLAB software for training and testing of soil loss is shown in Figs. 1a, b respectively. Figure 1 shows that SVM model results of training and testing phase are almost identical for each sample, thus suggesting prediction ability of SVM model is very good.

The regression model for the training and testing parts is shown in Figs. 2a, b respectively. From Fig. 2, we can note that the soil loss prediction of the SVM model is quite close to experimental soil loss.



(a) (b)

Figure 1. Soil loss study by SVM model (a) Training; (b) Testing phase



(a) (b)

Figure 2. Regression analysis of soil loss using SVM model for (a) training; (b) testing phase

The correlation value "R" obtained for training is 0.9123 and for the testing, it is 0.8993. RMSE values are 0.7294 and 0.6863 for training and testing, respectively, whereas MAE values are 0.2301 and 0.5657 respectively for training and testing (Table 2).

These results suggest the high accuracy of the SVM model in the soil loss prediction with minimum error.

Table 2. Summary of the prediction performance of the best simulations for the training and testing datasets

Part	Values	RMSE	MAE	R	Error Std
Train dataset	Max	0.7294	0.2301	0.9123	0.7308
Test dataset	Max	0.6863	0.5657	0.8993	0.7117

4.2. Importance of input factors using Partial dependence plots (PDP)

In this study, PDP was estimated (Fig. 3) for 8 variable inputs which correspond to 8 input variables of SVM model (ambient temperature T_{air} (°C), rainfall (mm), Antecedent Moisture Conditions (AMC) (mm), rainfall intensity (mm/h), slope, vegetation cover, (%) soil temperature T_{soil} (°C), moisture of soil (%). Soil loss (g/m^2) was plotted against each input variables. For the ambient temperature T_{air}, the values varied from [0 to -0.25], for the soil temperature T_{soil} it varies from [0 to 0.125], for the soil temperature T_{soil} it varies from [0 to -0.7], for the rainfall it varies from [0 to 0.65]. The value varies from [0 to -0.15], [0 to 1.2], [0 to 0.2] and [0 to -0.8] for AMC, rainfall intensity, slope and vegetation cover, respectively. Based on the value of each range, the input effect on the soil loss is maximum for rainfall intensity followed by vegetation cover, moisture, rainfall, ambient temperature, slope, AMC and soil temperature. This order shows that the most important input effect is of the rainfall intensity followed by vegetation cover. This is according to similar studies by other workers (Bissonnais et al., 2004). Further, PDP investigation shows two kinds of input effect on the soil loss: (i) soil temperature, rainfall, rainfall intensity and slope have positive effect on the soil loss (ii) ambient temperature, moisture of soil, AMC and vegetation cover have negative effect on soil loss. These observations also similar to other researchers (Gholami et al., 2018).

5. Conclusions

In this study, a well-known machine learning algorithm Support Vector Machine (SVM) has been used in the development of the soil loss prediction model. Data on published literature was used in the model study. Regression analysis shows that the soil loss prediction of the SVM model is quite close to experimental soil loss. Partial Dependence Plots (PDP) was used to investigate the dependence of prediction results of eight input variables used in the model study. Rainfall intensity and vegetation cover were found to be the two most important affecting input parameters for the soil loss prediction.

The accuracy of the proposed SVM was assessed by using three statistical performance evaluation indicators namely Person correlation coefficient (R), Root Mean Squared Error (RMSE), Mean Squared Error (MAE). The correlation value “R” obtained for training is 0.9123 and for the testing, it is 0.8993. RMSE values are 0.7294 and 0.6863 for training and testing, respectively, whereas MAE values are 0.2301 and 0.5657 respectively for training and testing. These results suggest the high accuracy of the SVM model in the soil loss prediction with minimum error. Therefore, the SVM model can be used for the accurate prediction of soil loss due to erosion not only in the study area but also in other areas considering local variable geo-environmental parameters.

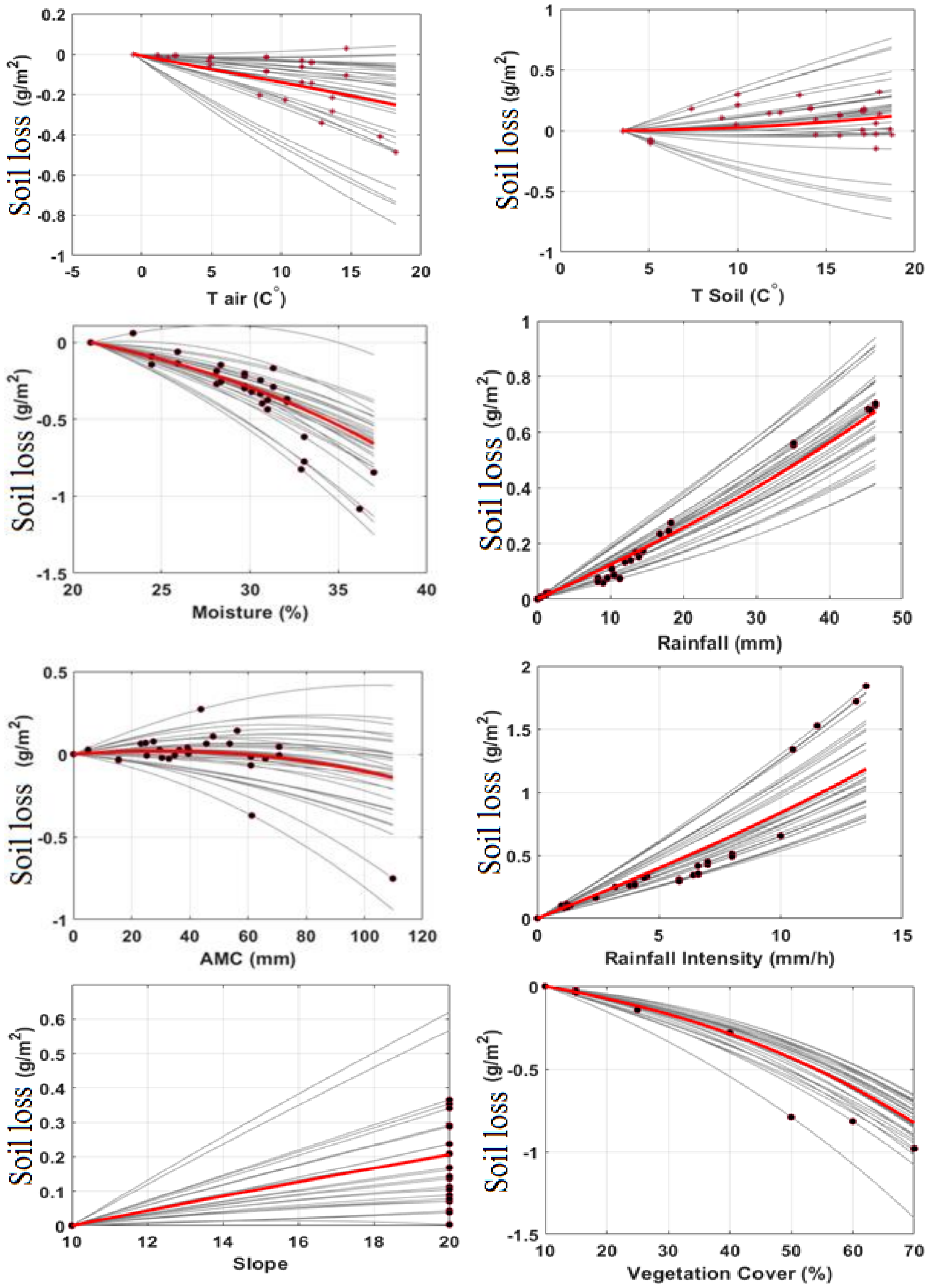


Figure 3. Partial Dependence Plots (PDP) of the input variables used in this study

References

https://doi.org/10.1017/S0021859610000158.

Bissonnais Y.L., Dubreuil N., Daroussin J., 2004. Modélisation et cartographie de l'aléa d'érosion des sols à l'échelle régionale. Étude et Gestion des Sols, 16.

Borrelli P., Van Oost K., Meusburger K., Alewell C., Lugato E., Panagos P., 2018. A step towards a holistic assessment of soil degradation in Europe: Coupling on-site erosion with sediment transfer and carbon fluxes. Environ Res., 161, 291–298. https://doi.org/10.1016/j.envres.2017.11.009.

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

Epstein E., Grant W.J., Struchtemeyer R.A., 1966. Effects of Stones on Runoff, Erosion, and Soil Moisture. Soil Science Society of America Journal, 30, 638–640.
https://doi.org/10.2136/sssaj1966.03615995003000050029x.

Fox D.M., Bryan R.B., 2000. The relationship of soil loss by interrill erosion to slope gradient. Catena, 38, 211–222.
https://doi.org/10.1016/S0341-8162(99)00072-7.

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.
https://doi.org/10.1016/j.catena.2017.12.027.

Kinnell P.I.A., 1981. Rainfall Intensity-Kinetic Energy Relationships for Soil Loss Prediction. Soil Science Society of America Journal, 45, 153–155. https://doi.org/10.2136/sssaj1981.03615995004500010033x.

Kiran S., Lal B., Tripathy S.S., 2016. Shear Strength Prediction of Soil based on Probabilistic Neural Network. Indian Journal of Science and Technology, 9, https://doi.org/10.17485/ijst/2016/v9i41/99188.

Klik A., Eitzinger J., 2010. Impact of climate change on soil erosion and the efficiency of soil conservation practices in Austria. The Journal of Agricultural Science, 148, 529–541.

Kuo Y.L., Jaksa M.B., Lyamin A.V., Kaggwa W.S., 2009. ANN-based model for predicting the bearing capacity of strip footing on multi-layered cohesive soil. Computers and Geotechnics, 36, 503–516. https://doi.org/10.1016/j.compgeo.2008.07.002.

Le Bissonnais Y., Montier C., Jamagne M., Daroussin J., King D., 2002. Mapping erosion risk for cultivated soil in France. Catena, 46, 207–220. https://doi.org/10.1016/S0341-8162(01)00167-9.

Le L.M., Ly H.-B., Pham B.T., Le V.M., Pham T.A., Nguyen D.-H., Tran X.-T., Le T.-T., 2019. Hybrid Artificial Intelligence Approaches for Predicting Buckling Damage of Steel Columns Under Axial Compression. Materials, 12, 1670. https://doi.org/10.3390/ma12101670.

Le T.-T., Pham B.T., Ly H.-B., Shirzadi A., Le L.M., 2020. Development of 48-hour Precipitation Forecasting Model using Nonlinear Autoregressive Neural Network, in: Ha-Minh, C., Dao, D.V., Benboudjema, F., Derrible, S., Huynh, D.V.K., Tang, A.M. (Eds.), CIGOS 2019, Innovation for Sustainable Infrastructure, Lecture Notes in Civil Engineering. Springer Singapore, 1191–1196.

Ly H.-B., Le L.M., Duong H.T., Nguyen T.C., Pham T.A., Le T.-T., Le V.M., Nguyen-Ngoc L., Pham B.T., 2019a. Hybrid Artificial Intelligence Approaches for Predicting Critical Buckling Load of Structural Members under Compression Considering the Influence of Initial Geometric Imperfections. Applied Sciences, 9, 2258. https://doi.org/10.3390/app9112258.

Ly H.-B., Le L.M., Phi L.V., Phan V.-H., Tran V.Q., Pham B.T., Le T.-T., Derrible S., 2019b. Development of an AI Model to Measure Traffic Air Pollution from Multisensor and Weather Data. Sensors, 19, 4941.
https://doi.org/10.3390/s19224941.

Martz L.W., de Jong E., 1987. Using cesium-137 to assess the variability of net soil erosion and its association with topography in a Canadian Prairie landscape. CatenA, 14, 439–451. https://doi.org/10.1016/0341-8162(87)90014-2.

McDowell R.W., Sharpley A.N., 2002. The effect of antecedent moisture conditions on sediment and

phosphorus loss during overland flow: Mahantango Creek catchment, Pennsylvania, USA. Hydrological Processes, 16, 3037–3050.
https://doi.org/10.1002/hyp.1087.

Nearing M.A., Govers G., Norton L.D., 1999. Variability in Soil Erosion Data from Replicated Plots. Soil Science Society of America Journal, 63, 1829–1835.
https://doi.org/10.2136/sssaj1999.6361829x.

PengT., Wang S., 2012. Effects of land use, land cover and rainfall regimes on the surface runoff and soil loss on karst slopes in southwest China. Catena, 90, 53–62. https://doi.org/10.1016/j.catena.2011.11.001.

Pham B.T., Le L.M., Le T.-T., Bui K.-T.T., Le V.M., Ly H.-B., Prakash I., 2020. Development of advanced artificial intelligence models for daily rainfall prediction. Atmospheric Research, 237, 104845. https://doi.org/10.1016/j.atmosres.2020.104845.

Qi C., Ly H.-B., Chen Q., Le T.-T., Le V.M., Pham B.T., 2019. Flocculation-dewatering prediction of fine mineral tailings using a hybrid machine learning approach. Chemosphere, 125450. https://doi.org/10.1016/j.chemosphere.2019.125450.

Samui P., 2008. Prediction of friction capacity of driven piles in clay using the support vector machine. Canadian Geotechnical Journal, 45, 288–295.

Thanh T.T.M., Ly H.-B., Pham B.T., 2020. A Possibility of AI Application on Mode-choice Prediction of Transport Users in Hanoi, in: CIGOS 2019, Innovation for Sustainable Infrastructure, Lecture Notes in Civil Engineering. Springer, Singapore, 1179–1184.
https://doi.org/10.1007/978-981-15-0802-8_189.

Vapnik V., 1999. The Nature of Statistical Learning Theory, 2nd edition. ed. Springer, New York.

Vapnik V., Chapelle O., 2000. Bounds on error expectation for support vector machines. Neural Computation, 12, 2013–2036.

Wang G., Wente S., Gertner G.Z., Anderson A., 2002. Improvement in mapping vegetation cover factor for the universal soil loss equation by geostatistical methods with Landsat Thematic Mapper images. International Journal of Remote Sensing, 23, 3649–3667. https://doi.org/10.1080/01431160110114538.

Wischmeier W.H., Smith D.D., 1958. Rainfall energy and its relationship to soil loss. Eos, Transactions American Geophysical Union, 39, 285–291. https://doi.org/10.1029/TR039i002p00285.

Wu Y., Ouyang W., Hao Z., Lin C., Liu H., Wang Y., 2018. Assessment of soil erosion characteristics in response to temperature and precipitation in a freeze-thaw watershed. Geoderma, 328, 56–65. https://doi.org/10.1016/j.geoderma.2018.05.007.