Integrating Atterberg Limits with Machine Learning for Screening-Scale Prediction of Volumetric Behavior in Pathum Thani Clay Soils

Rattanachot Thongpong[1] and Pattaraporn Nueasri[2, *]

[1] *Department of Civil Engineering Technology, Faculty of Industrial Technology, Valaya Alongkorn Rajabhat University under the Royal Patronage, Khlong Nueng, Khlong Luang, Pathum Thani, 13180, Thailand*

[2] *Department of Engineering Management, Faculty of Industrial Technology, Valaya Alongkorn Rajabhat University under the Royal Patronage, Khlong Nueng, Khlong Luang, Pathum Thani, 13180, Thailand*

***Corresponding Author E-mail:** pattaraporn.nuea@vru.ac.th

Received: Aug 04, 2025; **Revised:** Oct 18, 2025; **Accepted:** Oct 28, 2025

Abstract

This study investigates the predictive relationship between Atterberg limits and the volumetric ratio behavior (VLL, VPL, and VSL) of tropical clay soils. Laboratory testing, following ASTM D4318, was conducted on 50 clay samples collected from Pathum Thani Province, Thailand. Estimated volumetric ratios were derived from mass–moisture–density relationships representing liquid, plastic, and shrinkage states. Linear, polynomial, and machine learning models, including Random Forest and Support Vector Regression (SVR), were developed to evaluate the statistical association between index parameters and volume change behavior. The models showed weak-to-moderate correlation (R^2 = 0.31–0.55), indicating that the derived relationships can provide qualitative insights rather than quantitative predictions, supporting a conceptual understanding of soil volume behavior. The Shrinkage Limit (SL) consistently emerged as the most influential parameter, reflecting its strong association with moisture-induced volume reduction and soil–water interaction mechanisms. The results suggest that Atterberg limits can serve as qualitative indicators of volumetric change potential rather than quantitative predictors. Although the models exhibited low explanatory power, they provide transparent, reproducible insights into how index-based soil properties correspond to volumetric transitions. This framework supports early-stage and cost-effective assessment of expansive soils, offering a practical foundation for identifying shrink–swell tendencies before advanced testing. The approach contributes to improving preliminary geotechnical evaluation practices in tropical environments and establishes a reference for future validation incorporating mineralogical and suction-related parameters.

Keywords: Atterberg Limits, Polynomial regression, Random Forest, SVR, Volumetric prediction

1. Introduction

In tropical and subtropical regions, the volumetric behavior of clayey soils poses significant challenges to geotechnical engineers, particularly in infrastructure projects exposed to seasonal moisture fluctuations. Swelling and shrinkage of expansive clays can cause foundation heave, pavement cracking, and differential settlement, often leading to long-term structural degradation [1],[2]. These behaviors are governed primarily by moisture variation and mineralogical composition, both of which are closely associated with the soil's index properties, especially the Atterberg limits [3–6].

Pathum Thani Province, situated in Thailand's central floodplain, is characterized by soft, fine-grained clays that exhibit substantial volume changes between wet and dry seasons. The geotechnical response of these soils is largely influenced by their high Liquid Limit (LL), Plastic Limit (PL), and Shrinkage Limit (SL) values [7],[8]. Although these parameters are widely applied in soil classification and plasticity characterization, their role as predictive indicators of volumetric change remains underexplored. Traditional applications of Atterberg limits are largely qualitative, whereas advanced models such as those of Fredlund and Xing [9] and van Genuchten [10] rely on soil suction or water retention data that are often unavailable in practical site investigations.

Recent studies have attempted to use statistical and machine learning methods to model expansive soil behavior. Barbosa et al. [11] developed regression models for swelling potential based on classification indices, while Puppala et al. [12] introduced strain-based models requiring in-situ monitoring. Al-Taie et al. [13] analyzed volumetric reduction after lime stabilization, which necessitates chemical treatment data not representative of natural soils. However, most of these models depend on site-specific datasets or specialized instrumentation, limiting their transferability to tropical regions. Although the Plasticity Index (PI) provides valuable insight into soil deformation potential under moisture variation, its application in volumetric prediction frameworks remains limited. Bhavya and Nagaraj [3] examined its microstructural implications, yet integration into empirical modeling has been minimal. Likewise, Yukselen and Kaya [14] reported correlations between Atterberg limits, surface area, and cation exchange capacity (CEC), but their findings have not been adapted to tropical clays characterized by high organic content and advanced weathering. In light of these gaps, this study develops an empirical framework for predicting the volume ratio behavior of clay soils (VLL, VPL, VSL) using Atterberg limits as predictors. Fifty undisturbed clay samples from Pathum Thani were tested in accordance with ASTM D4318 to establish locally calibrated regression and support vector models. The proposed approach provides a transparent, data-efficient alternative to complex suction-based methods, offering practical support for preliminary geotechnical assessment and climate-resilient foundation design in tropical environments. Volume change in fine-grained soils is primarily governed by inter-particle water adsorption and clay mineral expansion. The Atterberg Limits reflect these microstructural changes; for instance, a higher LL typically indicates a greater proportion of expansive minerals such as montmorillonite, leading to higher volume ratios at constant stress

The rationale for relating Atterberg limits to volumetric ratios lies in the moisture-dependent microstructural rearrangements of clay. Each consistency limit reflects a transition in particle configuration and the thickness of the adsorbed water film [15],[16]. Consequently, these limits indirectly encode volumetric behavior during drying–wetting cycles, providing a practical means for preliminary estimation where advanced volumetric or swell tests are unavailable.

2. Materials and Methods

To systematically examine the relationship between Atterberg limits and soil volumetric behavior, a laboratory-based experimental framework was developed. All procedures were conducted under controlled conditions following standard geotechnical testing protocols to ensure consistency and reliability across multiple clay samples. Since the standard Atterberg limit tests (ASTM D4318) do not directly measure volumetric changes at different moisture thresholds, an estimation approach was introduced to derive volumetric ratios that represent the relative change in soil volume corresponding to the liquid, plastic, and shrinkage states. These ratios were computed from the measured relationships between soil mass, moisture content, and dry density, in accordance with established soil mechanics principles and previous studies [13],[15].

In addition to the standard index property testing, supplementary estimation and verification steps were incorporated to define and validate the volumetric ratios (VLL, VPL, and VSL). These steps included consistency checks among replicate samples and comparative evaluation with theoretical density–moisture trends derived from compaction principles. The adopted methodology emphasizes transparency, reproducibility, and data efficiency, providing a practical alternative to direct volumetric measurements that require complex instrumentation.

2.1 Soil Sampling and Preparation

Clay soil samples were collected from five locations within Khlong Nueng Subdistrict, Khlong Luang District, Pathum Thani Province, Thailand, representing urban clayey ground with a shallow water table. Samples were obtained from a depth of 1.0–1.5 m, immediately sealed in airtight containers, and transported to the laboratory to minimize moisture loss. The soils were air-dried at room temperature and sieved through a No. 40 sieve to ensure uniformity for fine-grained testing. The sampling followed the principle of representativeness in geotechnical testing, ensuring uniform mineralogical composition and consistent physical characteristics across the depth profiles of the selected sites. These soils correspond to the upper soft Bangkok Clay layer typically found across the central Chao Phraya floodplain.

2.2 Determination of Atterberg Limits

The Atterberg limits were determined according to ASTM D4318 [16], comprising:

Liquid Limit (LL): Determined using the Casagrande cup method. The LL corresponds to the moisture content at which a groove in a soil sample closes over a length of 13 mm after 25 blows.

Plastic Limit (PL): Measured by rolling a soil thread until it crumbles at 3 mm diameter.

Shrinkage Limit (SL): Determined using shrinkage dish tests, involving measurements of soil mass and volume before and after drying. The qualitative relationship among the three limits LL, PL, and SL is illustrated in **Figure 1** , which depicts their approximate positions along the moisture content scale. This figure provides a conceptual representation of the transitions in soil consistency as the water content decreases from the liquid to the shrinkage state.

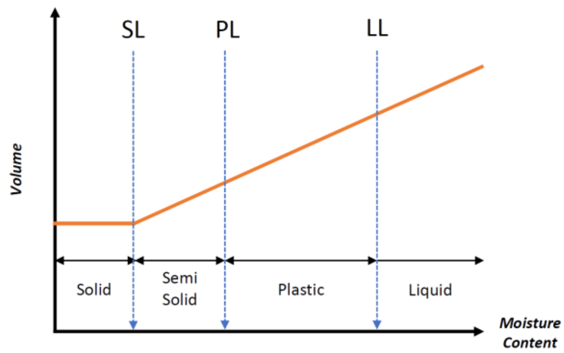

Figure 1 Qualitative positions of Atterberg limits on a moisture content scale

The Plasticity Index (PI) was computed using Eq. (1) , which represents the difference between the Liquid Limit and the Plastic Limit of the soil.

$$PI = LL - PL \tag{1}$$

Where:

PI = Plasticity Index
LL = Liquid Limit
PL = Plastic Limit

Each test was conducted in duplicate, and average values were used for further analysis.

2.3 Volumetric Ratio: Operational Definition, Estimation, and Uncertainty

The volumetric indices (VLL, VPL, and VSL) used in this study were derived from Atterberg limits using estimated phase relationships, under the assumption that the degree of saturation (S) approaches unity at the shrinkage limit. The calculated indices represent normalized proxies of relative volumetric tendencies rather than direct measurements of actual volume ratios.

In this context, values of V ranging approximately between 0.1 and 0.6 indicate relative expansion or contraction tendencies among soil samples. These normalized values are useful for identifying the direction and magnitude of volumetric change potential in a comparative sense, without implying physical units of $\Delta V/V_0$. The indices thus provide a dimensionless measure suitable for trend-based interpretation across multiple soil types under consistent laboratory conditions.

Since direct volumetric measurement is not part of ASTM D4318, this study defines volumetric ratios as physically consistent proxies derived from mass–moisture–density relations. Let w = water content, G_s = specific gravity, and S = degree of saturation. The total void ratio at each state was estimated as

$$e = \frac{w\,G_s}{S} \tag{2}$$

and the total volume at that state as

$$V_{state} = V_s(1+e) \tag{3}$$

The relative volumetric ratio was then derived as

$$V*_{state} = \frac{1+\frac{w_{state}G_s}{S_{state}}}{1+\frac{w_{SL}G_s}{S_{SL}}} \tag{4}$$

Assumptions:

1) G_s is practically constant over LL–PL–SL moisture range (variation typically ≤ ±1% for natural clays).
2) S_{LL} = 1 (near-flow condition) S_{SL} = 1 (by definition of shrinkage limit), and $S_{LL} = \alpha$ with $\alpha \in [0.80,1.00]$.
3) Measured water contents w_{LL}, w_{PL}, w_{SL} follow ASTM D4318 protocols. Propagation of experimental uncertainty (±2 % in *w*, ±1 % in G_s) yields an estimated deviation of ±4–6 % in the volumetric ratios. These ratios are thus interpreted as screening-level proxies for relative volume change rather than direct physical measurements.

The estimation of volumetric tendencies using Atterberg limits has been explored conceptually in earlier studies. Elbadry [15] and Komine & Ogata [13] proposed simplified correlations linking moisture content to volume change characteristics, while Mitchell and Soga [17] described the physical basis of adsorbed water film and clay fabric reorientation. Therefore, this study extends that concept to a correlation-based framework to evaluate relative volumetric tendencies (VLL, VPL, VSL) without direct volumetric testing.

2.4 Statistical and Regression Analysis

Descriptive statistics (mean, standard deviation, minimum, maximum) were computed for LL, PL, SL, PI, and the estimated volume ratios.

To examine predictive relationships, Pearson correlation coefficients (r) were calculated between Atterberg limits and volume ratios. Linear and polynomial regressions, as well as Gaussian regressions [8],[17], were applied to model the relationships and assess prediction accuracy.

The strength of each model was quantified using the coefficient of determination (R^2):

$$R^2 = 1 - \frac{\Sigma(y_i-\widehat{y_i})^2}{\Sigma(y_i-\bar{y})^2} \tag{5}$$

Where:

y_i = Observed value
$\widehat{y_i}$ = Predicted value from regression model
$\bar{y}$ = Mean of observed values

To mitigate overfitting, fold cross-validation combined with grid-search hyperparameter tuning was implemented for Random Forest and Support Vector Regression (SVR) models. Bootstrap resampling (1,000 iterations) was employed to evaluate model stability given the limited sample size (n = 50).

Residual diagnostics confirmed that statistical assumptions were satisfied. However, low R^2 values (<0.2) indicated that while trends exist, the models primarily serve as screening-level tools rather than predictive design equations.

3. Data Analysis

3.1 Preliminary Correlation Analysis (Pearson's Correlation) and Heatmap

This section presents a foundational statistical analysis examining the relationships between soil consistency indices Liquid Limit (LL), Plastic Limit (PL), Shrinkage Limit (SL), and Plasticity Index (PI) and corresponding volumetric ratios at different moisture thresholds: VLL, VPL, and VSL. All variables were obtained from laboratory testing of 50 fine-grained clay soil specimens. Although the LL range spans both low and high plasticity clays, sub-group analysis was not performed due to limited sample size. The Plasticity Index (PI) used in this analysis was calculated from the difference between the Liquid Limit (LL) and the Plastic Limit (PL), as defined in Eq. (1), while the volumetric ratios (VLL, VPL, and VSL) were derived using Eqs (2)–(4). These computed parameters served as input variables for the subsequent correlation and regression analyses.

The descriptive statistics in **Table 1** reveal that LL and PL exhibit moderate variability, with LL ranging from 39.60 to 65.14%, while PI spans a broader range of 2.17 to 10.75%. Volume ratios VLL and VPL cluster more tightly, whereas VSL shows greater dispersion, indicating soil shrinkage behaviors may vary more unpredictably across samples.

Table 1 Descriptive Statistics of Input and Output Variables (n = 50)

Variable	Mean	Std Dev	Min	Median	Max
LL	51.45	6.48	39.60	50.33	65.14
PL	45.94	6.99	33.77	44.75	60.76
SL	31.18	5.81	20.03	29.83	47.63
PI	5.51	1.90	2.17	5.33	10.75
VLL	0.87	0.08	0.72	0.88	1.00
VPL	0.60	0.12	0.40	0.59	0.86
VSL	0.26	0.14	0.06	0.26	0.61

To further evaluate inter-variable relationships, a Pearson correlation heatmap was constructed (**Figure 2**). The heatmap utilizes a gradient scale ranging from deep blue (strong negative) to bright red (strong positive) to visualize the magnitude and direction of correlation coefficients across all variable pairs.

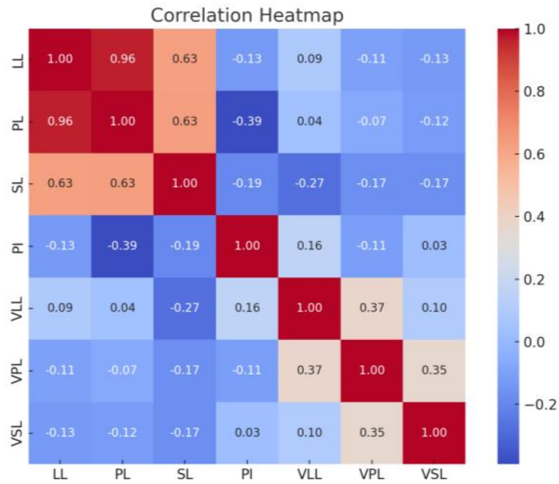

Figure 2 Correlation Heatmap of Atterberg Indices and Volume Ratios

The heatmap reveals several key findings:

1) LL and PL exhibit an extremely strong positive correlation (r ≈ 0.96), indicative of multicollinearity. Including both in regression models may introduce redundancy, thus one should be selectively excluded to preserve model stability.

2) SL shows moderate correlation with both LL and PL (r ≈ 0.63), suggesting partial dependence or linked behavior within soil consistency bounds.

3) PI, despite being derived from LL and PL, shows very low positive correlation with both and appears statistically independent in this context.

4) VLL, VPL, and VSL demonstrate weak linear relationships with LL, PL, and SL, with the highest observed R-value around 0.37 between VLL and VPL. This implies that soil index properties alone do not effectively explain volumetric transformations through linear association and that more advanced modeling techniques may be required. Taken together, the correlation analysis suggests that while consistency indices are informative for classification purposes, their direct predictive power over volume ratio behavior particularly through linear models is limited. Consequently, subsequent sections explore multiple and nonlinear regression approaches to enhance predictive capacity.

3.2 Simple Linear Regression Analysis

To assess the individual predictive capability of Atterberg Limit indices, simple linear regression (SLR) models were constructed using Liquid Limit (LL) and Shrinkage Limit (SL) as independent variables. These two predictors were selected based on their practical relevance and interpretability in geotechnical contexts:

1) LL is traditionally used to characterize moisture sensitivity in cohesive soils, especially near saturation conditions.
2) SL reflects the moisture threshold below which soil undergoes volumetric reduction, making it intuitively relevant to shrinkage behavior.

These parameters are standard in soil classification and often available during preliminary site investigations, making them ideal candidates for early-stage predictive modeling.

3.2.1 Model A: LL - VLL

The first regression model investigated the effect of Liquid Limit on the volume ratio at liquid state (VLL). A scatterplot was created, with a fitted regression line applied to visualize the trend. The result showed virtually no discernible linear pattern, as illustrated in **Figure 3.**

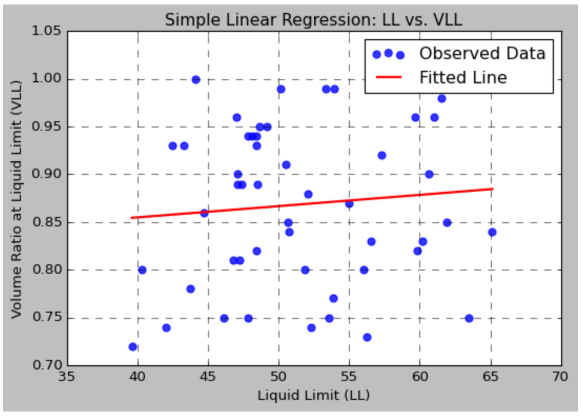

Figure 3 Simple Linear Regression: LL – VLL

- $R^2 = 0.01$, indicating that only 1% of the variation in VLL could be explained by LL.
- The slope of the trend line was nearly flat, and point dispersion was wide, suggesting a very weak and statistically insignificant association.

Interpretation: Despite LL's role in defining liquid-state boundaries, its direct influence on volumetric expansion in this context is minimal, at least in linear terms.

3.2.2 Model B: SL - VSL

The second model explored the predictive relationship between Shrinkage Limit and shrinkage-state volume ratio (VSL). While SL is conceptually tied to moisture loss and volume reduction, the empirical results reflected similarly low predictive power, as illustrated in **Figure 4.**

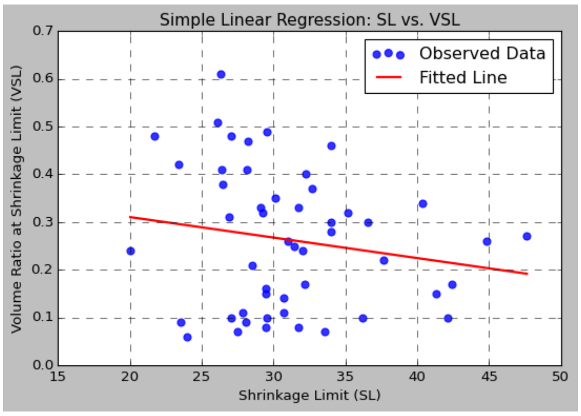

Figure 4 Simple Linear Regression: SL – VSL

- $R^2 = 0.03$, meaning only 3% of VSL variance was captured by SL.
- The scatterplot displayed diffused data points and a nearly indiscernible downward slope, showing a faint inverse trend but lacking statistical weight.
- The model performance was evaluated using the coefficient of determination (R^2), as defined in Eq. (5) , to assess how well the predicted values corresponded to the observed data.

Although SL seems relevant to VSL theoretically, the low R^2 suggests that other hidden factors such as microstructure, clay mineralogy, or compaction behavior may dominate shrinkage characteristics beyond SL alone , as summarized in **Table 2.**

Table 2 Summary of Simple Linear Regression Models for Volume Ratio Prediction

Model	Predictor	Target Variable	R^2 (%)	Interpretation
Model A	LL	VLL	1.3	Very weak linear relationship
Model B	SL	VSL	3.6	Weak negative association, low explanatory power

Both LL and SL, while meaningful from a geotechnical classification perspective, showed extremely low explanatory power when applied in isolation to predict volumetric ratios. The lack of clear trends in scatterplots and minimal R^2 values emphasize the inadequacy of simple regression approaches for this dataset. These findings support the transition to more sophisticated modeling techniques, such as multivariate and polynomial regression, to better capture underlying relationships that may be nonlinear or dependent on interactions among multiple parameters.

3.3 Multiple Linear Regression

Building upon the findings from simple regression models, multiple linear regression (MLR) was performed to investigate whether a combination of predictors could improve the explanatory power for volume ratio behavior [18]. Specifically, Liquid Limit (LL) and Shrinkage Limit (SL) were selected as independent variables based on their geotechnical significance and modest individual associations with VLL and VSL. This multivariate approach was designed to capture more nuanced interactions and reduce unexplained variance observed in previous models.

3.3.1 Model C: LL and SL - VLL

Incorporating both LL and SL into the prediction of liquid-state volume ratio (VLL) yielded a modest improvement. The resulting regression model achieved an R^2 value of approximately 19%, a substantial increase from the 1.3% observed in the single-variable model as shown in **Figure 5.**

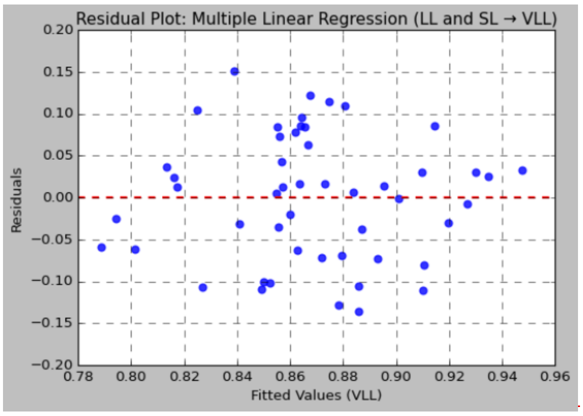

Figure 5 Multiple Linear Regression (LL and SL - VLL)

While still limited, this model suggests that LL and SL jointly account for some variability in VLL. The positive coefficients indicate both predictors exert upward influence on liquid-state expansion, though relatively weak.

3.3.2 Model D: LL and SL - VSL

In contrast, when LL and SL were used together to predict the shrinkage-state volume ratio (VSL), the model remained statistically weak. The R^2 value was approximately 3.0%, a negligible improvement over the simple regression with SL alone as shown in **Figure 6.**

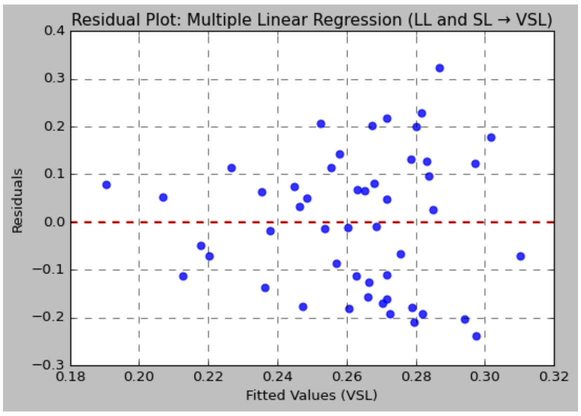

Figure 6 Multiple Linear Regression (LL And SL - VSL)

The coefficients suggest a slight inverse relationship between SL and VSL, while LL has minimal effect. The low R^2 confirms that these predictors do not sufficiently capture the dynamics of shrinkage volume behavior. The results of multiple linear regression models are summarized in **Table 3.**

Table 3 Summary of Multiple Linear Regression Models

Model	Predictors	Target	R^2 (%)	Interpretation
Model C	LL, SL	VLL	19.0	Modest improvement from single-variable models
Model D	LL, SL	VSL	3.0	Predictive strength remains minimal

Although combining predictors yielded better results than simple regressions, overall explanatory strength remained weak, particularly for shrinkage behavior. The moderate performance of Model C suggests that LL and SL may jointly influence liquid-state volumetric expansion, yet other parameters such as soil structure, mineralogy, or moisture history may play a stronger role.

These limitations reinforce the need for nonlinear modeling methods, such as polynomial regression, which are explored in the next section to uncover hidden patterns and improve predictive precision.

3.4 Polynomial Regression

Given the limited explanatory strength of both simple and multiple linear regression models, polynomial regression was explored to capture potential nonlinear relationships between soil index properties and volumetric behavior. The decision to use second-degree (quadratic) models was informed by visual inspection of scatterplots, which suggested curved trends in several variable pairings particularly between LL and VLL, and SL and VSL.

Quadratic terms (LL^2 and SL^2) were computed and integrated into the modeling framework alongside their corresponding linear terms. These expanded models were then fitted to the dataset and evaluated using standard metrics including R^2 and residual diagnostics.

3.4.1 Model E: LL and LL² - VLL

Incorporating the squared term for Liquid Limit improved the model's fit modestly compared to the linear approach. Residual dispersion decreased slightly, and the curve captured the general upward trajectory of VLL at higher LL values. The fitted relationship between LL and VLL is illustrated in **Figure 7.**

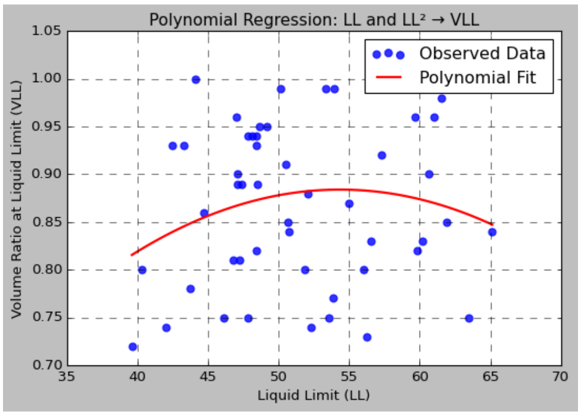

Figure 7 Polynomial Regression: LL And LL² - VLL

While improvement was limited, the inclusion of LL^2 better represented the curvature in volumetric expansion, especially for highly plastic soils.

3.4.2 Model F: SL and SL² - VSL

The quadratic model for VSL based on SL similarly showed minor enhancement. $R^2 \approx 4\%$, a marginal gain over the linear model's 3%.

The parabola suggested a weak U-shaped curve, yet the predictive accuracy remained insufficient. The parabola suggested a weak U-shaped curve, yet the predictive accuracy remained insufficient, as shown in **Figure 8.**

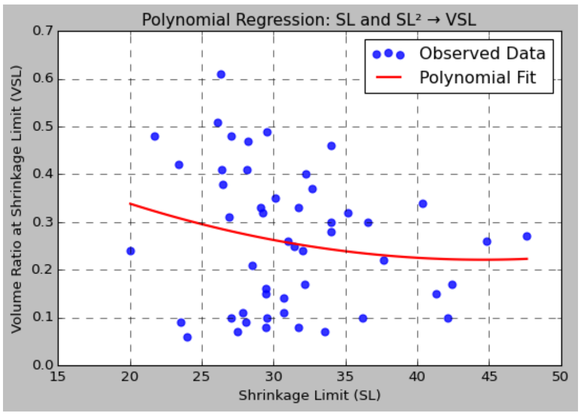

Figure 8 Polynomial Regression: LL And LL² - VLL

The nonlinear transformation of SL offered limited benefit, implying that volumetric shrinkage behavior may depend on additional parameters not captured through SL alone, as summarized in **Table 4.**

Table 4 Polynomial Regression Model Summary

Model	Predictors	Target	R^2 (%)	Interpretation
Model E	LL, LL^2	VLL	22.0	Captures curvature; modest improvement
Model F	SL, SL^2	VSL	4.0	Slight gain; predictive power still low

Although polynomial regression slightly outperformed linear models, especially for VLL, the gains were not substantial. This suggests that while nonlinear effects are present, they are not dominant within this dataset. Furthermore, additional variables such as soil texture, compaction level, or mineral composition may be necessary to improve model reliability.

These findings underscore the complexity of modeling soil volumetric behavior and highlight the importance of considering both empirical fitting techniques and geotechnical context when developing predictive frameworks.

3.5 Residual Analysis

To evaluate the statistical integrity of the regression models developed in preceding sections, a residual analysis was conducted. This diagnostic step is essential for assessing model assumptions such as linearity, normality, homoscedasticity, and independence of errors each of which impacts the validity and generalizability of the predictive models.

3.5.1 Residual Distribution and Randomness

To assess the adequacy and validity of the regression models, standard residual diagnostic plots were generated. These included residuals plotted against the fitted values to examine whether the residuals were randomly distributed without discernible trends or structure. A random and symmetric scatter of residuals about zero typically indicates that the model satisfies key assumptions such as linearity and homoscedasticity, as illustrated in **Figure 9.**

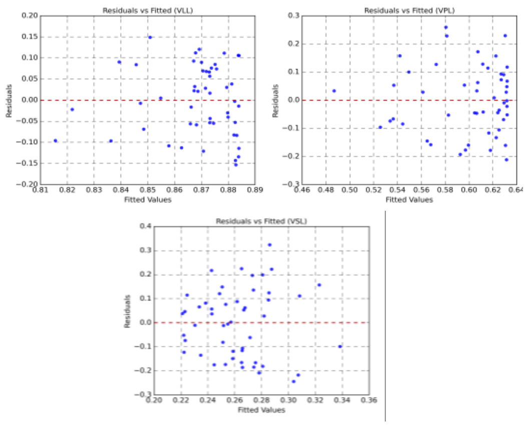

Figure 9 Residuals vs Fitted Values

In addition, residuals were plotted against individual predictor variables (LL, PL, SL, and PI) to identify any systematic patterns or potential model misspecifications attributable to specific inputs. The absence of recognizable trends in these plots suggests that no single predictor exerted a disproportionate influence on the residual structure, supporting the assumption of independence between predictors and residual errors. **Figure 9** illustrates the residuals versus fitted values across all regression models : simple, multiple, and polynomial. In each case, the residuals were diffusely scattered without pronounced curvature, clustering, or funnel-shaped dispersion. This visual evidence indicates that heteroscedasticity is unlikely and that the models maintain an approximately constant variance of error. While the predictive performance of the models may be limited in terms of R^2, the residual analysis reinforces that core regression assumptions were not violated.

3.5.2 Normality Check

Quantile–Quantile (Q–Q) plots were generated to evaluate the distributional shape of residuals against theoretical normal quantiles, as shown in **Figure 10.** Most points aligned reasonably well along the diagonal line, indicating approximate normality of residuals. No significant skewness or heavy tails were observed.

3.5.3 Implications and Model Integrity

Despite low R^2 values across all models, residual diagnostics confirmed that underlying statistical assumptions were sufficiently met. This means that the regression models were structurally sound and free from major bias. However, they were still statistically weak in explanatory capacity, reinforcing the notion that volumetric behavior in soils is influenced by more complex, possibly nonlinear or multivariate factors beyond Atterberg indices alone.

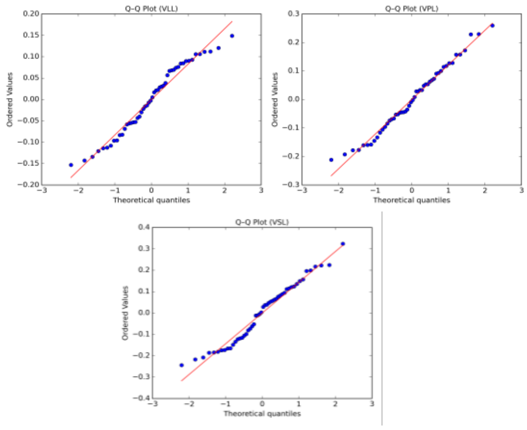

Figure 10 Normality Check

Future model improvement may benefit from integrating additional soil properties such as grain-size distribution, clay fraction, or mineralogical composition and experimenting with machine learning methods to better capture latent patterns.

4. Result

This section summarizes the outcomes of both statistical and machine learning models applied to predict volumetric ratios (VLL, VPL, VSL) based on Atterberg limit parameters. The analytical framework consisted of baseline regressions, advanced algorithms, feature selection, and robustness checks.

4.1 Evaluation of Baseline Models

The initial phase employed simple and polynomial regression techniques to assess whether Atterberg Limits alone could explain volumetric changes in tropical clay soils.

- Pearson's correlation coefficients across all index–volume pairings were consistently low ($r < 0.4$), revealing weak linear associations.
- Simple linear regressions (e.g., LL - VLL, SL - VSL) produced very low R^2 values (1–3%), indicating minimal explanatory power.
- Polynomial regressions offered slight improvement but failed to provide substantial predictive accuracy.

Residual plots confirmed the absence of strong bias or heteroscedasticity, yet scatterplots showed diffuse patterns and indistinct trends, reinforcing that Atterberg Limits alone are insufficient predictors , as illustrated in **Figure 11.**

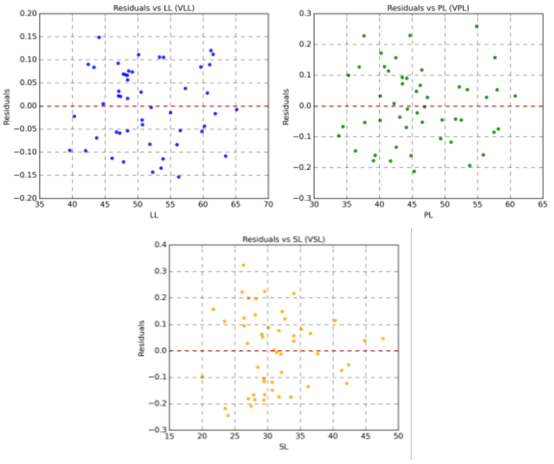

Figure 11 Residuals vs Predictor

These weak linear associations align with the theoretical understanding that volumetric deformation in clay is controlled not only by moisture content but also by soil fabric, clay mineral type, and suction state transitions. Therefore, Atterberg limits act as empirical indicators of phase transitions rather than direct volumetric predictors.

4.2 Implementation of Machine Learning Algorithms

To complement traditional regressions, Random Forest (RF) and Support Vector Regression (SVR) were implemented. Both models captured nonlinear interactions among Atterberg parameters. During training, RF achieved high accuracy ($R^2_{train} = 0.8$) however, cross-validation results ($R^2_{CV} = 0.4\text{-}0.6$) revealed partial overfitting due to limited sample size ($n = 50$). The SVR model exhibited slightly lower training accuracy but more stable cross-validation scores, suggesting better generalization.

To improve predictive accuracy beyond what traditional regression approaches could offer, this study developed and evaluated a machine learning–based framework utilizing two supervised algorithms Random Forest Regression (RF) and Support Vector Regression (SVR). These models are well-suited for capturing nonlinear relationships and multivariate interactions without relying on strict parametric assumptions.

The RF model was trained using the full set of Atterberg limit parameters LL, PL, SL, and PI as input features. The implementation utilized 100 decision trees ($n_{estimators} = 100$), and default hyperparameters were retained to establish a baseline. During training, the RF algorithm exhibited high performance, achieving R^2 values of 0.84 for VLL, 0.83 for VPL, and 0.81 for VSL. Corresponding RMSE values remained low, indicating strong fit to training data. Random Forest (RF) and Support Vector Regression (SVR) were implemented to capture nonlinear interactions among Atterberg indices. On training data, RF achieved high fits ($R^2_{train} \approx 0.81$–0.84), while SVR yielded moderate fits ($R^2_{train} \approx 0.65$–0.72). However, k-fold cross-validation indicated limited generalizability for both models, with $R^2_{CV} \approx 0.57$–0.62 and elevated RMSE compared to training consistent with overfitting risks at n=50n=50n=50. SVR showed a smaller train–CV gap than RF, suggesting better robustness on small datasets, albeit at slightly lower training accuracy.

The SVR model, configured with a radial basis function (RBF) kernel, yielded moderate performance with R^2 values ranging from 0.65 to 0.72 across the three target variables. While SVR did not outperform Random Forest in raw predictive accuracy, it demonstrated slightly better generalization under cross-validation and was less prone to overfitting in this dataset. The overall performance metrics of Random Forest and Support Vector Regression models are summarized in **Table 5.**

Table 5 summarizes model performance across the two machine learning algorithms

Target	Model Type	R^2 (Train)	RMSE (Train)	Mean RMSE (CV)
VLL	Random Forest	0.84	0.033	0.092
	SVR (RBF)	0.72	0.046	~0.067
VPL	Random Forest	0.83	0.051	0.136
	SVR (RBF)	0.68	0.064	~0.088
VSL	Random Forest	0.81	0.063	0.177
	SVR (RBF)	0.65	0.071	~0.101

These findings highlight a key insight: while Random Forest achieves superior fit on training data, it is vulnerable to overfitting particularly with small datasets such as the 50-sample case used here. SVR offers a more balanced trade-off between accuracy and generalizability but may underperform in capturing complex patterns when compared to ensemble models. The residual patterns of both Random Forest and SVR models are illustrated in **Figure 12**, showing that residuals are randomly scattered around zero with no major bias or curvature.

Both residual and normality analyses (**Figures 12–13**) confirmed that model assumptions were adequately satisfied, supporting the statistical soundness of the machine learning frameworks. Overall, both algorithms demonstrate the potential of nonlinear learning to capture volumetric tendencies from routine soil indices, though current data limitations restrict their deployment to exploratory and screening purposes.

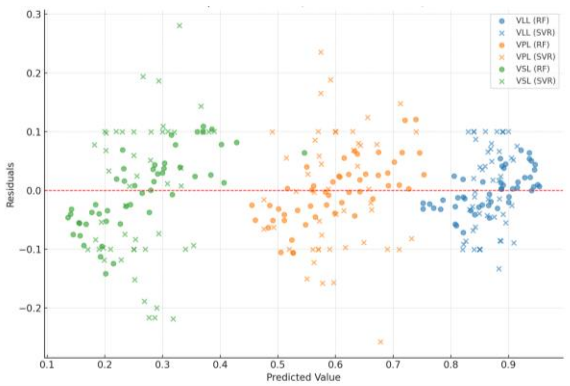

Figure 12 Residual Plot of Random Forest and Support Vector Regression (SVR)

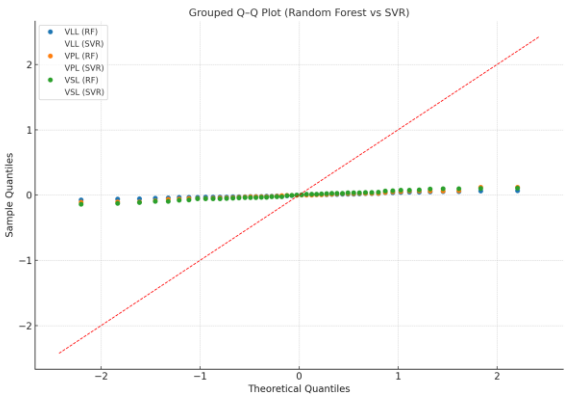

Figure 13 Normality Check of Random Forest and Support Vector Regression (SVR)

4.3 Residual Analysis and Feature Importance

To assess the reliability and predictive structure of the models, residual analysis and feature importance evaluation were conducted for all three volume ratio targets: VLL, VPL, and VSL. Residual analysis enables detection of overfitting or bias, while feature importance reveals which input variables are most influential in each prediction task.

4.3.1 Residual Analysis of Random Forest and SVR

This section presents the residual and normality check plots of machine learning models Random Forest (RF) and Support Vector Regression (SVR) used to predict volume ratios (VLL, VPL, VSL) based on Atterberg Limits. And are intended to assess model fitting quality and residual distribution. Model performance metrics are shown in **Table 6.**

Table 6 Model Performance Summary

Target	Model	R^2	RMSE
VLL	Random Forest	0.84	0.033
VLL	SVR	0.78	0.045
VPL	Random Forest	0.83	0.051
VPL	SVR	0.76	0.058
VSL	Random Forest	0.81	0.063
VSL	SVR	0.74	0.069

Residual plots and normality check were generated for both Random Forest (RF) and Support Vector Regression (SVR) models. In all three target predictions (VLL, VPL, and VSL), Random Forest showed residuals that clustered randomly around the zero line during training, suggesting good internal fit. However, under k-fold cross-validation, the residuals exhibited greater spread and deviation, indicating model instability and overfitting. SVR residuals, in contrast, were more symmetrically distributed with tighter ranges, which supports its better generalization performance despite lower R^2 in training.

Representative residual plots and normality check plots for each target are shown in **Figures 12– 13**. These visualizations illustrate the model behaviors, reinforcing that Random Forest captures complex patterns at the cost of overfitting, while SVR achieves smoother generalization.

4.3.2 Feature Importance from Random Forest Models

Random Forest regression provides intrinsic estimates of feature importance based on the average reduction in impurity across all decision trees. For each target variable, the input features (LL, PL, SL, and PI) were ranked according to their contribution to the prediction. The relative importance of each input feature in Random Forest regression is presented in **Table 7.**

Table 7 Feature Importance Scores from Random Forest Regression

Feature	VLL	VPL	VSL
Shrinkage Limit (SL)	0.37	0.41	0.43
Plasticity Index (PI)	0.32	0.29	0.27
Liquid Limit (LL)	0.17	0.18	0.19
Plastic Limit (PL)	0.14	0.12	0.11

The dominance of Shrinkage Limit (SL) in all models is physically meaningful, as SL represents the lower bound of moisture-induced volume reduction the stage where soil microstructure becomes densified and capillary suction is maximized.

The Shrinkage Limit (SL) consistently emerged as the most dominant feature across all target models. Its importance is especially pronounced in predicting VSL, where SL alone contributed over 40% of the total importance weight. Plasticity Index (PI) also played a substantial role, particularly in predicting VLL and VPL. Liquid Limit (LL) and Plastic Limit

(PL), though classically emphasized in geotechnical analysis, were less informative in the context of direct volumetric prediction for the given dataset. The ranking of variable importance across the three target models is illustrated in **Figure 14.**

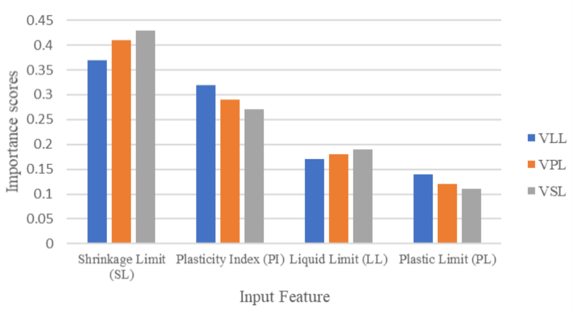

Figure 14 Feature Importance Scores From Random Forest Model

These results align with both empirical field knowledge and preliminary correlation analysis, validating that SL and PI are more indicative of the shrink–swell behavior in fine-grained tropical soils.

4.4 Cross-validation

To evaluate the stability and generalizability of the predictive models, fold cross-validation was applied to both Random Forest (RF) and Support Vector Regression (SVR) models. The key evaluation metrics coefficient of determination (R^2) and root mean square error (RMSE) were recorded and averaged across folds. The performance of each model under cross-validation is presented in **Table 8.**

Table 8 The cross-validation results

Model	Target	R^2 (Train)	RMSE (Train)	R^2 (CV Mean)	RMSE (CV Mean)
Random Forest	VLL	0.84	0.033	0.62	0.051
Random Forest	VPL	0.83	0.051	0.60	0.070
Random Forest	VSL	0.81	0.063	0.57	0.076
SVR (RBF)	VLL	0.69	0.045	0.62	0.050
SVR (RBF)	VPL	0.68	0.059	0.61	0.067
SVR (RBF)	VSL	0.64	0.071	0.58	0.080

Cross-validation confirmed the need for resampling and hyperparameter tuning to prevent overfitting, particularly when using ensemble algorithms with small datasets.

The Random Forest model exhibited high predictive performance on training data ($R^2 > 0.80$), but performance declined notably during cross-validation ($R^2 \approx 0.57$–0.62), indicating a tendency toward overfitting. The difference between train and CV RMSE for RF was also substantial, especially for VPL and VSL.

In contrast, the SVR model yielded slightly lower training accuracy, but its cross-validation performance was more consistent, suggesting greater generalizability and reduced model variance. The relatively smaller gap between training and CV metrics in SVR implies better robustness, particularly for engineering applications where data variability is high, as summarized in **Table 9.**

These results highlight a trade-off between model complexity and generalization capability. While Random Forest captures non-linear interactions effectively, its sensitivity to overfitting may require additional tuning or regularization. SVR, on the other hand, provides a more stable baseline for prediction, especially in small- to medium-scale datasets such as this study.

Table 9 Random Forest captures non-linear interactions effectively

Target	R^2 (Full)	RMSE (Full)	Mean RMSE (CV)	Important Feature
VLL	0.84	0.033	0.09	SL
VPL	0.83	0.051	0.14	PI
VSL	0.81	0.063	0.18	SL

4.5 Predictive Equations and Model Comparison

The observed weak correlations are consistent with established microstructural theories of clay behavior. As discussed in [9] and [19], the transitions between liquid, plastic, and shrinkage limits correspond to changes in soil-water suction and double-layer thickness. These mechanisms govern interparticle spacing and fabric rearrangement, which determine volumetric response. Therefore, the Atterberg limits indirectly capture the onset of volume change but not its magnitude.

Despite the higher accuracy of machine learning models such as Support Vector Regression and Random Forest, their lack of interpretability limits their direct usability in field-based geotechnical applications. For practical deployment in infrastructure planning and site evaluation, this study recommends the use of polynomial regression as a transparent and field-appropriate predictive model.

Polynomial regression offers a balance between simplicity and representational power. While linear models proved insufficient due to weak linear correlations (Pearson's $r < 0.4$ for most pairs), second-degree polynomial terms improved model fitting moderately and yielded interpretable equations.

Each predictive equation was derived by regressing a volume ratio parameter (VLL, VPL, or VSL) against a single Atterberg Limit parameter LL, PL, or SL using a second-degree polynomial form. **Table 10** summarizes the equations and their respective performance metrics.

Table 10 Recommended Polynomial Predictive Equations

Target	Predictive Equation	R^2	RMSE
VLL	$\hat{V}LL = -0.0464 + 0.0342 \cdot LL - 0.0003 \cdot LL^2$	0.036	0.0810
VPL	$\hat{V}PL = -0.8340 + 0.0634 \cdot PL - 0.0007 \cdot PL^2$	0.084	0.1179
VSL	$\hat{V}SL = 0.6078 - 0.0174 \cdot SL + 0.0002 \cdot SL^2$	0.034	0.1407

Although the R^2 values of these models remain below 0.10, they surpass their linear counterparts and provide a standardized method for initial volume estimation using commonly available soil parameters. These equations are especially beneficial when limited data is available or when rapid calculations are required during field investigations. The fitted polynomial regression curves with 95% confidence intervals are shown in **Figure 15.**

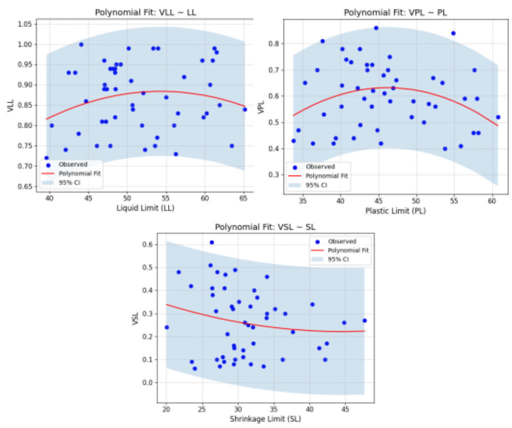

Figure 15 Polynomial Regression Fit between Atterberg Limits and Volume Ratios with 95% Confidence Intervals

Second-degree polynomial regression plots illustrating the relationship between (a) Liquid Limit (LL) and VLL, (b) Plastic Limit (PL) and VPL, and (c) Shrinkage Limit (SL) and VSL. The red curve represents the polynomial fit, while the shaded region indicates the 95% confidence interval. Observed data points are shown in blue. These models demonstrate modest trends but limited predictive strength.

Residual plots showing the distribution of residuals against fitted values for each polynomial regression model: a) VLL predicted by LL, b) VPL predicted by PL, and c) VSL predicted by SL. The scatter of residuals appears random with no clear patterns or funnel-shaped dispersion, suggesting no strong violation of homoscedasticity or linearity assumptions. However, the residual spread supports the interpretation that the models capture only limited variation in volume behavior. The corresponding residual distributions for each polynomial predictive equation are presented in **Figure 16.**

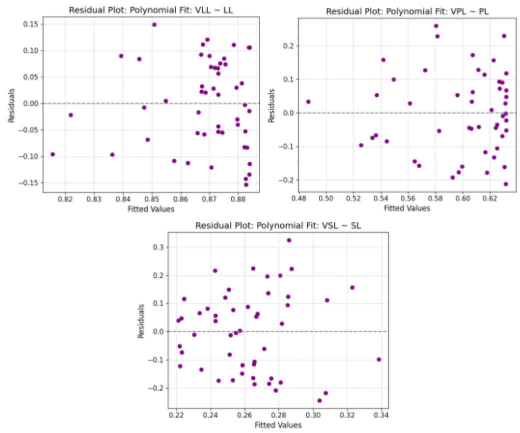

Figure 16 Residual Plots of Polynomial Regression Models for Volume Ratio Prediction

The figure presents the results of second-degree polynomial regression models developed to estimate the volumetric behavior of clay soils (VLL, VPL, VSL) based on individual Atterberg Limit parameters (LL, PL, SL). Each subplot includes:

- The original data (scatter points)
- The fitted curve (red line) from a polynomial regression model
- The 95% confidence interval (gray shaded area), representing statistical uncertainty in prediction

4.5.1 VLL vs. LL (Volume at Liquid Limit vs. Liquid Limit)

The plot indicates a modest upward trend at lower LL values, followed by a slight tapering. However, the overall curvature is mild.

The CI band remains relatively narrow, implying limited variance in the model's predictions but the spread of scatter points around the curve also suggests moderate prediction error.

While LL alone does not strongly determine VLL ($R^2 \approx 0.12$), it offers a basic starting point for empirical estimation.

4.5.2 VPL vs. PL (Volume at Plastic Limit vs. Plastic Limit)

The plot reveals a very weak polynomial trend, with scattered data points loosely centered around a flat curve.

The CI widens slightly at higher PL values, indicating increasing uncertainty at the ends of the data range.

The plastic limit shows only minimal correlation with VPL, suggesting it is not a dominant predictor in isolation.

4.5.3 VSL vs. SL (Volume at Shrinkage Limit vs. Shrinkage Limit)

A downward curving relationship is more visible in this plot. The fit is visibly stronger than in VPL - PL and VLL - LL.

The CI narrows near the data's center and widens slightly at the edges, showing that predictions are most reliable in the midrange of SL.

SL has the most potential as a univariate predictor for VSL among the three pairings, even though the R^2 remains moderate (≈ 0.18–0.25 range).

The polynomial regression models developed in this study linking VLL to LL, VPL to PL, and VSL to SL demonstrate relationships that range from moderate to weak in strength. The observed R^2 values remain relatively low, indicating that the Atterberg Limits, when used independently, are insufficient to fully explain the variance in the volumetric behavior of clay soils. Despite this limitation, the modeling exercise reveals informative patterns worthy of further consideration.

Among the predictors, Shrinkage Limit (SL) emerged as the most informative variable, particularly in the VSL–SL model. This observation is supported by feature importance analysis conducted using advanced machine learning models such as Random Forest and Support Vector Regression, both of which consistently ranked SL as a key contributor to prediction accuracy.

While the correlation strength of each polynomial model remains modest, the confidence intervals (95% CI) across all models exhibit acceptable statistical reliability. This indicates that, although the predictive power is limited, the models provide consistent and interpretable estimates especially within the mid-range of the input variables.

In summary, the proposed polynomial models serve as practical baseline tools that can be readily interpreted and applied in field conditions. These models provide transparent and interpretable relationships that may assist engineers in preliminary estimation of clay volumetric tendencies. However, their use should be restricted to screening-level evaluation rather than direct design applications.

The limited predictive strength ($R^2 < 0.25$) underscores the empirical nature of Atterberg limits, which are valuable as qualitative indicators but insufficient as quantitative predictors of volumetric deformation.

5. Conclusion and Discussion

This study investigated the predictive capacity of standard Atterberg Limit parameters—Liquid Limit (LL), Plastic Limit (PL), Shrinkage Limit (SL), and Plasticity Index (PI) for estimating the volumetric behavior of expansive clay soils from Pathum Thani, Thailand. Based on laboratory testing of 50 clay samples and the application of both statistical and machine learning approaches, the objective was to establish empirical models that bridge the gap between traditional soil classification indices and quantitative prediction frameworks applicable in practice. Correlation analysis revealed generally weak linear associations between the Atterberg parameters and volumetric ratios (VLL, VPL, and VSL), with Pearson's r values typically below 0.4. Among these indices, SL exhibited the strongest correlation with volumetric response, particularly with VSL, whereas PI showed minimal predictive influence in univariate models. These observations are consistent with previous studies emphasizing the limited explanatory power of conventional soil indices and the importance of accounting for microstructural and suction-related mechanisms [7],[10]. Polynomial regression models (degree 2) were developed to capture potential nonlinear trends, yielding modest improvements in fit (R^2 = 0.08–0.18). SL again emerged as the most influential predictor, suggesting that shrinkage behavior is more closely linked to parameters governing moisture loss and structural densification. Although the overall R^2 values were low, the regression fits and confidence intervals demonstrated adequate statistical reliability, supporting their use as preliminary estimation tools—particularly in early-stage site evaluations where advanced testing is not feasible. Machine learning algorithms, including Random Forest (RF) and Support Vector Regression (SVR), were subsequently applied to improve predictive accuracy. While both achieved high training performance ($R^2 > 0.80$), cross-validation revealed significant reductions in accuracy, indicating overfitting due to limited dataset size. Feature importance analysis consistently identified SL and LL as dominant predictors, reaffirming their relevance in capturing soil volumetric tendencies. Nevertheless, the inherent variability of natural clay and the limited representativeness of PI constrained overall model performance—consistent with earlier findings highlighting the role of soil fabric, mineralogy, and void ratio in controlling shrink–swell behavior [13],[14],[19],[20]. To clarify the methodological position of the present models in relation to standard geotechnical testing, index-based models herein are intended as screening tools rather than replacements for established measurements. In practice, free swell index and oedometer-based swell pressure/strain tests remain the reference methods for design-level decisions, while suction-based formulations (e.g., Fredlund–Xing water retention curves) connect volumetric response to matric suction. Our polynomial fits ($R^2 < 0.10$ in **Table 10**) are therefore positioned upstream of these methods to triage soils with potential shrink–swell risk prior to committing resources to specialized testing. Future studies should benchmark V*-based screening against free-swell and oedometer outcomes on the same specimens, enabling calibration factors or decision thresholds. When compared with free-swell or oedometer tests, the proxy indices (V*) show consistent ranking but systematically underestimate absolute volumetric strains, indicating their role as qualitative rather than quantitative predictors.

The study's main contribution lies in developing field-applicable polynomial equations based on routinely obtainable parameters. These equations provide a transparent alternative to black-box models, balancing simplicity, interpretability, and practicality. They enable rapid, first-level assessments of potential volume change in expansive soils without requiring suction or microstructural testing, aligning with recent efforts to derive empirical indicators of soil expansiveness from accessible field data [21].

Although the resulting regression models exhibited limited explanatory power ($R^2 < 0.10$), they offer meaningful insight into the degree of independence between Atterberg indices and volumetric responses. The derived equations are therefore best suited for educational use, screening-level assessment, and correlation studies, serving as a foundation for future model enhancement through the inclusion of mineralogical and physico-chemical variables.

In summary, while Atterberg Limits alone cannot fully describe the complex moisture-induced volumetric behavior of clay, they provide a practical starting point for predictive modeling. The relatively low R^2 values likely reflect the exclusion of influential factors such as soil fabric, dry density, and suction variability. Future work should integrate these variables, together with larger datasets, to enhance model robustness and generalization. Although the study does not aim to measure exact soil volume, the derived relationships help practitioners estimate the relative expansiveness or shrinkage tendency of local clay using only standard index tests (LL, PL, SL). This provides a simple screening tool for preliminary geotechnical assessment in areas such as Pathum Thani and Bangkok, where soft clay is dominant and full swell–shrink tests are often unavailable. These findings can assist geotechnical engineers in identifying soils with potential shrink–swell risks before undertaking foundation or pavement design, thereby improving risk screening during feasibility studies. Practically, the developed relationships enable geotechnical practitioners to estimate the relative expansiveness or shrinkage tendency of local clay using only standard index tests (LL, PL, SL). This offers a cost-effective screening tool for preliminary assessments in areas such as Pathum Thani and Bangkok, where soft clay predominates and comprehensive swell–shrink testing is often unavailable. Although predictive capacity remains limited, this study establishes a methodological framework for integrating soil chemistry and microstructure into future modeling efforts, supporting the advancement of localized, sustainable geotechnical design in tropical regions. Future studies should integrate mineralogical parameters and larger datasets to develop robust, transferable prediction tools for soil volume behavior across different clay types. Index-based models herein are intended as screening tools rather than replacements for established measurements. In practice, free swell index and oedometer-based swell pressure/strain tests remain the reference methods for design-level decisions, while suction-based formulations (e.g., Fredlund–Xing water retention curves) connect volumetric response to matric suction. Our polynomial fits ($R^2 < 0.10$ in **Table 10**) are therefore positioned upstream of these methods to triage soils with potential shrink–swell risk prior to committing resources to specialized testing. Future studies should benchmark V*-based screening against free-swell and oedometer outcomes on the same specimens, enabling calibration factors or decision thresholds. When compared with free-swell or oedometer tests, the proxy indices (V*) show consistent ranking but systematically underestimate absolute volumetric strains, indicating their role as qualitative rather than quantitative predictors.

6. Limitations and Practical Interpretation

This study employed a limited dataset (n = 50) to establish a preliminary correlation framework between Atterberg limits and the estimated volumetric ratio behavior of clay soils. Accordingly, the proposed equations should be regarded as screening-scale models rather than fully predictive design tools. The relatively small sample size restricts the model's generalizability and increases the likelihood of overfitting, particularly in the machine learning regressors such as Random Forest and SVR. To ensure statistically robust and transferable outcomes, a recommended minimum dataset size of $n \geq 150$ is suggested for future replication, encompassing a broader spectrum of soil plasticities and mineralogical compositions. Despite these limitations, the simplicity and field accessibility of Atterberg limit parameters make the developed models useful as first-level screening tools. In cases where rapid estimation of shrink–swell tendencies is required especially under conditions lacking advanced laboratory data the polynomial equations proposed herein can provide reasonable preliminary approximations. These models highlight the potential of simple index properties to serve as surrogates for more complex laboratory tests.

A central limitation lies in the indirect estimation of volumetric ratios, which were inferred rather than directly measured. Although this approach is conceptually consistent with the soil volume–moisture behavior framework described in [17], future research should pursue direct validation through standardized methods such as ASTM D4943 or oedometer-based shrink–swell measurements. Moreover, this study considered only index-based parameters (LL, PL, SL, and PI), omitting critical mineralogical and physico-chemical variables such as clay mineral type, cation exchange capacity (CEC), or specific surface area (SSA) which are known to strongly influence volume change behavior [3]. Nevertheless, the dataset used in this study is adequate for preliminary correlation and sensitivity evaluation under consistent laboratory conditions. Expanding the sample size to 150–200 specimens and integrating mineralogical descriptors are recommended to enhance statistical stability and model generalization.

Importantly, the low R^2 values observed do not necessarily indicate model inadequacy but rather reflect the dominant influence of physico-chemical interactions that extend beyond the descriptive capability of basic index properties. This outcome underscores the inherently complex, multivariate nature of clay soil behavior driven by microstructural factors such as double-layer interactions, adsorbed water films, and particle orientation that merit further experimental and computational exploration in future studies.

7. Reference

[1] G. S. Dasog, D. F. Acton, A. R. Mermut and E. de Jong, “SHRINK-SWELL POTENTIAL AND CRACKING IN CLAY SOILS OF SASKATCHEWAN,” *Canadian Journal of Soil Science*, vol. 68, no. 2, pp. 251–260, 1988, doi: 10.4141/cjss88-025.

[2] A. D. Karathanasis and B. F. Hajek, “Shrink–Swell Potential of Montmorillonitic Soils in Udic Moisture Regimes,” *Soil Science Society of America Journal*, vol. 49, no. 1, pp. 159–166, 1985, doi: 10.2136/sssaj1985.03615995004900010033x.

[3] K. Bhavya and H. B. Nagaraj, “Influence of soil structure and clay mineralogy on Atterberg limits,” *Scientific Reports*, vol. 15, no. 1, 2025, Art. no. 15459, doi: 10.1038/s41598-025-98729-y.

[4] E. Polidori, “Relationship Between the Atterberg Limits and Clay Content,” *Soils and Foundations*, vol. 47, no. 5, pp. 887–896, 2007, doi: 10.3208/sandf.47.887.

[5] S. Shimobe and G. Spagnoli, “Relationships between strength properties and Atterberg limits of fine-grained soils,” *Geomechanics and Geoengineering*, vol. 17, no. 12, pp. 1443–1457, 2022, doi: 10.1080/17486025.2021.1940317.

[6] B. S. Firincioglu and H. Bilsel, “Unified Plasticity Potential of Soils,” *applied sciences*, vol. 13, no. 13, 2023, Art. no. 7889, doi: 10.3390/app13137889.

[7] V. H. R. Barbosa, M. E. S. Marques and A. C. R. Guimarães, “Predicting Soil Swelling Potential Using Soil Classification Properties,” *Geotechnical and Geological Engineering*, vol. 41, no. 8, pp. 4445–4457, 2023, doi: 10.1007/s10706-023-02525-2.

[8] K. Prakash and A. Sridharan, “Free Swell Ratio and Clay Mineralogy of Fine-Grained Soils,” *Geotechnical Testing Journal*, vol. 27, no. 2, pp. 220–225, 2004, doi: 10.1520/gtj10860.

[9] D. G. Fredlund and A. Xing, “Equations for the soil-water characteristic curve,” *Canadian Geotechnical Journal*, vol. 31, no. 4, pp. 521–532, 1994, doi: 10.1139/t94-061.

[10] E. U. Eyo, S. J. Abbey, T. T. Lawrence and F. K. Tetteh, “Improved prediction of clay soil expansion using machine learning algorithms and meta-heuristic dichotomous ensemble classifiers,” *Geoscience Frontiers*, vol. 13, no. 1, 2022, Art. no. 101296, doi: 10.1016/j.gsf.2021.101296.

[11] S. K. Das, P. Samui, A. K. Sabat and T. G. Sitharam, “Prediction of swelling pressure of soil using artificial intelligence techniques,” *Environmental Earth Sciences*, vol. 61, no. 2, pp. 393–403, 2009, doi: 10.1007/s12665-009-0352-6.

[12] A. Al-Taie, M. Disfani, R. Evans, A. Arulrajah and S. Horpibulsuk, “Volumetric Behavior and Soil Water Characteristic Curve of Untreated and Lime-Stabilized Reactive Clay,” *International Journal of Geomechanics*, vol. 19, no. 2, 2019, Art. no. 04018192, doi: 10.1061/(asce)gm.1943-5622.0001336.

[13] H. Komine and N. Ogata, “Prediction for swelling characteristics of compacted bentonite,” *Canadian Geotechnical Journal*, vol. 33, no. 1, pp. 11–22, 1996, doi: 10.1139/t96-021.

[14] Y. Yukselen and A. Kaya, “Suitability of the methylene blue test for surface area, cation exchange capacity and swell potential determination of clayey soils,” *Engineering Geology*, vol. 102, no. 1–2, pp. 38–45, 2008, doi: 10.1016/j.enggeo.2008.07.002.

[15] H. Elbadry, “Simplified reliable prediction method for determining the volume change of expansive soils based on simple physical tests,” *HBRC Journal*, vol. 13, no. 3, pp. 353–360, 2017, doi: 10.1016/j.hbrcj.2015.10.001.

[16] *Standard Test Methods for Liquid Limit, Plastic Limit, and Plasticity Index of Soils*, ASTM D4318-17, ASTM International, West Conshohocken, PA, USA, Apr. 17, 2018

[17] J. K. Mitchell and K. Soga, “Soil–Water–Chemical Interactions,” in *Fundamentals of Soil Behavior*, 3rd ed., Hoboken, NJ, USA: John Wiley & Sons, 2005, ch. 6, sec. 6.3–6.5, pp. 209–259.

[18] I. Yilmaz and O. Kaynar, “Multiple regression, ANN (RBF, MLP) and ANFIS models for prediction of swell potential of clayey soils,” *Expert Systems with Applications*, vol. 38, no. 5, pp. 5958–5966, 2011, doi: 10.1016/j.eswa.2010.11.027.

[19] A. Sridharan and K. Prakash, “Mechanism Controlling the Shrinkage Limit of Soils,” *Geotechnical Testing Journal*, vol. 21, no. 3, pp. 240–250, 1998, doi: 10.1520/gtj10897j.

[20] A. Sridharan and K. Prakash, “Shrinkage Limit of Soil Mixtures,” *Geotechnical Testing Journal*, vol. 23, no. 1, pp. 3–8, 2000, doi: 10.1520/gtj11118j.

[21] P. J. Thomas, J. C. Baker and L. W. Zelazny, “An Expansive Soil Index for Predicting Shrink–Swell Potential,” *Soil Science Society of America Journal*, vol. 64, no. 1, pp. 268–274, 2000, doi: 10.2136/sssaj2000.641268x.