Comparison of Machine Learning Models in Predicting Dam Stability

Theshikan Sundararajah
Department of civil engineering
University of Moratuwa
Sri Lanka
theshikans.21@uom.lk

N.M.M. Nihaaj
Department of Civil Engineering
University of Moratuwa
Moratuwa, Sri Lanka
nihajnoor@gmail.com

L.L. Ekanayake
Department of Civil Engineering
University of Moratuwa
Moratuwa, Sri Lanka
lesly@uom.lk

I. Introduction

Granular filters are integral to the safety and operational performance of soil embankment dams, as they effectively mitigate internal erosion and piping phenomena. These filters function as protective barriers designed to retain base soil particles while simultaneously facilitating the safe passage of the seepage water. An effective filter design will have pore sizes that are sufficiently small to prevent the migration of underlying soil particles, while allowing seepage to continuously dissipate. Failing to provide optimum filter and base particle combination caused nearly 50% of dam failures occur due to internal erosion Foster et al (2000).

Concurrently, the filter material must maintain adequate coarse particles to promote free drainage and prevent excessive pore water pressure, making a clear understanding of the interactions between base soil and filter materials essential for effective filter systems in embankment dam construction. Although numerous relationships have been proposed regarding particle size distributions, the complexity of combining filter and base soil gradations has prevented the development of a perfect criterion. To better address these complexities, machine learning provides a powerful approach, enabling the analysis of nonlinear and interdependent soil property patterns using experimental datasets and helping identify the most accurate models for predicting soil behavior and stability

II. Literature Review

Internal stability of sandy soils has been evaluated using several indices proposed by different researchers. Kezdi (1969) introduced the internal stability index I_r, defined as the ratio D_{15c}/d_{85f}, where the particle size distribution (PSD) curve is divided into filter and base soil components. Here, D_{15c} corresponds to the particle diameter at 15% by weight of filter particles, and d_{85f} refers to the particle diameter at 85% by weight of base particles, providing a measure to assess the soil's internal stability. Kenney et al. (1985) proposed the use of $(H/F)_{min}$ as a discriminant index to identify the potential for internal erosion within soil. In this formulation, H represents the mass fraction smaller than a random particle diameter D, while F is the mass fraction between particles of size D and 4D. Earlier, Istomina (1957) suggested the coefficient of uniformity (C_u) as an indicator for evaluating the internal stability of sandy soils, which remains a widely referenced parameter in soil mechanics for its simplicity and effectiveness.

Zhang et al. (2022) reviewed ML applications for predicting soil properties and introduced a tool integrating six algorithms, demonstrating its effectiveness through MDD prediction and model ranking. Kharnoob et al. (2025) used hybrid ANN models optimized with PSO and GA to estimate soil permeability. Prajapati and Dey (2025) applied multiple ML algorithms to estimate filter dimensions for earthen embankment dams, demonstrating that these models capture the complex relationships between soil properties and filter design more effectively than traditional empirical methods. Together, these studies highlight the value of optimized and comparative ML approaches in enhancing soil property prediction

Wang et al. (2025) developed high-performance ML models to predict soil internal stability using eight key grain-size and gradation features, including d_5–d_{90}, Cu, $(H/F)_{min}$, and $(D_{c15}/d_{f85})_{max}$. They compared several algorithms—Logistic Regression, SVM, Random Forest, XG Boost, Light GBM, and Cat Boost to identify the most accurate model for binary stability classification. Building on previous ML-based studies, their work focuses on determining the best-performing model for predicting soil stability using factors commonly applied in assessing optimal filter–base soil compatibility.

III. Materials and Methods

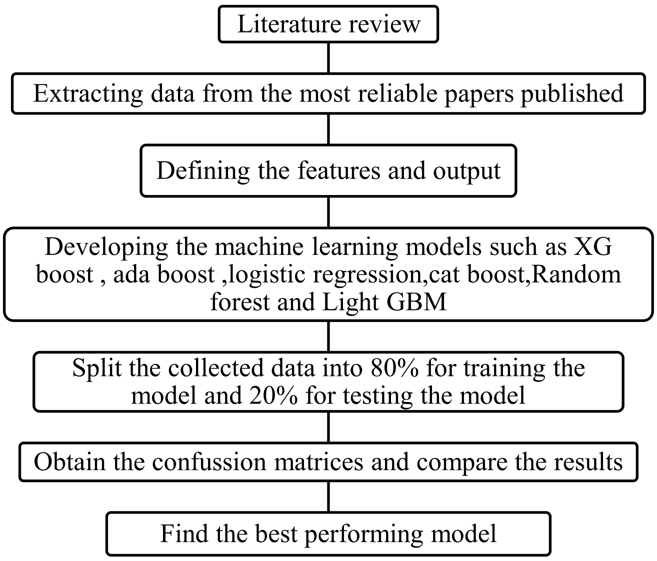

Fig. 1. Methodology flow chart

Above diagram shows the methodology flow diagram of this research. Almost 160 experimental data were extracted from reliable literature sources After extracting the data it was categorized as features and output the model which consist of the C_u and $(D_{15c}/ d_{85f})_{max}$, and $(H/F)_{min}$. these three parameters of the soil type and the experimental results whether the soil

ERU Symposium 2025 - DOI 10.31705/ERU.2025.28

is stable or unstable during seepage condition is extracted and these data are used to train and test machine learning models. Fig2 shows the three features and the outcome of the machine learning model. The performance of logistic regression, Ada boost, XG boost, Cat Boost, Light GBM and Random Forest models were analyzed through a confusion matrix

Features		Outcome
$(H/F)_{min}$ C_u $(d_{15c}/D_{85f})_{max}$	—	Soil stability (stable -1/unstable -0)

Fig. 2. Figure mentioning features and outcome of the data

IV. Results and Discussion

All six machine learning models were successfully trained and evaluated, with each model generating and displaying its respective confusion matrix. The Table I, Table II, Table III, Table IV, Table V , Table VI, shows the confusion matrices of Logistic regression, Ada Boost, XG Boost, Random Forest, Cat Boost and Light GBM respectively. Remarkably, All models perfectly classified Stable instances, showing zero Type I error—meaning none misclassified an actual unstable case as stable, and all correctly identified every unstable instance.

TABLE I. Results of the logistic regression model

		Predicted		
		Stable	*Unstable*	
Actual	**Stable**	20 (TS)	1 (FU)	Accuracy: 0.97
				Precision: 1.00
	Unstable	0 (FS)	11 (TU)	Sensitivity: 0.95
				Specificity: 1.00

TABLE II. Results of the Ada Boost model

		Predicted		
		Stable	*Unstable*	
Actual	*Stable*	19 (TS)	2 (FU)	Accuracy: 0.94
				Precision: 1.00
	Unstable	0 (FS)	11 (TU)	Sensitivity: 0.90
				Specificity: 1.00

TABLE III. Results of the XG Boost model

		Predicted		
		Stable	*Unstable*	
Actual	*Stable*	18 (TS)	3 (FU)	Accuracy: 0.91
				Precision: 1.00
	Unstable	0 (FS)	11 (TU)	Sensitivity: 0.86
				Specificity: 1.00

TABLE IV. Results of the Random Forest Model

		Predicted		
		Stable	*Unstable*	
Actual	*Stable*	18 (TS)	3 (FU)	Accuracy: 0.91
				Precision: 1.00
	Unstable	0 (FS)	11 (TU)	Sensitivity: 0.86
				Specificity: 1.00

TABLE V. Results of the Cat boost model

		Predicted		
		Stable	*Unstable*	
Actual	*Stable*	17 (TS)	4 (FU)	Accuracy: 0.88
				Precision: 1.00
	Unstable	0 (FS)	11 (TU)	Sensitivity: 0.81
				Specificity: 1.00

TABLE VI. Results of Light GBM

		Predicted		
		Stable	*Unstable*	
Actual	*Stable*	17 (TS)	4 (FU)	Accuracy: 0.88
				Precision: 1.00
	Unstable	0 (FS)	11 (TU)	Sensitivity: 0.81
				Specificity: 1.00

However, the models mainly differed in Type II error—the misclassification of stable samples as unstable, reflected in false unstable cases. Logistic Regression had the fewest, showing the best class separation. AdaBoost produced two false unstable cases, XG Boost and Random Forest three each, while Cat Boost and Light GBM had the most, indicating slightly lower precision. These differences reflect variations in model decision boundaries and sensitivity to class imbalance.

TABLE VII. Summary of Results Obtained

Model	Accuracy	Precision	Sensitivity	Specificity
Random Forest	0.91	1.00	0.86	1.00
Ada Boost	0.94	1.00	0.90	1.00
XG Boost	0.91	1.00	0.86	1.00
Light GBM	0.88	1.00	0.81	1.00
Cat Boost	0.88	1.00	0.81	1.00
Logistic regression	0.94	1.00	0.95	1.00

V. Conclusion

Given the Logistic Regression model demonstrated the best performance, achieving the highest overall accuracy. The AdaBoost model followed closely, showing slightly lower accuracy compared to Logistic Regression. Both XG Boost and Random Forest models exhibited similar performance levels, while the more advanced gradient boosting models, Light GBM and Cat Boost, produced comparable yet relatively lower accuracy values. This reduction in performance among the advanced models may be attributed to the limited size of the available dataset, which can restrict their ability to generalize effectively and fully leverage their complex learning capabilities.

References

[1] V. S. Istomina, *Filtration Stability of Soils.* Moscow–Leningrad, Russia: Gostroizdat, 1957.

[2] T. C. Kenney and D. Lau, “Internal stability of granular filters,” *Canadian Geotechnical Journal*, vol. 22, no. 2, pp. 215–225, 1985.

[3] A. Kézdi, *Soil Physics: Selected Topics*, Developments in Geotechnical Engineering, vol. 25. Amsterdam, Netherlands: Elsevier Scientific Publishing, 1979.

[4] M. Foster, R. Fell, and M. Spannagle, “The statistics of embankment dam failures and accidents,” *Canadian Geotechnical Journal*, vol. 37, no. 5, pp. 1000–1024, 2000

[5] Wang, H.-L. and Yin, Z.-Y., “High performance prediction of soil compaction parameters using multi expression programming,” *Engineering Geology*, vol. 276, 105758, Oct. 2020

[6] P. Zhang, Z.-Y. Yin, and Y.-F. Jin, "Machine Learning-Based Modelling of Soil Properties for Geotechnical Design: Review, Tool Development and Comparison," *PolyU Institutional Research Archive*, 2022

[7] Y. Wang, H. Wang, and J. Wang, “High performance prediction of soil internal stability using machine learning-based models,” *Available at SSRN*,