JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Comparative Analysis of Slope Stability Prediction using Logistic Regression and Random Forest Approach

[1]Saurabh Kumar Anuragi, [2]D.Kishan, [3]S.K.Saritha

[1]Research Scholar, [2]Associate Professor, [3]Assistant Professor
[1,2]Department of Civil Engineering, [3]Department of Computer Science and Engineering
[1,2,3]Maulana Azad National Institute of Technology, Bhopal (M.P.), India, 462003

Abstract : The study presents a comparative analysis of three prediction models, Logistic Regression (LR), Random Forest (RF), and Multi-Layer Perceptron Neural Networks (MLPNN). It employs a trial and error method for optimizing hyperparameters in slope stability prediction. The study utilizes a dataset comprising 108 slope cases with various influencing factors, including unit weight (Y), cohesion (c), angle of internal friction (ϕ), slope angle (β), height (H), pore water pressure coefficient (ru), and factor of safety (FS) as inputs, while slope status (S) serves as the output variable. Based on the confusion matrix and ROC curves, the RF model demonstrated superior performance over the other models. The utilization of RF enhances the capacity and efficiency of slope deformation prediction models, establishing it as the most accurate tool for forecasting slope stability.

IndexTerms **- Machine Learning, Slope Stability, Limit Equilibrium, Logistic Regression, Neural Networks, Artificial Intelligence.**

I. Introduction

Slopes represent one of the predominant geological features in both natural landscapes and human-engineered environments. However, the inherent instability of slopes, often triggered by a combination of natural and man-made factors, including collapses, landslides, and mudslides, has resulted in substantial loss of life, property, and economic resources. Accurately assessing slope stability is pivotal in the success of slope engineering endeavors. Given the complex and nonlinear nature of slope dynamics, which involve myriad random and intricate elements, the relationships between slope stability and its influencing factors are highly nonlinear [1].

Conventional methods for analyzing slope stability, such as the limit equilibrium method and numerical analysis techniques like the numerical manifold method (NMM), discontinuous deformation analysis (DDA), and phase-field model (PFM), have encountered significant limitations. These methodologies often grapple with "unclear mechanisms" and "inaccurate models," impeding their ability to accurately forecast the stability of complex slopes [2]–[5]. Moreover, the computational demands associated with these traditional methods pose challenges, especially in scenarios where rapid stability assessments are imperative, such as in large-scale engineering projects like hydroelectric endeavors.

The advancements in science and technology have facilitated the application of machine learning techniques in slope stability analyses. These approaches involve scrutinizing available slope data to understand and predict the interplay between slope stability and its influencing factors. Yaser et.al (2023) [6] carried out comparative analysis utilizing a range of machine learning techniques, including multilayer perceptron (MLP), decision tree (DT), support vector machines (SVM), and random forest (RF) algorithms. The objective revolves around predicting the Factor of Safety (F.S) for earth slopes, aiming to enhance stability analysis and facilitate the adoption of suitable stabilization measures. To achieve this, a dataset comprising 100 earth slopes sourced from diverse locations across Iran is employed, subsequently partitioned into distinct training and testing sets. The overarching aim is to furnish precise and dependable calculations of the Factor of Safety, thereby fortifying the accuracy of stability assessments and enabling the effective implementation of appropriate measures to bolster earth slope stability.
Bui et.al (2019) [7] used machine learning to predict slope failure safety factors, comparing regression methods like multi-layer perceptron (MLP), Gaussian process regression (GPR), multiple linear regression (MLR), simple linear regression (SLR), and support vector regression (SVR). It replaces traditional slope analysis with more advanced machine learning-based design tools. The aim is to optimize these models using a dataset of 630 analyses split into training (504) and testing (126) sets. Statistical indices reveal MLP as the top performer among the machine learning models. Mahmoodzadeh et.al (2022) [8] performed a comparison between six machine learning techniques, Gaussian process regression (GPR), support vector regression, decision trees, long-short term memory, deep neural networks, and K-nearest neighbors were performed. The study concluded most of the features in this study have significant contributions to slope stability.

Choobbasti et.al (2009) [9] utilized ANN systems consisting of multilayer perceptron networks are developed to predict slope stability in a specified location, based on the available site investigation data from Noabad, Mazandaran, Iran. The results were compared with limit equilibrium and found to be better perfroming model for slope stability prediciton. Das et.al (2011) [10] developed different ANN models to classify slope and to predict foctor of safety. On comparing the model with support vector machine and genetic programming, ANN outperformed.

In recent years, the integration of intelligent computational techniques has significantly impacted various aspects of geotechnical applications. These applications span a wide spectrum, including critical areas such as landslide susceptibility assessment [11]–[14], settlement prediction for shallow foundations [15], [16], estimation of bearing capacity for pile foundations [17]–[19], determination of physical properties of soil [20], assessment of soil quality [21], and evaluation of liquefaction potential [22]. Moreover, as the accessibility and availability of slope-related parameters have increased, supervised learning methods have emerged as a robust approach for predicting slope stability. The utilization of these methods has yielded noteworthy and promising results in the realm of slope stability prediction. This trend marks a significant advancement in geotechnical engineering, offering enhanced tools and methodologies to foresee and manage slope stability concerns through the application of computational intelligence.

.

II. DATA VISUALIZATION METHODOLOGY

The study uses the dataset from [23], that consists the magnitude of unit weight (Y), cohesion (c), angle of internal friction (ϕ), slope angle (β), height (H), pore water pressure coefficient (ru),factor of safety (FS) as input and status of slope (S) as output i.e. stable (1) or unstable (0). The FS serves as a comprehensive tool for assessing the stability of a slope. Before commencing the analysis, the dataset was normalized using Eq.(1) to enhance the performance of the model. Normalizing the data to a standardized range is instrumental in enhancing the model's capacity to generalize and make precise predictions when presented with new, unseen data. This normalization procedure plays a pivotal role in ensuring a more robust and reliable model performance.

$$y_{normalization} = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (1)$$

where, y is a normalized input parameter, x is the original input parameter, and x_{max} and x_{min} are the maximum and minimum parameters, respectively.

The basic descriptive statistical analysis for the dataset is shown in Table 1. The correlation between input variables and the slope's status is vividly depicted through the correlation Table 2 and pairplots found in Figure 2. These visual representations offer comprehensive insights into the correlation coefficients and the marginal frequency distributions of each parameter. They effectively showcase the intricate pairwise relationships existing among these parameters. The diagonal histograms within the pairplots vividly illustrate the distribution of numerical values associated with each feature. Table 2 specifically highlights intriguing correlations, notably demonstrating a positive relationship between the Status feature and cohesion, while showcasing a negative correlation with the slope's height. These revelations hint at the varying impacts different variables exert on the Status feature, contributing to a deeper understanding of the dynamics at play.

Furthermore, the dataset's characteristics are visually presented through Violin plots, showcased in Figure 1. These plots serve to explain the distribution and density of data across distinct categories or groups. Notably, the width of the violin graphically represents data density at different points, with thicker sections indicating higher density and thinner parts suggesting regions of lower density. The horizontal line within the violin signifies the median value of the dataset, offering a pivotal reference point to gauge the central tendency of the data.

Additionally, the methodology's workflow and process are briefly outlined in the form of a flow chart, depicted in Figure 3. This visual aid offers a structured representation of the steps involved in the study's approach, providing a clear roadmap of the methodology's progression and stages. These visual aids collectively enrich the understanding of the dataset's characteristics, interrelationships between variables, and the procedural pathway followed in the study's methodology.

Table 1 : Description of Dataset

index	Y (kN/m^3)	C (kPa)	ϕ (^{0})	β (^{0})	H (m)	ru	FS
Standard deviation	3.68	22.40	11.25	9.67	47.37	0.17	0.40
min	12.00	0.00	0.00	16.00	3.60	0.00	0.63
mean	20.01	14.58	26.25	33.44	44.28	0.19	1.29
max	28.44	150.05	45.00	53.00	214.00	0.50	2.31
Average	19.96	10.48	26.63	33.26	40.73	0.21	1.28
Variance	12.53	116.73	116.57	90.23	1845.94	0.03	0.16
count	108.00	108.00	108.00	108.00	108.00	108.00	108.00

Table 2 : Correlation Matrix of dataset

index	ϒ	C	ϕ	β	H	ru	FS	Status
ϒ	1	0.327	0.378	0.186	0.472	-0.030	0.252	0.326
C		1.000	0.238	0.311	0.397	-0.233	0.136	0.234
ϕ			1.000	0.564	0.244	0.048	0.304	0.429
β				1.000	0.188	-0.120	-0.202	0.101
H					1.000	-0.207	-0.180	-0.042
ru						1.000	-0.221	-0.121
FS							1.000	0.780
Status								1.000

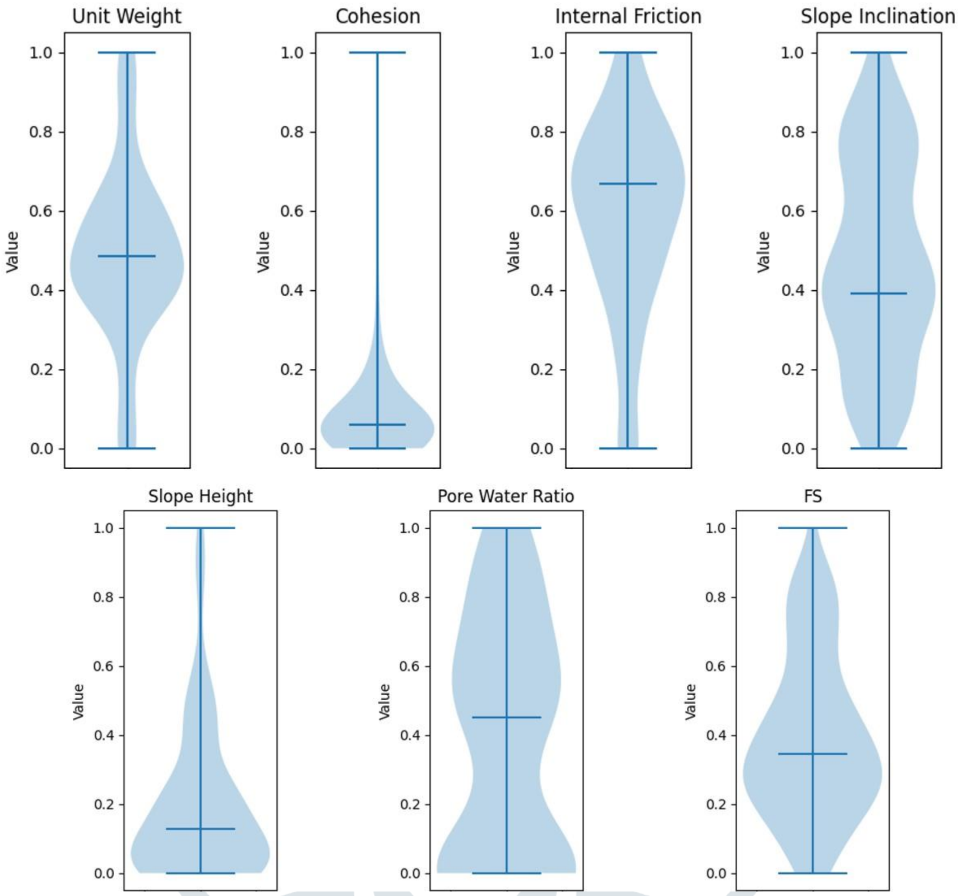

Figure 1 : Violin Plots showing distribution of slope cases

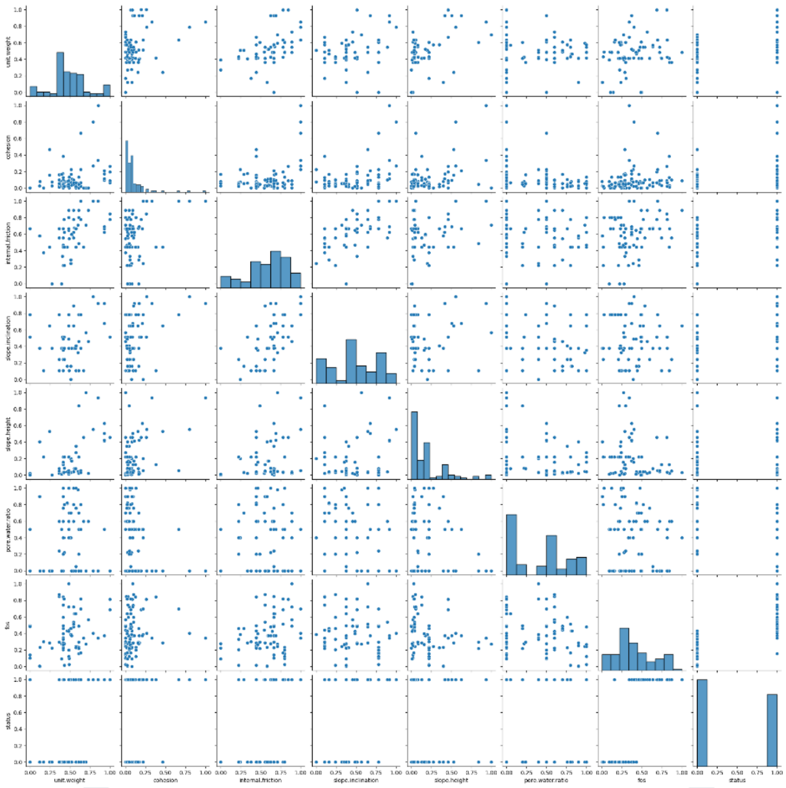

Figure 2 **:** Pairplots of variables

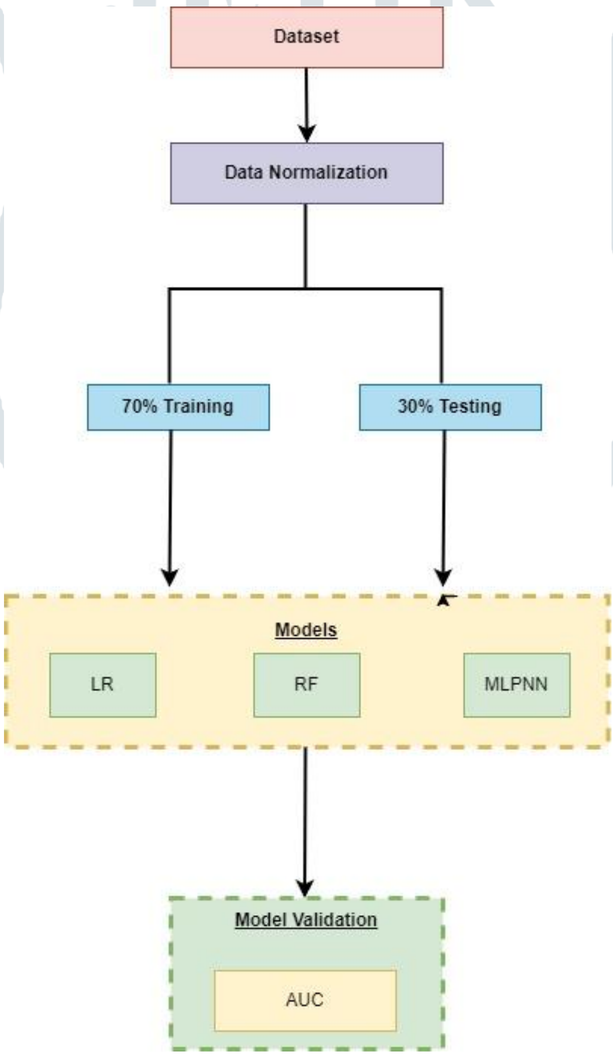

Figure 3 **:** Methodology of the Study

III. Results and Discussion

The study implements the use of Logistic Regression (LR), Random Forest (RF) and Multi-Layer Perceptron Neural Networks (MLPNN) for the prediction of slope stability. The dataset consists of 108 data sample that consists the magnitude of unit weight

(ϒ), cohesion (c), angle of internal friction (ϕ), slope angle (β), height (H), pore water pressure coefficient (ru), factor of safety (FS) as input and status of slope (S) as output i.e. stable (1) or unstable (0). The dataset is divided into 70:30 (70% training and 30% testing). The hyperparameters taken in the analysis are shown in Table 3. The performance of the model is evaluated using AUC and ROC curve. The optimal hypermeters of different algorithms for best AUC score were obtained by trial and error, is shown in Table 4. The Confusion Matrix of each model is shown in Figure 4. The confusion matrix evaluates the performance of classification model by providing summary of predicted and classified output by the model. Various metrics such as accuracy, precision, recall, specificity and f1 score are used to evaluate the model performance and understand its strengths and weaknesses in classifying instances. The confusion matrix provides a detailed breakdown of the model's predictions, making it a useful tool in machine learning evaluation and model selection. The optimal hyper-parameters for each algorithm obtained are displayed in Table 4.

Table 3 **:** Hyperparameters taken for analysis

Algorithm	Hyperparameters
LR	C = [50,100,150,200,250,300], max_iter = 500
RF	n_estimators = [100,200,300,400,500,600], criterion = 'entropy'
MLPNN	max_iter = [100,200,300,400,500,600], learning_rate_init=0.001

Table 4 **:** Optimal Hyperparameters for best AUC score

Algorithm	Optimal Hyperparameters	AUC
LR	C=50	0.879
RF	n_estimators = 200	1
MLPNN	max_iter = 400	0.969

Based on Figure 4, it can be observed that LR has the highest number of misclassifications, with a maximum of 2. Following MLPNN have 1 misclassification while RF has no misclassifications out of the total 33 instances in the testing dataset. Thus, lesser the misclassification more the accuracy of model. Therefore, from Table 4 , the model with highest value of AUC is RF = 1 and lowest value of AUC of LR = 0.879. Among the three models, RF outperformed other models but the performance of LR and MLPNN is slightly similar however comparatively lower than RF. The performance comparison of different algorithms used in this study is illustrated in Fig (5) through ROC curves. The results presented in Fig (5) indicate that Random Forest (RF) exhibited exceptional performance with a perfect Area Under the Curve (AUC) score of 1, indicating its remarkable ability to discriminate between classes. Logistic Regression (LR) demonstrated strong predictive power with an AUC of 0.879, emphasizing its effectiveness in capturing essential patterns within the data. Furthermore, both MLPNN displayed impressive AUC scores of 0.969, highlighting their robustness and reliability in handling complex relationships within the dataset.

Models	Actual Values	Predicted Values: 0	Predicted Values: 1
LR	0	15	2
LR	1	2	14
RF	0	17	0
RF	1	0	16
MLPNN	0	17	0
MLPNN	1	1	15

Figure 4 **:** Confusion Matrix of models for best AUC score

These results highlight the significance of choosing machine learning algorithms that align with the unique characteristics of the problem at hand. The exceptional performance of Random Forest (RF) emphasizes its suitability for tasks where precision and accuracy take precedence. Despite a slight lag, MLPNN remains a valuable option for scenarios where achieving a balance between interpretability and predictive strength is key. Furthermore, the similar performances of RF and MLPNN demonstrate the adaptability of these algorithms, positioning them as strong contenders for a wide array of predictive modeling endeavors.

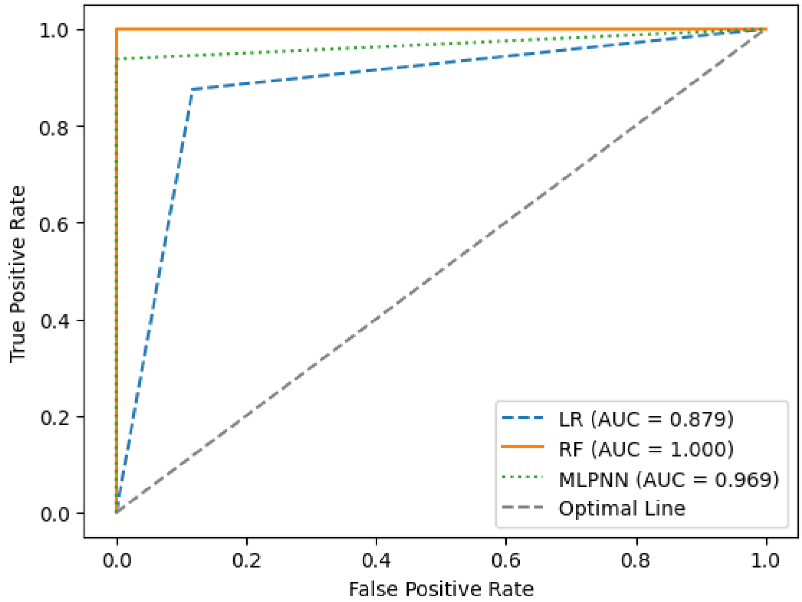

Figure 5 **:** ROC Curve of Models

IV. Conclusion

In conclusion, this research explored slope stability prediction using LR, RF, and MLPNN algorithms, constructing models based on geological conditions and failure mechanisms. The study utilized a dataset of 108 slope cases with key influencing factors. Hyperparameters were fine-tuned through trial and error, and evaluation metrics like the confusion matrix, AUC score, and ROC curve were employed.

Among the models, RF demonstrated superior performance in predicting slope stability compared to others, as evidenced by the confusion matrix and ROC curves. This suggests RF's potential as a reliable tool for such predictions. The research not only offers comparative insights into predictive models but also emphasizes the importance of robust evaluation methods. These findings can pave the way for further studies and practical applications in slope stability analysis, aiding in informed decision-making and risk reduction for unstable slopes.

References

[1] O. Nawari, R. Hartmann, and R. Lackner, “Stability analysis of rock slopes with the direct sliding blocks method,” Int. J. Rock Mech. Min. Sci., vol. 34, no. 3, p. 220.e1-220.e8, Apr. 1997, doi: 10.1016/S1365-1609(97)00206-2.
[2] M. R. Arvin, A. Zakeri, and M. Bahmani Shoorijeh, “Using Finite Element Strength Reduction Method for Stability Analysis of Geocell-Reinforced Slopes,” Geotech. Geol. Eng., vol. 37, no. 3, pp. 1453–1467, Jun. 2019, doi: 10.1007/s10706-018-0699-0.
[3] R. P. Bhandary, A. Krishnamoorthy, and A. U. Rao, “Stability Analysis of Slopes Using Finite Element Method and Genetic Algorithm,” Geotech. Geol. Eng., vol. 37, no. 3, pp. 1877–1889, Jun. 2019, doi: 10.1007/s10706-018-0730-5.
[4] J. M. Duncan and S. G. Wright, “The accuracy of equilibrium methods of slope stability analysis,” Eng. Geol., vol. 16, no. 1, pp. 5–17, Jul. 1980, doi: 10.1016/0013-7952(80)90003-4.
[5] “State of the Art: Limit Equilibrium and Finite-Element Analysis of Slopes | Journal of Geotechnical Engineering | Vol 122, No 7.” Accessed: May 01, 2023. [Online]. Available: https://ascelibrary.org/doi/abs/10.1061/(ASCE)0733-9410(1996)122:7(577)
[6] Y. A. Nanehkaran et al., “Comparative Analysis for Slope Stability by Using Machine Learning Methods,” Appl. Sci., vol. 13, no. 3, Art. no. 3, Jan. 2023, doi: 10.3390/app13031555.
[7] D. Tien Bui, H. Moayedi, M. Gör, A. Jaafari, and L. K. Foong, “Predicting Slope Stability Failure through Machine Learning Paradigms,” ISPRS Int. J. Geo-Inf., vol. 8, no. 9, Art. no. 9, Sep. 2019, doi: 10.3390/ijgi8090395.
[8] A. Mahmoodzadeh, M. Mohammadi, H. Farid Hama Ali, H. Hashim Ibrahim, S. Nariman Abdulhamid, and H. R. Nejati, “Prediction of safety factors for slope stability: comparison of machine learning techniques,” Nat. Hazards, vol. 111, no. 2, pp. 1771–1799, Mar. 2022, doi: 10.1007/s11069-021-05115-8.
[9] A. J. Choobbasti, F. Farrokhzad, and A. Barari, “Prediction of slope stability using artificial neural network (case study: Noabad, Mazandaran, Iran),” Arab. J. Geosci., vol. 2, no. 4, pp. 311–319, Nov. 2009, doi: 10.1007/s12517-009-0035-3.
[10] S. K. Das, R. K. Biswal, N. Sivakugan, and B. Das, “Classification of slopes and prediction of factor of safety using differential evolution neural networks,” Environ. Earth Sci., vol. 64, no. 1, pp. 201–210, Sep. 2011, doi: 10.1007/s12665-010-0839-1.
[11] B. Feizizadeh, M. S. Roodposhti, T. Blaschke, and J. Aryal, “Comparing GIS-based support vector machine kernel functions for landslide susceptibility mapping,” Arab. J. Geosci., vol. 10, no. 5, p. 122, Mar. 2017, doi: 10.1007/s12517-017-2918-z.

[12] "Full article: Comparison of four kernel functions used in support vector machines for landslide susceptibility mapping: a case study at Suichuan area (China)." Accessed: May 01, 2023. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/19475705.2016.1250112

[13] "A Comparative Study of Landslide Susceptibility Mapping Using SVM and PSO-SVM Models Based on Grid and Slope Units." Accessed: May 01, 2023. [Online]. Available: https://www.hindawi.com/journals/mpe/2021/8854606/

[14] C. Ballabio and S. Sterlacchini, "Support Vector Machines for Landslide Susceptibility Mapping: The Staffora River Basin Case Study, Italy," Math. Geosci., vol. 44, no. 1, pp. 47–70, Jan. 2012, doi: 10.1007/s11004-011-9379-9.

[15] M. A. Shahin, H. R. Maier, and M. B. Jaksa, "Predicting Settlement of Shallow Foundations using Neural Networks," J. Geotech. Geoenvironmental Eng., vol. 128, no. 9, pp. 785–793, Sep. 2002, doi: 10.1061/(ASCE)1090-0241(2002)128:9(785).

[16] "Settlement prediction of shallow foundations on granular soils using B-spline neurofuzzy models - ScienceDirect." Accessed: May 01, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0266352X03000946

[17] "Prediction of pile bearing capacity using support vector machine: International Journal of Geotechnical Engineering: Vol 5, No 1." Accessed: May 01, 2023. [Online]. Available: https://www.tandfonline.com/doi/abs/10.3328/IJGE.2011.05.01.95-102

[18] I.-M. Lee and J.-H. Lee, "Prediction of pile bearing capacity using artificial neural networks," Comput. Geotech., vol. 18, no. 3, pp. 189–200, Jan. 1996, doi: 10.1016/0266-352X(95)00027-8.

[19] "Modeling Pile Capacity Using Support Vector Machines and Generalized Regression Neural Network | Journal of Geotechnical and Geoenvironmental Engineering | Vol 134, No 7." Accessed: May 01, 2023. [Online]. Available: https://ascelibrary.org/doi/full/10.1061/%28ASCE%291090-0241%282008%29134%3A7%281021%29

[20] A. Besalatpour, M. Hajabbasi, S. Ayoubi, A. Gharipour, and A. Jazi, "Prediction of soil physical properties by optimized support vector machines," Int. Agrophysics, vol. 26, no. 2, pp. 109–115, Apr. 2012, doi: 10.2478/v10247-012-0017-7.

[21] "A comprehensive support vector machine-based classification model for soil quality assessment - ScienceDirect." Accessed: May 01, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167198715001543

[22] "'APPLICATION OF A SUPPORT VECTOR MACHINE FOR LIQUEFACTION ASSESSMENT' by Ching-Yinn Lee and Shuh-Gi Chern." Accessed: May 01, 2023. [Online]. Available: https://jmstt.ntou.edu.tw/journal/vol21/iss3/10/

[23] A. Manouchehrian, J. Gholamnejad, and M. Sharifzadeh, "Development of a model for analysis of slope stability for circular mode failure using genetic algorithm," Environ. Earth Sci., vol. 71, no. 3, pp. 1267–1277, Feb. 2014, doi: 10.1007/s12665-013-2531-8.