

Binder Chemistry and Its Impact on Geopolymer Compaction: Predicting Compaction Parameters

T.J. Fahee
 Department of Civil Engineering
 University of Moratuwa
 Moratuwa, Sri Lanka
 tjfahee@gmail.com

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

Keywords—Geopolymer, Soil Stabilization, Machine Learning, Binder Chemistry, Compaction

I. INTRODUCTION

Soil stabilization enhances the engineering performance of weak soil [1], [2]. Chemical stabilization is a common method that incorporates additives to form strong interparticle bonds [2]. Traditional binders like ordinary Portland cement (OPC) and lime have been widely used for stabilization, but their production causes about 7% of global greenhouse gas emissions and 11% of energy intake [1], [2], [3]. Geopolymer, which is also referred to as alkali-activated binders (AAB), is a sustainable alternative to OPC in soil stabilization [1], [2], [4], [5]. Geopolymer involves the alkali activation of aluminosilicate precursors like fly ash (FA), rice husk ash (RHA), and ground granulated blast furnace slag (GGBS) [1], [2], [3]. The use of geopolymers enables a substantial reduction in process-related carbon emissions, approximately 60-80% compared to conventional OPC production [4]. Each precursor contains a unique combination of chemical compounds that strongly influences the geopolymerization reactions and the resulting material properties.

Binder chemistry governs geopolymer performance through major oxides such as SiO_2 , Al_2O_3 , CaO , Na/Al , and Si/Al ratios, which influence geopolymerization and gel formation [1], [5]. Key geotechnical parameters like maximum dry density (MDD) and optimum moisture content (OMC) help to evaluate stabilization efficiency [1]. However, predicting these parameters from binder chemistry remains complex. Instead of relying on random trial-and-error to determine the optimal composition, it is essential to establish a systematic mechanism for formulating effective soil reinforcement. Therefore, this study aims to develop a standardized predictive model to estimate compliance with required mechanical properties, moving beyond inefficient trial-and-error methods [1], [2].

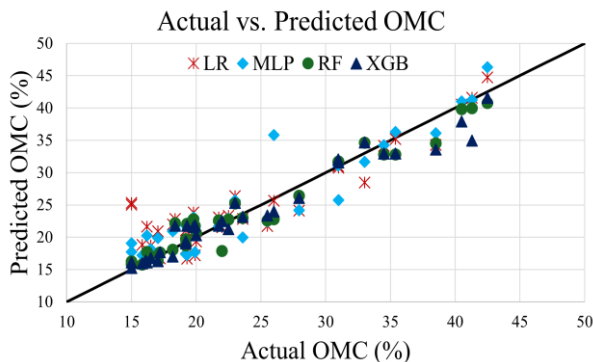


Fig. 1. Predicted Vs experimental values of OMC for different machine learning algorithms

II. LITERATURE REVIEW

Reference [6] mentioned that the binder chemistry determines the performance of geopolymerization through the polycondensation reaction of aluminosilicate oxides. Reference [5] mentioned that this enhanced performance is attributed to the formation of robust structures where C-S-H and C-A-S-H products coexist with N-A-S-H/K-A-S-H geopolymer arrangements contributing to better reactivity, more effective formation of cementitious products, and greater strength development. References [3] and [6] mentioned that the geopolymerization kinetics and resulting strength are mainly governed by the silicate-to-aluminate (Si/Al) and sodium-to-aluminum (Na/Al) ratios. Reference [5] found that stabilizing expansive soil generally increases MDD by 14.7-176.2% and reduces OMC by 22.8-55.6%, mainly due to binder filling of soil pores and the lubrication effect of the AAB coating. However, certain aluminosilicates like RHA and MK can slightly reduce MDD by 0.5-9.3% and increase OMC by 2.3-31.6%. Lower AAB contents may even produce higher OMC and lower MDD than untreated soil, showing that AAB-stabilized clays do not always follow typical compaction trends.

Beyond experimental approaches, recent studies have increasingly employed machine learning (ML) techniques to model complex binder-soil interactions. Reference [4] employed ML models such as random forest (RF), support vector regression (SVR), multilayer perceptron (MLP), gradient boosting decision tree (GBDT), and extreme gradient boosting (XGBoost) for predicting the unconfined compressive strength (UCS) of GGBS-FA-based geopolymer stabilized soil. Reference [1] employed an MLP model to create a composition-informed mix design framework based on the chemical composition of precursors and activators. Since compaction characteristics such as MDD and OMC are

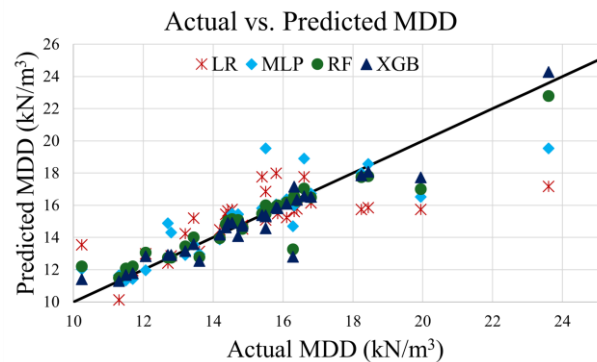


Fig. 2. Predicted Vs experimental values of MDD for different machine learning algorithms

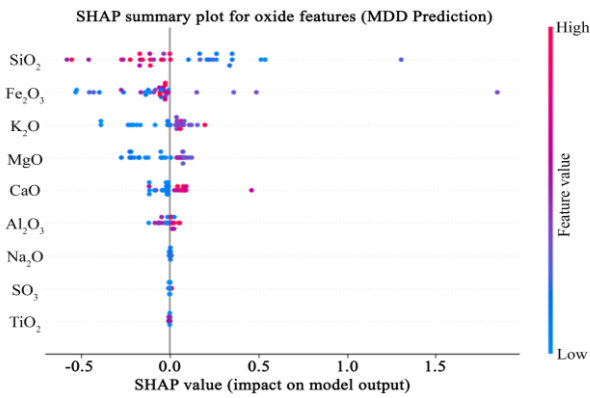


Fig. 3. SHAP summary plot for MDD prediction

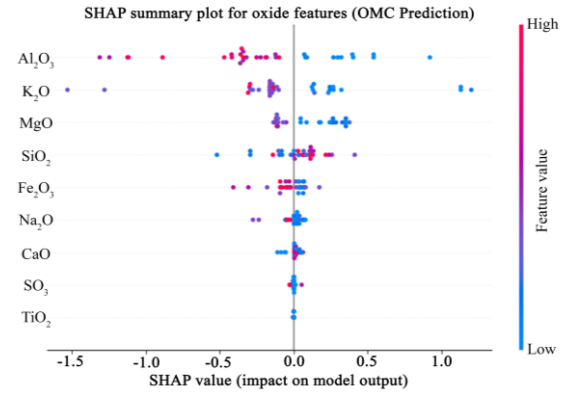


Fig. 4. SHAP summary plot for OMC prediction

the most critical parameters required in field applications, developing a simple and reliable machine learning model capable of predicting these properties from binder and soil properties is essential for practical implementation. This research intends to predict the compaction parameter along with the binder chemistry and soil properties. This model can reduce the requirement of making field trial and error experiments.

III. DATA COLLECTION

A dataset was compiled from peer-reviewed literature, containing 161 records for OMC and MDD from 14 sources. To ensure data diversity, the dataset encompasses various soil types (plastic index between 19% - 59.5%) and binder sources (class C and class F fly ash, and rice husk ash). Each data entry contained 14 input features: soil properties (LL, PL, PI, SG), oxide compositions (SiO₂, Al₂O₃, Fe₂O₃, CaO, MgO, Na₂O, K₂O, TiO₂, and SO₃), and binder percentage. These features were selected based on their established role in geopolymerization kinetics and soil-binder physical interaction [5]. All literature determines OMC and MDD with changing binder percentage solely. Thus, binder percentage is considered a feature. OMC and MDD were modelled separately. Data were normalized to eliminate scale bias, and an 80:20 split was used for training and testing.

IV. MODEL TRAINING AND TESTING

Four regression algorithms were employed: Linear Regression (LR), RF, MLP, and XGBoost using Python (scikit-learn). To enhance model reproducibility and robustness, hyperparameters were optimized using grid search with 5-fold cross-validation. Each output variable was trained independently with identical inputs. Model accuracy was evaluated using root mean squared error (RMSE) and coefficient of determination (R²). A higher R² value indicates a stronger correlation between predicted and actual data, while lower RMSE values reflect reduced prediction deviation. Table I contains the evaluated results of MSE, RMSE and R² values of all models for OMC and MDD.

TABLE I. R² AND RMSE VALUES FOR THE LR, RF, MLP AND XGBOOST

		LR	RF	MLP	XGBoost
RMSE	OMC (%)	3.5808	1.9057	2.9703	2.0065
	MDD (kNm ⁻³)	1.8265	0.9143	1.4173	0.8435
R ²	OMC (%)	0.8128	0.9470	0.8712	0.9412
	MDD (kNm ⁻³)	0.5061	0.8762	0.7026	0.8947

V. DISCUSSION AND CONCLUSION

For OMC, the RF achieved the highest accuracy (R² = 0.9470, RMSE = 1.95), closely followed by XGBoost (R² = 0.9412, RMSE = 2.01). For MDD, XGBoost achieved the best performance (R² = 0.8947, RMSE = 0.84), followed by RF (R² = 0.8762, RMSE = 0.91). In contrast, the LR exhibited weak predictive ability, particularly for MDD, with an R² of 0.5061. This is visually evident in Fig. 1 and Fig. 2, where LR predictions show significant deviation from the line of best fit. Overall, XGBoost and RF demonstrated the most balanced and consistent performance. To interpret these complex interactions, SHAP analysis was conducted on the XGBoost model. As shown in Fig. 3, SiO₂ negatively impacts MDD due to its low specific gravity, while Fe₂O₃ and CaO enhance density. Similarly, Fig. 4 identifies Al₂O₃ as the dominant negative predictor on OMC prediction, indicating that reactive alumina consumes pore fluid and lowers hydrophilicity. A similar negative correlation was observed with K₂O. These results confirm that nonlinear models better capture the complex influence of oxide composition and compaction parameters. Practically, this allows engineers to use local binders for compatibility without extensive lab testing and directly contributes to sustainable practices aligned with SDG 9 (Industry, Innovation & Infrastructure) and SDG 13 (Climate Action).

REFERENCES

- [1] J. Zhang, C. E. Choi, Z. Liang, and R. Li, "A generic framework for mix design of geopolymer for soil stabilization: Composition-informed machine learning model," *Comput Geotech*, vol. 170, Jun. 2024, doi: 10.1016/j.compgeo.2024.106322.
- [2] A. A. Disu and P. K. Kolay, "A Critical Appraisal of Soil Stabilization Using Geopolymers: The Past, Present and Future," *International Journal of Geosynthetics and Ground Engineering*, vol. 7, no. 2, Jun. 2021, doi: 10.1007/s40891-021-00267-w.
- [3] M. M. A. L. N. Maheepala, M. C. M. Nasvi, D. J. Robert, C. Gunasekara, and L. C. Kurukulasuriya, "Mix design development for geopolymer treated expansive subgrades using artificial neural network," *Comput Geotech*, vol. 161, Sep. 2023, doi: 10.1016/j.compgeo.2023.105534.
- [4] J. Hu, Z. Li, X. Qin, S. Dong, T. Zhao, and H. Li, "Machine learning-based optimization of geopolymer-stabilized soil formulation using blast furnace slag and fly ash," *Mater Today Commun*, vol. 48, Sep. 2025, doi: 10.1016/j.mtcomm.2025.113500.
- [5] M. M. A. L. N. Maheepala, M. C. M. Nasvi, D. J. Robert, C. Gunasekara, and L. C. Kurukulasuriya, "A comprehensive review on geotechnical properties of alkali activated binder treated expansive soil," *J Clean Prod*, vol. 363, Aug. 2022, doi: 10.1016/j.jclepro.2022.132488.
- [6] D. C. Nwonu, "Exploring soil geopolymer technology in soft ground improvement: a brief excursion," *Arabian Journal of Geosciences*, 2021, doi: 10.1007/s12517-021-06832-x/Published.