

Machine Learning (ML) Prediction of Soil Erosion Rate Index Enhanced by Domain-Guided Feature Engineering

Ahmed M. Khalil, Mousa F. Attom, Rami A. Hawileh and Mohammad M. Yamin

Abstract— Accurate prediction of the Erosion Rate Index (I_{HET}) is essential for evaluating internal erosion susceptibility in earth structures such as embankment dams and engineered soil systems. Internal erosion is one of the main causes of failure in such structures, and its consequences can be severe. Experimental determination of I_{HET} is expensive and time-consuming, motivating the development of reliable data-driven models based on readily measurable soil properties. This study presents an interpretable and data-efficient linear modelling framework for predicting I_{HET} using a set of physical and mechanical soil parameters, namely Liquid Limit (LL), Plasticity Index (PI), Initial Water Content (w_i), Dry Density (γ_d), Internal Friction Angle (ϕ), and Cohesion (C). Two predictive scenarios are investigated. In the baseline model, I_{HET} is estimated directly from the six original parameters, achieving a test Coefficient of Determination (R^2) of 0.71 and a Root Mean Square Error (RMSE) of 0.39. To enhance predictive performance, a feature-engineered model is developed by expanding the original feature space to nineteen parameters through physically meaningful ratios, logarithmic transformations, and interaction terms. This structured feature expansion results in a substantial improvement in prediction accuracy, yielding a test R^2 of 0.82 and reducing RMSE to 0.31. For comparison, a Machine Learning (ML) model was also applied to the six-parameter scenario; however, they showed lower prediction capability, with Gradient Boosting (GB) performing best among them (test $R^2 = 0.79$, RMSE = 0.33). This confirms that the feature-engineered linear approach consistently outperformed the tested learning-based models. Overall, the results show that systematic feature engineering significantly improves I_{HET} prediction while preserving model simplicity, providing an efficient and interpretable alternative to more complex learning-based models.

Keywords— Machine Learning, Soil Erosion Rate Index, Feature Engineering, Interpretable Linear Modeling, Feature-Space Optimization.

Manuscript received Jan. 27, 2025. This work was supported by the American University of Sharjah (AUS), Civil Engineering Department under Grant PDFA25.

A. M. Khalil is Postdoctoral Fellow in American University of Sharjah (AUS), Sharjah, 26666 UAE

M. F. Attom is Professor in American University of Sharjah (AUS), Sharjah, 26666 UAE

R. A. Hawileh is Professor in American University of Sharjah (AUS), Sharjah, 26666 UAE

M. M. Yamin is Associate Professor in Minnesota State University (MSU), Mankato, 56001 USA

I. INTRODUCTION

Internal erosion is one of the leading causes of embankment dam failure [1], [2]. It occurs when fine soil particles are progressively detached and transported by seepage flow through defects or cracks within the dam body, potentially resulting in sinkholes, increased seepage, muddy discharge, and loss of structural integrity [2]. The susceptibility of soils to internal erosion depends on several physical and mechanical properties, including fines content, plasticity, water content, dry density, shear strength, and mineralogical composition [2], [3]. The Erosion Rate Index (I_{HET}) is widely used to quantify soil resistance to erosion under hydraulic shear stress and typically ranges from 0 to 6, indicating that erosion rates may vary by up to six orders of magnitude among different soils [3], [4]. Experimental methods such as the Slot Erosion Test and the Hole Erosion Test are commonly employed to determine I_{HET} and critical shear stress, with the latter favored for its simplicity despite known flow non-uniformity limitations [3]-[5]. Reliable prediction of I_{HET} is therefore essential for erosion risk assessment and the development of mitigation measures in embankment dams [2],[4],[6].

The I_{HET} is commonly determined through laboratory-based experimental methods, most notably the Hole Erosion Test (HET), the Jet Erosion Test (JET), and flume-based erosion tests [1], [7], [8]. While these techniques provide valuable insight into soil erosion behavior, they are generally associated with high costs, lengthy testing procedures, and limited scalability. Tests such as the HET and laboratory flume experiments require specialized equipment, careful sample preparation, and continuous monitoring, making large experimental campaigns difficult to implement [1], [7], [8]. Similarly, JET involve complex setups and data interpretation procedures that further increase time and resource demands [7]. At the same time, laboratory experimentation remains essential for defining fundamental physical, mechanical, and dynamic soil properties, despite the cost and effort associated with cyclic triaxial and resonant column testing [9], [10]. To address these challenges, data-driven approaches such as Artificial Neural Network (ANN) models have been introduced to estimate I_{HET} based on basic soil properties, offering faster predictions and improved scalability once trained [1]. However, these models typically rely on extensive experimental datasets for calibration and often function as black-box predictors with limited physical interpretability. As a result, there remains a need for

predictive approaches that can reduce experimental effort while maintaining both accuracy and engineering insight [1], [7].

In recent years, Machine Learning (ML) techniques have been increasingly applied to model soil erosion behavior and erosion-related indices, demonstrating strong predictive capability across different spatial scales and soil conditions [11]-[25]. ANNs, including generalized feed-forward networks and adaptive neuro-fuzzy inference systems, have been successfully used to estimate erosion rates and soil erodibility, often achieving high accuracy compared to traditional statistical approaches [11]-[13]. In parallel, tree-based ensemble methods such as Random Forest (RS), Gradient Boosting (GB), and Extreme Gradient Boosting (EGB) have shown strong performance in erosion susceptibility mapping and erosion rate prediction, benefiting from their ability to capture nonlinear relationships and variable interactions [14]-[20]. Other Learning-based models, including Support Vector Machines (SVM), Convolutional Neural Networks (CNNs), and relevance vector machines, have also been explored, particularly for large-scale erosion assessment using topographic, vegetation, and remote sensing data [21]-[24]. While these studies confirm the effectiveness of ML for soil erosion prediction, many existing models prioritize predictive accuracy and rely on complex architectures, often at the expense of interpretability and physical transparency, which limits their direct applicability in engineering practice.

Despite the growing body of research applying ML techniques to soil erosion prediction, several important gaps remain, particularly in terms of model interpretability and feature engineering [25]-[36]. Many existing studies prioritize predictive accuracy and rely on complex learning architectures, while offering limited insight into how individual input variables influence erosion behavior [27]-[29]. Although interpretability tools such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) have been introduced in some cases, their use is inconsistent and often lacks a comprehensive analysis that links model outputs to physical erosion mechanisms [25], [26],[30]-[31]. In addition, feature selection and engineering strategies in erosion-related studies are frequently based on conventional methods, with limited attention given to handling correlated variables or developing physically meaningful feature transformations [31]-[33]. Furthermore, feature engineering is often tailored to specific environmental or climatic conditions, which restricts the general applicability of developed models across different soil systems [25],[34]-[36]. These limitations highlight the need for predictive frameworks that integrate systematic, physically informed feature engineering with transparent and interpretable modelling approaches.

The objective of this study is to develop a reliable and interpretable framework for predicting the IHET without relying on complex or computationally intensive learning-based models. To this end, an interpretable linear model is first established using six fundamental physical and mechanical soil parameters. These parameters are routinely

measured and widely used in geotechnical engineering practice. Their use reflects experimentally based soil characterization approaches reported in related geotechnical studies [37]-[40]. Building on this baseline, a systematic feature-engineering strategy is introduced to expand the original feature space to nineteen physically meaningful variables, while retaining model simplicity. The predictive performance of the proposed feature-engineered linear model is benchmarked against representative ML approaches, namely Gradient Boosting (GB) and a shallow ANN. The results demonstrate that enhancing the feature space within a linear modelling framework can outperform more complex learning-based models, while maintaining transparency and physical interpretability. The proposed approach provides a practical and efficient tool for erosion assessment and preliminary engineering design, offering a balanced alternative between empirical testing and black-box prediction models.

II. DATASET DESCRIPTION AND PREPROCESSING

A. Dataset Description and Soil Parameters

The dataset used in this study consists of experimentally derived soil samples reported by Attom et al. [41] and compiled to support data-driven prediction of the IHET. The dataset includes sixteen types of clayey soils prepared at different initial dry densities and water contents, resulting in a total of forty distinct soil samples, for which erosion parameters were determined independently. The input variables comprise six fundamental soil parameters, grouped into physical properties: LL, PI, w_i , γ_d , and mechanical properties, namely ϕ , and C . These parameters are routinely measured in geotechnical engineering practice and are known to influence soil erosion behaviour. The target variable is IHET, which quantifies the susceptibility of soil to internal erosion under hydraulic shear stress and serves as a key indicator for erosion risk assessment in earth structures.

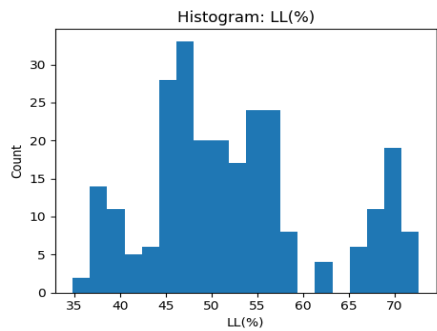
B. Data Preprocessing and Synthetic Expansion

Before model development, the dataset was prepared to ensure consistency and reliability. The cleaned data were divided into training and testing subsets using an 80/20 split. A fixed random seed was used to ensure reproducibility and allow fair comparison among different models. To improve model robustness and mitigate the limitations associated with the small dataset size, the dataset was synthetically expanded using controlled Gaussian noise perturbation and SMOTE-style interpolation applied to all features. Similar augmentation strategies have been successfully applied in engineering and geotechnical studies where Gaussian noise was shown to improve model robustness and generalization when applied in a controlled manner [42], [43]. In addition, SMOTE-based interpolation, either alone or combined with Gaussian noise, has been reported to enhance predictive stability and accuracy in geotechnical and industrial applications [44]-[45]. These studies emphasize the importance of controlling noise levels and combining augmentation techniques to improve

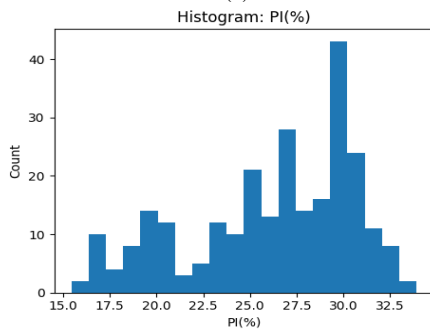
generalization without degrading performance [42],[45]. Accordingly, the synthetic data in this study were used to support model generalization rather than to artificially inflate predictive performance. All reported results are therefore based

on predictions obtained from the unseen test dataset.

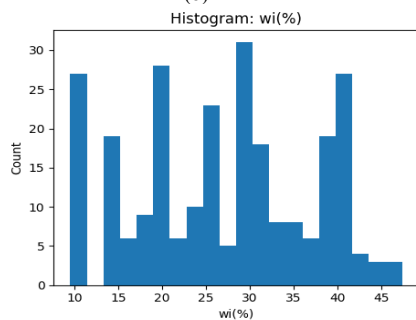
C. Exploratory Data Analysis and Visualization



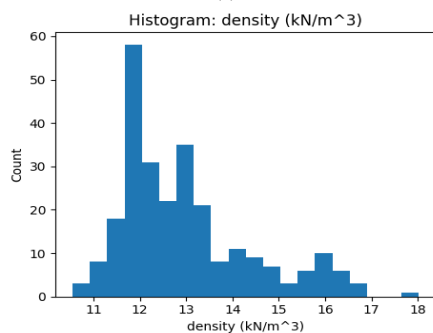
(a)



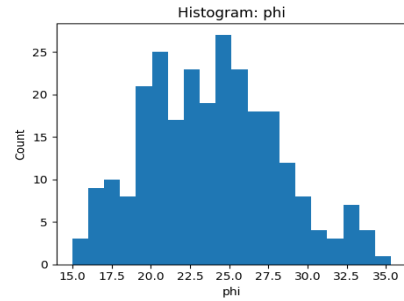
(b)



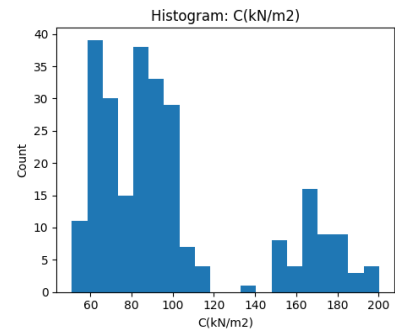
(c)



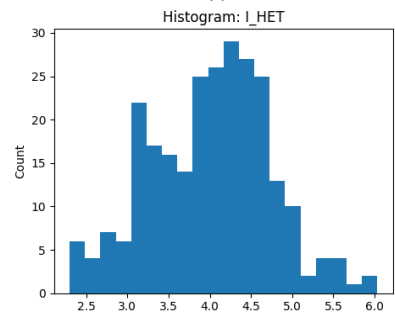
(d)



(e)



(f)



(g)

Fig. 1. Histograms of the six physical and mechanical soil parameters (LL, PI, wi, γ_d , ϕ , and C) and the IHET.

Figure 1 presents the distributions of the six physical and mechanical soil parameters together with IHET. The histograms indicate noticeable variability in all input variables, reflecting the heterogeneous nature of the clayey soils considered in this study. Several parameters exhibit non-uniform and skewed distributions, while IHET also shows a broad spread across its range. This variability suggests that soil erosion behaviour is influenced by multiple interacting factors rather than a single dominant parameter.

Figure 2 shows the Pearson correlation heatmap between the soil parameters and IHET. The results indicate that no single input variable exhibits a strong linear correlation with IHET, while moderate correlations are observed among some soil properties themselves. These patterns suggest the presence of nonlinear relationships and interactions between variables, which may not be fully captured using the original input

parameters alone. Consequently, the observed distributions and correlation structure motivate the use of feature engineering to introduce physically meaningful transformations and interaction terms that can better represent the underlying erosion mechanisms.

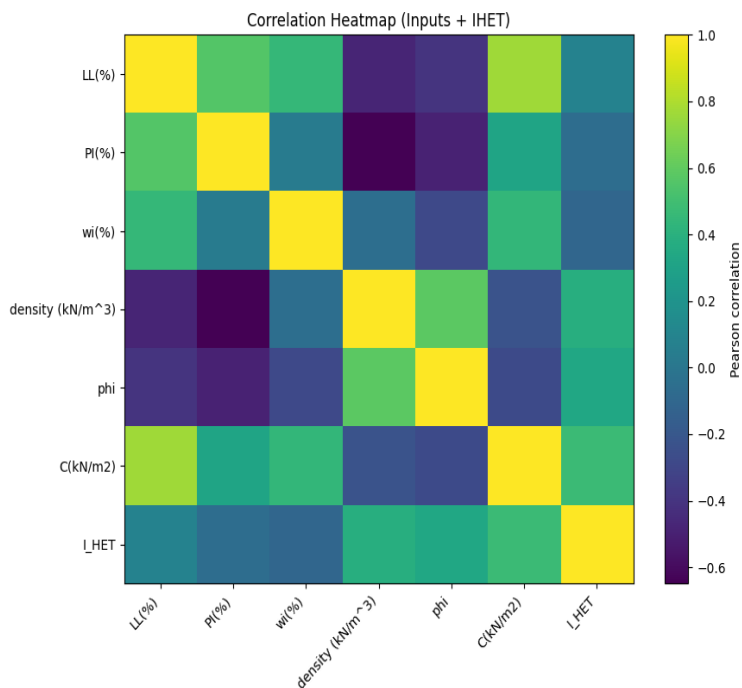


Fig. 2. Correlation heatmap.

I. METHODOLOGY

A. Feature-Engineered Linear Model

To improve the predictive capability of the baseline linear model while preserving interpretability, a systematic feature engineering strategy was adopted. The motivation behind this approach is to enrich the original input space with physically meaningful transformations that can capture nonlinear soil behavior and variable interactions without resorting to complex or black-box learning models. Rather than increasing model complexity, the proposed strategy enhances the descriptive power of a simple linear formulation through informed feature construction. Starting from the six fundamental physical and mechanical soil parameters: LL, PI, w_i , γ_d , ϕ , and C, the feature space was expanded to a total of nineteen variables. The engineered features were grouped into three categories: ratio-based features, logarithmic transformations, and interaction terms. Each category was selected to reflect known geotechnical relationships governing soil plasticity, strength, and erosion resistance. Ratio-based features were introduced to represent relative soil consistency and plasticity characteristics, which are commonly used indicators in soil classification and erosion assessment. Logarithmic transformations were applied to reduce skewness, stabilize variance, and account for the nonlinear influence of soil and erosion parameters across wide

value ranges. Interaction terms were included to capture coupled effects between key mechanical and erosion-related variables, allowing the linear model to implicitly represent nonlinear behavior while maintaining full transparency. The complete set of base and feature-engineered parameters used in the proposed linear model, including ratio-based, logarithmic, and interaction features, is summarized in Table 1.

TABLE I: SUMMARY OF INPUT AND FEATURE-ENGINEERED PARAMETERS USED IN THE LINEAR MODEL.

Category	Feature Type	Number
Base Inputs	Physical & Mechanical Parameters	6
Feature-Engineered	Ratio-Based Features	2
Feature-Engineered	Log-Transformed Features	8
Feature-Engineered	Interaction Features	3
Total	-	19

The final feature set integrates the original six soil parameters with a series of physically informed engineered variables. Ratio-based features quantify relative plasticity and consistency, logarithmic transformations address nonlinearity and variance stabilization, and interaction terms capture coupled mechanical and erosion effects. By embedding these transformations within a linear regression framework, the model gains the capacity to represent complex nonlinear soil behavior while preserving full transparency and interpretability. This approach provides a balance between predictive accuracy and mechanistic understanding, offering a practical alternative to more opaque machine learning techniques.

B. Benchmark Machine Learning Models

For benchmarking purposes, a Gradient Boosting (GB) regression model was trained using the same six physical and mechanical soil parameters. The model employs an ensemble of decision trees optimized sequentially to reduce prediction error and capture nonlinear relationships in the data. The GB model was used strictly for performance comparison and not for optimization, allowing an objective assessment of the proposed feature-engineered linear approach.

II. RESULTS AND DISCUSSION

A. Model Performance Comparison

Figures 3 and 4 present a comparative evaluation of the predictive performance of the baseline linear model, the feature-engineered linear model, and the Gradient Boosting (GB) model using the test dataset. The baseline linear model, based on six original soil parameters, achieved a test R^2 of 0.71 with an RMSE of 0.39, indicating moderate predictive capability. By contrast, the feature-engineered linear model, which expands the feature space to nineteen physically meaningful variables, demonstrates a clear improvement in performance, achieving the highest test R^2 of 0.82 and the lowest RMSE of 0.31.

Although the GB model shows competitive performance

(test $R^2 = 0.79$ and $RMSE = 0.33$), it remains inferior to the feature-engineered linear approach. This behavior is consistent with findings reported in literature, where feature-engineered linear or shallow models often outperform more complex learning algorithms when datasets are limited and governed by physical constraints, owing to their higher data efficiency, robustness, and improved generalization [46]-[48]. Moreover, the ability to incorporate domain knowledge through physically meaningful feature transformations enhances predictive accuracy while preserving interpretability, which is critical for engineering decision-making [46], [49]-[50].

These results highlight that systematic feature engineering within a linear modeling framework can outperform more complex learning-based models while maintaining simplicity, transparency, and physical interpretability. This finding reinforces the effectiveness of feature-space enrichment as a practical and reliable alternative to black-box machine learning models for IHET prediction.

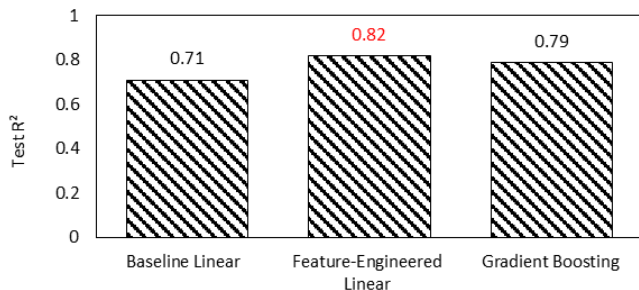


Fig. 3. Comparison of test coefficient of determination (R^2) for the baseline linear model (6 parameters), feature-engineered linear model (19 parameters), and Gradient Boosting model (6 parameters).

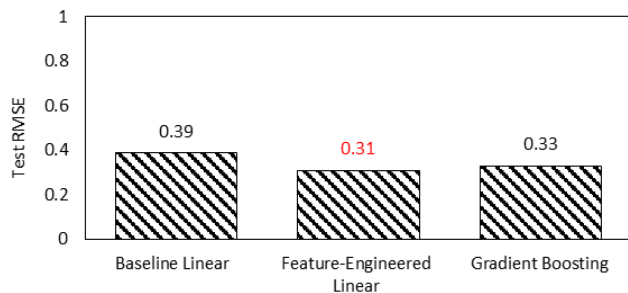


Fig. 4. Comparison of test root mean square error (RMSE) for the baseline linear model (6 parameters), feature-engineered linear model (19 parameters), and Gradient Boosting model (6 parameters).

III. CONCLUSION

This study proposed an interpretable linear modeling framework for predicting the IHET using fundamental physical and mechanical soil parameters. A baseline linear model based on six routinely measured soil properties provided reasonable predictive accuracy, confirming their relevance for erosion assessment. By introducing a systematic feature engineering strategy, the feature space was expanded to nineteen physically meaningful variables, leading to a clear improvement in model performance. The feature-engineered linear model achieved the highest predictive accuracy, increasing the test R^2 from 0.71 to 0.82 and reducing the RMSE from 0.39 to 0.31. When

benchmarked against learning-based models, including Gradient Boosting, the proposed approach consistently outperformed the more complex models despite its simpler structure. These results demonstrate that enhancing the feature space within a linear framework can be more effective than increasing model complexity, particularly for physically constrained and moderately sized geotechnical datasets. Overall, the findings confirm that systematic, physically informed feature engineering significantly improves IHET prediction while preserving model simplicity, transparency, and computational efficiency. The proposed approach offers a practical and interpretable alternative to black-box machine learning models for erosion assessment and preliminary geotechnical design.

IV. ACKNOWLEDGMENTS

The authors gratefully acknowledge the support provided by the American University of Sharjah (AUS) in facilitating this research.

REFERENCES

- [1] Attom, M., Haweeleh, R., Yamin, M. and Astillo, P.V., 2025. Predicting the Erosion Rate Index of Soils Using ANN. In *Management Science and Industrial Engineering* (pp. 37-44). IOS Press.
- [2] Ali, A.K., Hunt, S.L. and Tejral, R.D., 2021. Embankment breach research: Observed internal erosion processes. *Transactions of the ASABE*, 64(2), pp.745-760.
- [3] Wan, C.F. and Fell, R., 2004. Laboratory tests on the rate of piping erosion of soils in embankment dams. *Geotechnical testing journal*, 27(3), pp.295-303.
- [4] Wan, C.F. and Fell, R., 2004. Investigation of rate of erosion of soils in embankment dams. *Journal of geotechnical and geoenvironmental engineering*, 130(4), pp.373-380.
- [5] Farrar, J.A., Torres, R.L. and Erdogan, Z., 2007. Bureau of reclamation erosion testing for evaluation of piping and internal erosion of dams. In *Geotechnics of soil Erosion* (pp. 1-10).
- [6] Chi, F., Carvajal, C., Peyras, L. and Breul, P., 2025. Probabilistic analysis of three different internal erosion mechanisms-Application to a real earth dam. *Engineering Geology*, p.108186.
- [7] Wahl, T.L., 2021. Methods for analyzing submerged jet erosion test data to model scour of cohesive soils. *Transactions of the ASABE*, 64(3), pp.785-799.
- [8] Witt, O. and Westrich, B., 2003. Quantification of erosion rates for undisturbed contaminated cohesive sediment cores by image analysis. *Hydrobiologia*, 494(1), pp.271-276.
- [9] Khalil, A., Khan, Z., Attom, M., Fattah, K., Ali, T. and Mortula, M., 2023. Continuous Evaluation of Shear Wave Velocity from Bender Elements during Monotonic Triaxial Loading. *Materials*, 16(2), p.766.
- [10] Khalil, A., Khan, Z.H., Attom, M., El Emam, M. and Fattah, K., 2022. Dynamic properties of calcareous sands from urban areas of Abu Dhabi. *Applied Sciences*, 12(7), p.3325.
- [11] Gholami, V., Sahour, H. and Amri, M.A.H., 2021. Soil erosion modeling using erosion pins and artificial neural networks. *Catena*, 196, p.104902.
- [12] Nguyen, K.A., Chen, W., Lin, B.S. and Seeboonruang, U., 2020. Using machine learning-based algorithms to analyze erosion rates of a watershed in Northern Taiwan. *Sustainability*, 12(5), p.2022.
- [13] Abiye, W., Alebachew, E.D. and Dengiz, O., 2025. Harnessing machine learning and geospatial technologies for precise soil erodibility mapping and prediction. *Environmental Earth Sciences*, 84(11), p.291.
- [14] Karimi, S., Xu, B., Asgharpour, A., Shirazi, S.A. and Sen, S., 2020, July. Predicting solid particle erosion and uncertainty in elbows by artificial intelligence methods. In *Fluids Engineering Division Summer Meeting* (Vol. 83723, p. V002T04A014). American Society of Mechanical Engineers.

- [15] Khan, I., Khan, K., Bęcek, K. and Bilal, M.F., 2025. Evaluating Water-Induced Soil Erosion Using Machine Learning: XGBoost as the Most Effective Model. *Land Degradation & Development*.
- [16] Ollii, M.R., Ollii, A.K.Z., Ollii, A., Djau, R.A., Mokoagow, M.A., Kironoto, B.A., Bachtiar, B., Ollii, R.S.N. and Pakaya, R., 2025. Tree-based machine learning algorithms for soil erosion vulnerability (SEV) prediction in Sadding Watershed, south Sulawesi, Indonesia. *Journal of Water and Climate Change*, 16(4), pp.1459-1476.
- [17] Sadia, H., Sarkar, S.K. and Haydar, M., 2023. Soil erosion susceptibility mapping in Bangladesh. *Ecological Indicators*, 156, p.111182.
- [18] Avand, M., Mohammadi, M., Mirchooli, F., Kavian, A. and Tiefenbacher, J.P., 2023. A new approach for smart soil erosion modeling: integration of empirical and machine-learning models. *Environmental Modeling & Assessment*, 28(1), pp.145-160.
- [19] Mokarram, M. and Pourghasemi, H.R., 2024. Prediction of soil erosion using machine learning. In *Advanced Tools for Studying Soil Erosion Processes* (pp. 307-322). Elsevier.
- [20] Nguyen, K.A., Chen, W., Lin, B.S. and Seeboonruang, U., 2021. Comparison of ensemble machine learning methods for soil erosion pin measurements. *ISPRS International Journal of Geo-Information*, 10(1), p.42.
- [21] Vu Dinh, T., Hoang, N.D. and Tran, X.L., 2021. Evaluation of different machine learning models for predicting soil erosion in tropical sloping lands of Northeast Vietnam. *Applied and Environmental Soil Science*, 2021(1), p.6665485.
- [22] Dinh, T.V., Nguyen, H., Tran, X.L. and Hoang, N.D., 2021. Predicting Rainfall-Induced Soil Erosion Based on a Hybridization of Adaptive Differential Evolution and Support Vector Machine Classification. *Mathematical Problems in Engineering*, 2021(1), p.6647829.
- [23] Li, H., Miao, S., Qi, Y., Gao, H., Duan, H., Liu, C. and Gao, W., 2025. A Comprehensive Analysis of Soil Erosion in Coastal Areas Based on an Unmanned Aerial Vehicle and Deep Learning Approach. *Sustainability* (2071-1050), 17(3).
- [24] Priyadharshini, V.M., Aldehim, G., Negm, N. and Subathradevi, S., 2025. Integrating geospatial techniques and machine learning for assessing soil erosion and associated geomorphic risks. *Journal of South American Earth Sciences*, 156, p.105463.
- [25] Gholami, H., Darvishi, E., Moradi, N., Mohammadifar, A., Song, Y., Li, Y., Niu, B., Kaskaoutis, D. and Pradhan, B., 2024. An interpretable (explainable) model based on machine learning and SHAP interpretation technique for mapping wind erosion hazard. *Environmental Science and Pollution Research*, 31(56), pp.64628-64643.
- [26] Mahmud, T., Datta, N., Chakma, R., Das, U.K., Aziz, M.T., Islam, M., Salimullah, A.H.M., Hossain, M.S. and Andersson, K., 2024. An approach for crop prediction in agriculture: Integrating genetic algorithms and machine learning. *IEEE Access*.
- [27] Bolsynbek, M., Abdikerimova, G., Serikbayeva, S., Batyrkhanov, A., Shrymbay, D., Tazhurekova, Z., Zhidkulova, G. and Shraimanova, G., 2025. Soil erosion analysis based on machine learning method. *Bulletin of Electrical Engineering and Informatics*, 14(6), pp.4797-4811.
- [28] Bouamrane, A., Kantoush, S.A., Dahri, N., Bouamrane, A., Saber, M., Boutaghane, H., Abida, H., Sumi, T., Sohei, K. and Mabrouk, E., 2025. Simulation of erosion susceptibility in the Medjerda watershed under different climate change scenarios. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, pp.1-24.
- [29] Khan, I., Khan, K., Bęcek, K. and Bilal, M.F., 2025. Evaluating Water-Induced Soil Erosion Using Machine Learning: XGBoost as the Most Effective Model. *Land Degradation & Development*.
- [30] Zhou, S., Zhu, Q. and Li, S., 2025. Quantifying Elevation Changes Under Engineering Measures Using Multisource Remote Sensing and Interpretable Machine Learning: A Case Study of the Chinese Loess Plateau. *Remote Sensing*, 17(20), p.3451.
- [31] Li, Y., Zhang, J., McKinney, B.A., Karimi, S. and Shirazi, S.A., 2025, July. Enhancing Erosion Rate Prediction Through NPDR-Based Feature Selection and Machine Learning Models. In *Fluids Engineering Division Summer Meeting* (Vol. 88995, p. V001T01A006). American Society of Mechanical Engineers.
- [32] Zeng, X., Cai, S., Liu, C. and Zhou, F., 2025, August. Design and Development of Soil and Water Conservation Effect Evaluation Based on Feature Selector and Learning Algorithm. In *2025 Third International Conference on Networks, Multimedia and Information Technology (NMITCON)* (pp. 1-7). IEEE.
- [33] Li, Y., Zhang, J., McKinney, B., Karimi, S. and Shirazi, S.A., 2024, July. Linear Versus Non-Linear Machine Learning Feature Selection for Erosion Rate Prediction. In *Fluids Engineering Division Summer Meeting* (Vol. 88131, p. V002T05A005). American Society of Mechanical Engineers.
- [34] Gemechu, G.F. and Wei, W., 2025. An Explainable Machine Learning and Cloud-Based Remote Sensing Framework for Monitoring Terrace Degradation and Restoration in Mountain Landscapes. *Land Degradation & Development*.
- [35] Gholami, H., Mohammadifar, A., Fitzsimmons, K.E., Li, Y. and Kaskaoutis, D.G., 2023. Modeling land susceptibility to wind erosion hazards using LASSO regression and graph convolutional networks. *Frontiers in Environmental Science*, 11, p.1187658.
- [36] Minghao, Z., Wei, Q., Xiangdong, W., Zhe, Y., Xiaohong, X., Liheng, Z. and Xing, H., 2026. Assessment of gully erosion susceptibility in Northeast China's black soil region using new stacking model with multiple machine learning algorithms. *Soil and Tillage Research*, 257, p.106964.
- [37] Attom, M., El-Kadri, O. and Khalil, A., 2024, August. Optimization of Geothermal Energy Piles Design Parameters to Enhance Thermal Efficiency. In *2024 5th International Conference on Clean and Green Energy Engineering (CGEE)* (pp. 83-87). IEEE.
- [38] Khalil, A., Khan, Z., Attom, M. and Khalafalla, O., 2024. Evaluation of Ground Improvement with Dynamic Replacement and Rapid Impact Compaction of an Artificial Island in the UAE—A Case Study. In *Geo-Congress 2024* (pp. 116-125).
- [39] Attom, M., Al-Asheh, S., Yamin, M., Vandanapu, R., Al-Lozi, N., Khalil, A. and Eltayeb, A., 2025. Soil Improvement Using Plastic Waste—Cement Mixture to Control Swelling and Compressibility of Clay Soils. *Buildings*, 15(8), p.1387.
- [40] Khalil, A., Attom, M., Khan, Z., Astillo, P.V. and El-Kadri, O.M., 2024. Recent Advancements in Geothermal Energy Piles Performance and Design. *Energies*, 17(14), p.3386.
- [41] Attom, M.F., Vandanapu, R., Khan, Z., Yamin, M., Astillo, P.V., Eltayeb, A. and Khalil, A., 2024. Prediction of internal erosion parameters of clay soils using initial physical properties. *Water*, 16(2), p.232.
- [42] Ma, Y., Qu, D., Yan, E., Shi, H., Zhang, X. and Guo, J., 2025, September. Estimating soil thermal properties in ground source heat pump zones with artificial neural networks and enhanced data techniques. In *Second International Conference on Artificial Intelligence and Power System (AIPS 2025)* (Vol. 13786, pp. 255-262). SPIE.
- [43] Zhao, C., Mahmoudi, E., Luo, M., Jiang, M. and Lin, P., 2023. Unfavorable geology recognition in front of shallow tunnel face using machine learning. *Computers and Geotechnics*, 157, p.105313.
- [44] Liu, Y., Sun, T., Wu, K., Xiang, W., Zhang, J., Zhang, H. and Feng, M., 2025. Interpretability Analysis of Data Augmented Convolutional Neural Network in Mineral Prospectivity Mapping Using Black-Box Visualization Tools. *Natural Resources Research*, 34(2), pp.759-783.
- [45] Yoo, J., 2025. Enhancing Nickel Matte Grade Prediction Using SMOTE-Based Data Augmentation and Stacking Ensemble Learning for Limited Dataset. *Processes*, 13(3), p.754.
- [46] Batista, J.E., Pindur, A.K., Iba, H. and Silva, S., 2024, June. Measuring structural complexity of gp models for feature engineering over the generations. In *2024 IEEE Congress on Evolutionary Computation (CEC)* (pp. 1-8). IEEE.
- [47] Chang, Y.C., 2025. Multiple-Stream Models for a Single-Modality Dataset with Fractal Dimension Features. *Fractal and Fractional*, 9(4), p.248.
- [48] Batista, J.E., Pindur, A.K., Cabral, A.I., Iba, H. and Silva, S., 2025. Complexity, interpretability and robustness of GP-based feature engineering in remote sensing. *Swarm and Evolutionary Computation*, 92, p.101761.
- [49] Pei, T. and Luo, N., Prediction of Ultimate Bearing Capacity of Shallow Foundations Using Physically Constrained Neural Networks. In *Geotechnical Frontiers 2025* (pp. 11-20).
- [50] Horn, F., Pack, R. and Rieger, M., 2019, September. The autofeat python library for automated feature engineering and selection. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 111-120). Cham: Springer International Publishing.