University of Baghdad
College of Engineering
JOURNAL OF ENGINEERING

Journal of Engineering
journal homepage: www.jcoeng.edu.iq
Volume 31 Number 10 October 2025

Statistical Prediction of Soil Atterberg Limits from Grain Size Distribution Alone Using Excel Solver

Mohammed Yaseen Abdullah

Technical College of Engineering, Duhok Polytechnic University, Duhok, Iraq

ABSTRACT

Atterberg limits play a crucial role in soil investigations due to their ability to provide fundamental insights into the mechanical behavior of soils. The primary objective of this study was to derive empirical equations for determining the liquid limit (LL) and plastic limit (PL) of soils using only soil particle size distribution data. Microsoft Excel was used for statistical analysis, applying 10 different methodologies. The methodology integrated the clay fraction (CF) (<0.002 mm) as the main independent factor in all models. The only independent variable in model M1 is CF, while model M2 includes sand and silt, in addition to clay fractions, as independent factors. In the analytical process, the number of independent variables progressively increased from model M1 to model M10, systematically enhancing the predictive capacity of the models. Using this methodology, two robust equations were developed in model M6, which are capable of determining LL and PL, with R-squared values reaching the value 1.0 and an RMSE of approximately zero; however, it is important to note that while an RMSE of zero does indicate an R-squared of 1.0, an R-squared of 1.0 does not necessarily imply a zero RMSE. The results of this study show that full grain size analysis data alone may effectively predict LL and PL, negating the need for additional physical or chemical characteristics of the soil.

Keywords: Atterberg limits, Grain size distribution, Liquid limit, Plastic limit, Clay, Sand, Silt.

1. INTRODUCTION

The liquid limit (LL) and plastic limit (PL) are fundamental to engineering classification systems that define fine soils and play significant roles in distinguishing silt (M) and clay (CF). Furthermore, LL, PL, and plasticity index (PI) of soils are widely utilized, either independently or in conjunction with other soil properties, to correlate with engineering performance according to **(ASTM D4318, 2017)**.

All portions of soil particles affect soil properties because the rocks become the origins of soil particles, and all the processes, including the origin of soils, weathering, transportation, and deposition, make different types of soil in terms of grain size, shape, mineral

*Corresponding author
Peer review under the responsibility of University of Baghdad.
https://doi.org/10.31026/j.eng.2025.10.02

Article received: 30/01/2025
Article revised: 18/06/2025
Article accepted: 15/08/2025
Article published: 01/10/2025

composition, etc., and the grain decrease in size **(Ishibashi and Hazarika, 2015)**. In soil mechanics, LL and PL are basic properties of fine-grained soils, and the GSD curve directly provides information on the distribution of particle sizes within the soil sample, including the percentages of Gravel(G), Sand(S), Silt (M), and Clay Fraction (CF). Soil plasticity and GSD are essential factors in classifying soils based on standard classification systems, such as the Unified Soil Classification System (USCS) or the AASHTO system, which are based on the particle size distribution, according to **(ASTM D2487, 2017; AASHTO M145, 2024)**.

The literature review highlights two key analytical areas: the first focuses on deriving Atterberg limits, namely, LL and PL, based on soil grading parameters and particle size distribution. The second area critically assesses the predictive strength and capabilities of GSD in modeling diverse geotechnical soil behaviors, emphasizing its empirical correlations with fundamental engineering properties such as shear strength, compressibility, and hydraulic conductivity.

Several studies have focused on methods for estimating the liquid limit (LL) and plastic limit PL) based on soil grading parameters and particle size distribution. For example **(Jong et al., 1990)** used the stepwise linear regression method and demonstrated that the LL, PL, and plasticity index (PI) of the samples were most strongly associated with clay fraction (CF) content. Sand (S) and silt (M) were not utilized because sand was closely linked to the clay fraction, and both sand and silt exhibited no significant correlation with Atterberg properties **(Jong et al., 1990)**. Additionally, analyses using an Artificial Neural Network (ANN) by **(Borowiec and Wilk, 2017)** with two input elements (CF and S) and two output elements indicated an enhancement in the prediction of soil parameters LL and PL. Further research **(Deng et al., 2017)** examined the relationship between individual parameters, where individual nonlinear regression analysis revealed a significant positive correlation between soil LL and PL with CF, organic matter, cation exchange capacity (CEC), and free iron oxide, and a notably negative correlation with sand (S).

The prediction process for LL and PL by **(Sherzoy, 2017)** found that using the Support Vector Machine (SVM) or Adaptive Neuro-Fuzzy Inference System (ANFIS) model for predictions with three inputs (S, M, and CF) yielded results closer to the true value than to the analysis using only two inputs (M and CF). According to **(Nini, 2014)** both M and CF should be used together for LL measurement, and another study by Sen and Pal **(Sen and Pal, 2014)** concluded that the plasticity of soil increases as the finer fraction (M+CF) increases. Based on ANN analysis by **(Zolfaghari et al., 2015)**, the CF and calcium carbonate equivalent are the primary factors predicting the limits of Atterberg. In addition to CF, the type of clay and size of the particles have considerable effects on the Atterberg limits. In another study, a higher correlation was found between CF content and LL than with PL when Atterberg limits were individually correlated with CF, M, and S contents **(Knadel et al., 2021)**, neither the Atterberg limits nor the PI were significantly correlated with the M content. Another study found strong positive correlations between CF and all Atterberg limits **(Van Tol et al., 2016)**, and significant negative correlations exist between S content and all Atterberg limits.

Other studies have established a relationship between plasticity limits and CF. One study **(Polidori, 2007)** revealed that CF is directly proportional to LL, PL, and PI, with only LL and CF showing a regression line with a zero intercept value. Additionally, another study **(Rehman et al., 2019)** correlated clay type and CF to LL and PL using visible–near-infrared spectroscopy (VisNIR) and found a strong relationship between them. Furthermore, **(Moradi and Ebrahimi, 2013)** indicated that both LL and PL are influenced by the quantity

and type of CF in the soil and that high CF content leads to high LL and PL. Another study **(Polidori, 2003)** found a nonlinear regression relationship between LL and PL with the CF content.

Further investigations by **(Kayabali, 2011)** were conducted to examine the impact of the grain size of material passing sieve #200 (0.075 mm) versus sieve #40 (0.425 mm) on LL and PL; It was discovered that the LL and PL of the material passing through a #200 sieve were approximately 10 to 20 percent higher than those of the material passing through a #40 sieve for the same soil, resulting in higher LL and PL values owing to the increase in the specific surface caused by the decrease in grain size. Attempts were made by **(Seybold et al., 2008)** to relate the plasticity limits with CF and other soil parameters such as CEC, bulk density, free iron oxide, linear extensibility percent, and organic carbon, the results revealed that CF and CEC were the most highly correlated and important independent variables in predicting both LL and PI, compared to the other remaining variables.

Another study **(Dolinar and Škrabl, 2013)** explored how soil composition influences LL and PL and found that the Atterberg limits are mainly dependent on the size and proportion of (CF) clay minerals. In a separate study conducted by Khalaf and Issa **(Khalaf and Issa, 2021)**, which focused on five selected soil sites in Babylon, Iraq, the researchers aimed to explore the relationship between the soil's physical properties. They emphasized that particle size distribution is a key factor affecting physical characteristics such as plasticity and forms the essential foundation for understanding the link between these properties.

In a recent study conducted by **(Nawaz et al., 2022)**, a computational model leveraging artificial intelligence (AI) methodologies was developed to systematically integrate parameters S, CF, and M to more accurately model the consistency characteristics of fine-grained soils. The investigation revealed that the liquid limit (LL), a critical index property of soil plasticity, is predominantly influenced by particle-size distribution, particularly the proportional contributions of silt, clay fraction, and sand within the soil matrix. Grain Size Distribution (GSD) and particle arrangement play a crucial role in determining soil properties such as porosity, permeability, shear strength, compressibility, and bearing capacity. The impacts of the GSD and particle arrangement on the porosity and pore size distribution, which in turn affect permeability, **(Ubani et al., 2018)** predicted and found a strong positive association between permeability and grain size data and distribution, also another study by **(Díaz et al., 2022)** demonstrated that the permeability decreases with decreasing porosity in compacting aggregates, and the Kozeny equation was modified by taking the grain size grading into account.

GSD significantly affects other soil properties. For example, cohesion and the friction angle of the soil, which are influenced by the distribution of particle sizes, are essential for determining slope stability and foundation construction. The results of the study by **(Basson, 2023)** demonstrate that shear strength increases with particle size range; **(Van Hoa et al., 2021)** show that strength parameters increase as soil grain size increases; **(Ahmad and Uchimura, 2023)** demonstrate that soil moisture content, dry density, and grain size, influence soil shear strength; and **(Abdou, 2008)** proved that the angle of internal friction increases as each type of (S) particle size (fine, medium, or coarse) increases.

Soil with a high proportion of fine grains is highly compressible, whereas soil with a high proportion of coarse grains is less compressible. The compressibility of a material is influenced by the GSD and can be accurately modeled using a GSD index **(Sun et al., 2015)**. Additionally, **(Yong et al., 2017)** highlights the crucial role of GSD in the behavior of soils and granular media. Furthermore, **(He et al., 2021)** confirms that the characteristics of

gradation, including the shape of the GSD curve, have a significant impact on the deformation properties of materials with varying particle sizes.

The GSD of a soil can provide useful information about its California Bearing Ratio (CBR) value. Generally, soils with a well-graded GSD, meaning they contain a wide range of particle sizes from coarse to fine, tend to have higher CBR values compared to poorly-graded soils with a more uniform particle size distribution have lower CBR values. In geotechnical engineering, artificial intelligence (AI) approaches have been used by **(Taskiran, 2010)** to evaluate the (CBR) of soils using grain gradation (including CF + M, S, and G fractions), plasticity values, and compaction factors as input data. These procedures have proven to be capable of properly predicting the CBR of soils. Furthermore **(Duque et al., 2020)** discovered a substantial correlation between granular soils' CBR and their GSD. Using conventional laboratory data on the GSD of compacted soil, a method has been devised by **(Stephenson et al., 1976)** to predict the strength characteristics of the soil. However, the various characteristics present in natural soils, such as grain size, shape, and surface texture, all of which have a significant impact on the strength of the soil, are mostly responsible for the scatter in the recorded CBR values around the regression line. The natural fluctuations in soil qualities that occur also add to the observed uncertainty in the relationship between CBR and GSD.

A study by **(Abdullah, 2024)** found that Both Maximum dry density (MDD) and optimum moisture content (OMC) are based only on GSD data without any other chemical or physical properties of soil, with R^2 approximately about 1.0 in this case. **Fig. 1,** based on Abdullah's equations **(Abdullah, 2024)**, depicts the MDD and OMC in relation to typical GSD curves of soil. The novelty of the present study lies in its comprehensive approach to modeling the relationship between detailed GSD parameters-including clay fraction, sand and silt fractions, and specific passing diameters-and Atterberg limits using advanced regression techniques. By incorporating a broader range of predictors and refining the model structure, this research aims to improve prediction accuracy and provide a more robust tool for soil behavior characterization, addressing limitations in existing literature.

This study explores the correlation between soil fractions and soil plasticity, utilizing Grain Size Distribution (GSD) curves of soil obtained from sieve and hydrometer analyses to estimate and gain a better understanding of the soil plasticity.

MES is particularly suitable for geotechnical engineering applications because of its user-friendly environment, which allows for efficient data entry, manipulation, and visualization through built-in functions and customizable spreadsheets. This accessibility not only streamlines the analysis process but also enables rapid computation and graphical representation of complex soil parameters, directly from laboratory or field data. Furthermore, MES facilitates the minimization of the sums of squares of differences and maximization of the Pearson correlation coefficient (R^2) for specific datasets, making it highly effective for determining relationships between retaining fractions and corresponding soil consistency limits. Its widespread use in geotechnical practice and education further supports its reliability and versatility for such analyses. In this study, the Microsoft Excel Software (MES) was used to prepare and analyze the data with an easy-to-use interface without having to write code.

To derive the theoretical model equation for determining the LL and PL from the grain size data, more than 10 models were analyzed by MES using a generalized reduced gradient (GRG), and the model with the highest R^2 value and the smallest RMSE value was selected in the final step. While previous studies have explored the correlation between grain size

distribution (GSD) and Atterberg limits, most have relied on limited GSD parameters or simple linear models, often resulting in moderate predictive accuracy.

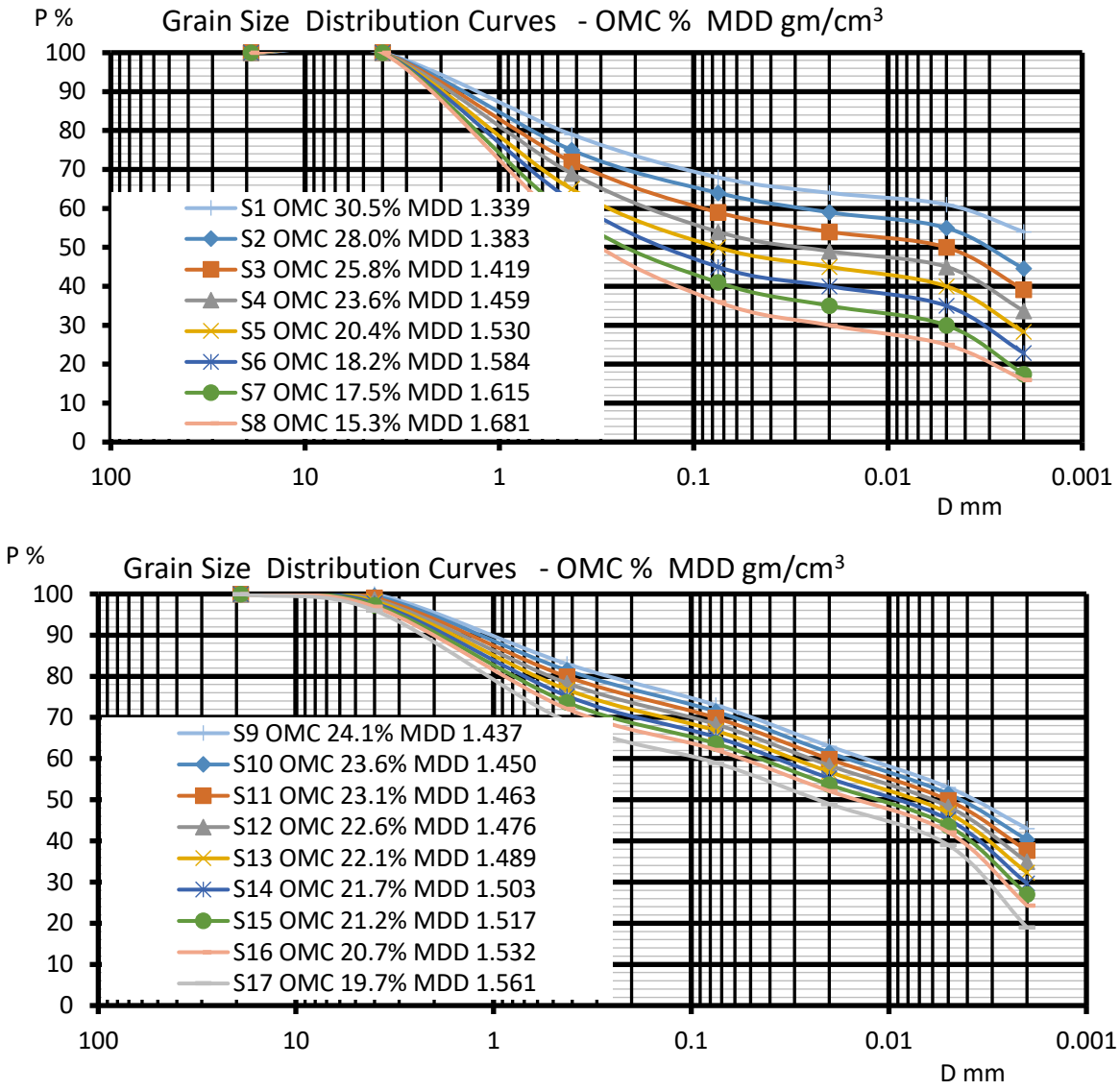

Figure 1. MDD and OMC in relation to typical GSD curves for different soils, after Abdullah's equations **(Abdullah, 2024).**

Moreover, the complex interactions between different grain size fractions and soil plasticity remain inadequately quantified, especially across diverse soil types. This gap restricts the reliable use of GSD data alone for precise soil classification and engineering assessments. Previous studies have attempted to model the correlation between grain size distribution (GSD) and Atterberg limits, with varying degrees of success. GSD parameters such as clay fraction (CF), sand and silt fractions (S, M), or specific grain size passing diameters have been used as predictors in regression models to estimate Atterberg limits like liquid limit (LL) and plastic limit (PL). These studies indicate that the correlation between GSD and Atterberg limits reflects the influence of soil texture on soil plasticity and related characteristics. However, these efforts also highlight the complexity of accurately predicting soil consistency parameters from grain size data alone. Building on this foundation, the present study aims

to develop regression models that better capture the relationship between GSD and soil plasticity, thereby improving classification accuracy and engineering assessments.

2. MATERIALS AND EXPERIMENTAL WORKS

Soil samples were collected from various locations in Mosul City, Iraq. Twenty samples were taken from a depth of 1.00 meters below the ground surface in the following Mosul city sectors: Qadisiyah 1, Qadisiyah 2, Yarmouk, Wahda, Hadbaa, Hawi Kanisa, and Shuhada. This selection ensures that the study captures spatial variability in soil properties, thereby providing a comprehensive understanding of the conditions relevant to geotechnical engineering applications in the city. Soil samples were subjected to specific gravity (Gs), particle-size analysis (gradation), and Atterberg limit tests in accordance with ASTM Specification guidelines. The specific gravity of the soil particles was determined using a pycnometer, which allows for accurate measurement of the mass and volume of soil solids. Particle-size distribution was assessed through sieve analysis for coarse fractions, utilizing a standard set of sieves and a mechanical shaker, and by hydrometer analysis for finer particles, which measures the rate of sedimentation in a suspension. The Atterberg limits, including the liquid limit (LL) and plastic limit (PL), were determined using the Casagrande apparatus for the percussion-cup (liquid limit) test and the rolling method for the plastic limit. These standardized laboratory procedures and apparatus ensure the reliability and reproducibility of the soil characterization results, providing essential data for geotechnical analysis and engineering design. The soil samples collected from the study area exhibit a wide range of geotechnical properties, as summarized in **Table 1**. The Liquid Limit (LL) values vary from 0.0% to 101.0%, with an average of 45.8%, while the Plastic Limit (PL) ranges from 0.0% to 50.0%, averaging 22.1%. The Plasticity Index (PI), which is a key indicator of soil plasticity and behavior, ranges from 0.0% to 51.0%, with an average value of 23.8%. Based on the PI values, the soils can be categorized into low plasticity (PI < 10%), medium plasticity (PI between 20% and 40%), and very high plasticity (PI > 40%) groups **(Das, 2013)**.

Table 1. Statistical properties of soil samples

	LL%	PL%	PI%	Gs	P#4	P#40	P#200	CF5	CF2
Min. limit	0.0	0.0	0.0	2.50	97	90.0	65.0	20.0	6.2
Max. limit	101.0	50.0	51.0	2.74	100	100	96.0	72.0	65.0
Median	50.5	22.5	25.0	2.71	100	96.8	89.5	56.0	40.5
Average	45.8	22.1	23.8	2.70	99.8	95.5	86.5	52.3	36.2
St. dev.	19.6	9.8	11.5	0.1	0.7	3.1	8.3	14.1	15.5
VAR.P	385.7	95.6	131.8	0.0	0.5	9.5	69.6	197.5	240.9

The majority of samples fall within the moderate plasticity range, indicating soils with moderate plastic behavior. Additional parameters such as specific gravity (Gs), particle size distribution-where P#4, P#40, and P#200 indicate the percentage passing each sieve-and clay fractions (CF5 and CF2 denote clay fractions with particle sizes less than 0.005 mm and 0.002 mm, respectively) further characterize the soil texture and structure, supporting a comprehensive understanding of the soil conditions in the study area.

3. STATISTICAL METRICS AND MODEL

In this research, the GRG Nonlinear engine is selected to find the solver for models. This Solving method alone, like virtually all nonlinear optimization algorithms, can find a locally optimal solution. In the "GRG," when it is making very slow progress Solver stops before finding a locally optimal solution, and their objective function is changing from one trial solution to another. This is similar to the principle of trial and error; therefore, the GRG usually needs more than one attempt to find the optimal solution.

The GRG (Generalized Reduced Gradient) algorithm in Excel Solver is used to solve smooth nonlinear optimization problems by iteratively adjusting variables to find a local optimum while satisfying constraints. It was chosen because it efficiently handles continuous variables and nonlinear relationships, offering faster convergence than the Evolutionary algorithm and greater suitability than the Simplex method, which is limited to linear problems. Overall, GRG provides an effective balance of accuracy and computational speed for this study's optimization needs. The GRG algorithm has several limitations. First, it requires a feasible starting point, which can be challenging to identify complex problems. The algorithm primarily handles smooth, differentiable nonlinear functions and may struggle or fail with non-differentiable or highly non-convex problems. It can also converge to local optima rather than the global optimum, especially in problems with multiple local minima**.**

The model's reliability was checked by both R^2 and root mean squared error (RMSE) metrics to evaluate the equation's performance. It is noted that a coefficient of determination **(Likhith et al., 2022)** with R^2=1 does not indicate zero RMSE, whereas zero RMSE does indicate R^2 = 1.0.

There is always a risk of error when estimating for more situations, so it is better to use a mathematical approach that removes the risk of bias. One of the Mathematical determinations of regression lines is the method of least squares, which is the most popular and used one **(Smith, 1986)**.

4. METHODOLOGY

In this paper, the master strategy is implemented by ten models (M1 to M10), all relying on the clay fraction (CF < 0.002 mm) as the main variable parameter, along with parameters obtained from sieve and hydrometer analyses. The models employ multiple linear regression to establish predictive equations relating Liquid Limit (LL) and Plastic Limit (PL) to grain size distribution (GSD) data and soil retention percentages. This approach allows quantifying the relationship between Atterberg limits and soil particle size characteristics.

The model strategy involves increasing the number of independent input parameters from model M1 to M10. Model M1 (with 1 input parameter) relied solely on the clay fraction (CF), which represents the percentage of soil particles smaller than 0.002 mm. Model M2 (with 3 input parameters) incorporated soil portion percentages defined as Sand (S), Silt (M), and Clay Fraction (CF). Specifically, S (Sand) refers to the coarse fraction of soil particles typically ranging from 0.075 mm to 4.75 mm, M (Silt) represents the fine fraction with particle sizes between 0.002 mm and 0.075 mm, and CF (Clay Fraction) denotes the finest particles less than 0.002 mm in diameter. Models M3 and M4 were based on six sieve sizes (#4, #8, #16, #40, #50, and #200) in addition to hydrometer parameters, which provide detailed particle size distribution data. Models M5 to M9 expanded the input to eight sieve sizes (#4, #8, #16, #30, #40, #50, #100, and #200) alongside hydrometer measurements for increased

resolution. Finally, Model M10 utilized twelve sieve sizes (#4, #8, #10, #16, #20, #30, #40, #50, #60, #100, #140, and #200), representing a comprehensive set of sieves tailored to capture the full range of grain sizes present in the soil samples.
In addition to increasing the number of sieve points in the ten models, another strategy involves gradually increasing the number of independent parameters obtained from hydrometer analysis to fifteen points in the final model (0.07, 0.06, 0.05, 0.04, 0.03, 0.02, 0.01, 0.009, 0.008, 0.007, 0.006, 0.005, 0.004, 0.003, and 0.002). All input parameter points from both sieve and hydrometer analyses are displayed and analyzed across the ten models using MES. **Table 2.** shows the total number of parameters used for each model, and the Initial RMSE of LL and PL at the first start before the solver run from MES.

Table 2. Details of number parameters for main models.

Sub-Model	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
Number of Parameters	1	3	14	16	18	20	21	22	24	29
Initial RMSE of LL	50	50	48	48	48	48	48	48	48	48
Initial RMSE of PL	24	24	23	23	23	23	23	23	23	23

5. RESULTS AND DISCUSSIONS

The initial model (M1) was formulated to derive empirical correlations for estimating the LL and PL using clay fraction (CF <0.002 mm) as the sole input parameter. Two regression equations were established under this framework: Eq. (1) for LL prediction yielded R^2 of 0.72 and RMSE of 10.75, while Eq. (2) for PL prediction produced an R^2 of 0.60 and an RMSE of 6.44. The modest predictive capacity of these equations, evidenced by the limited R^2 values and elevated RMSE magnitudes, reflects the inherent constraints of a single-parameter regression approach. This methodological limitation underscores the restricted explanatory power of CF alone in capturing the full variability of Atterberg limits, necessitating the inclusion of additional granulometric parameters to enhance model robustness.

$$LLp\% = 12360.8 * 10^{-4} * CF2 \qquad R^2=0.719, \quad M1 \qquad (1)$$
$$PLp\% = 5906.3 * 10^{-4} * CF2 \qquad R^2=0.598, \quad M1 \qquad (2)$$

Similarly, the correlation coefficient R^2 in M2 (3 input parameters S, M, and CF) very slightly improves in Eq. (3) to Eq. (4). The dependence of these equations on only three soil portions, S, M, and CF, limits the equation's ability to accurately forecast LL and PL.

$$LLp\% = (58.0S + 1036.3\,M + 11250.9\,CF2) * 10^{-4} \qquad R^2=0.722, \quad M2 \qquad (3)$$
$$PLp\% = (698.5S + 885.5\,M + 5107.7\,CF2) * 10^{-4} \qquad R^2=0.605, \quad M2 \qquad (4)$$

Fig. 2 displays the M2 equation results as well as the degree of accuracy in the relationship between the actual and predicted values of LL and PL. The distribution of points around the y=x line is random, suggesting that LL and PL cannot be strongly predicted using only the primary soil portions (S, M, CF). This is very clear from high values of RMSE=10.35 and 6.14 for LL and PL, respectively. This implies that other input parameters from GSD data should be taken into account and examined according to the research methodology of this paper.

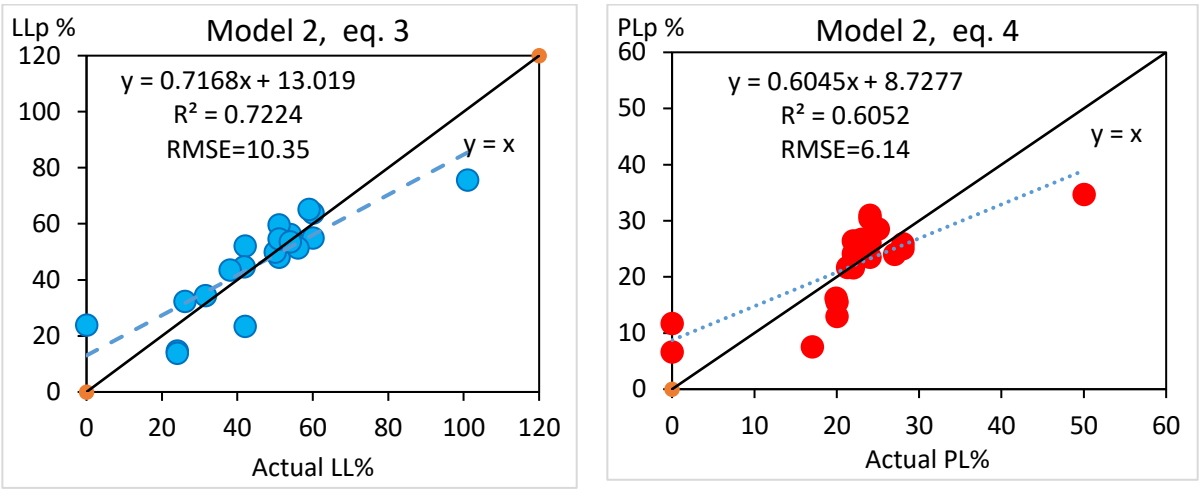

Figure 2. Model 2, Actual versus predicted LL, PL values from Eq. (3) and Eq. (4)

R^2 was utilized to assess the degree of correctness of the 10 main models. As the number of variable input points is increased from 1 to 29, the findings clearly show an improvement in regression, with the R^2 increasing and RMSE decreasing (see **Fig. 3**). Additionally, **Fig. 4** shows how adding more input points affects the R^2 coefficient's augmentation and RMSE decrease. Eq. (5) and Eq. (6) from model M6 predict LL and PL using twenty independent variables, alongside, inserted as ''a'' factor and ''b'' factor, respectively. These equations depend on the retaining (Rt) percent on each corresponding "j" Index indicating a size where the point of retaining is taken, where; $Rt_j = P_{i+1} - P_i$ and (P) is the passing from gran size distribution data. **Table 3** presents the calibrated coefficients (a-factors and b-factors) corresponding to the grain size fractions used in Eq. (5) and (6) for predicting LL and PL, respectively, within Model M6. A positive a-factor indicates that an increase in the retained percentage of the corresponding particle size fraction contributes to an increase in the predicted LL value. Conversely, a negative a-factor suggests a reducing effect on LL. Similarly, the b-factors represent the influence of each grain size fraction on PL. Notably, the trends of a-factors and b-factors are generally consistent in sign and magnitude, reflecting that the same particle size fractions tend to influence LL and PL similarly, although the magnitude of their effects varies

Also, **Table 3** breakdown highlights how the entire grain size distribution influences soil plasticity characteristics, emphasizing the importance of considering multiple size fractions rather than relying solely on one or three soil fractions, as referred to in models M1 and M2.

$$LLp\% = \left[a_0\, CF2 + \sum_{i=1}^{19} a_i Rt_j \right] * 10^{-2} \qquad a_0 = 221.0 \qquad \text{M6} \qquad (5)$$

$$PLp\% = \left[b_0\, CF2 + \sum_{i=1}^{19} b_i Rt_j \right] * 10^{-2} \qquad b_0 = 129.7 \qquad \text{M6} \qquad (6)$$

Eq. (5) and Eq. (6) demonstrate a functional relationship between LL, PL, respectively, and their GSD Besides, increasing R^2 up to 1.000 as the number of input points increases to 20 in the model M6. It has been observed that the relationships are more accurate due to RMSE decreases to 0.0064 and 0.0016 for LL and PL, respectively, referring to **Figs. 3 and 4** for a comparison between the 10 main models.

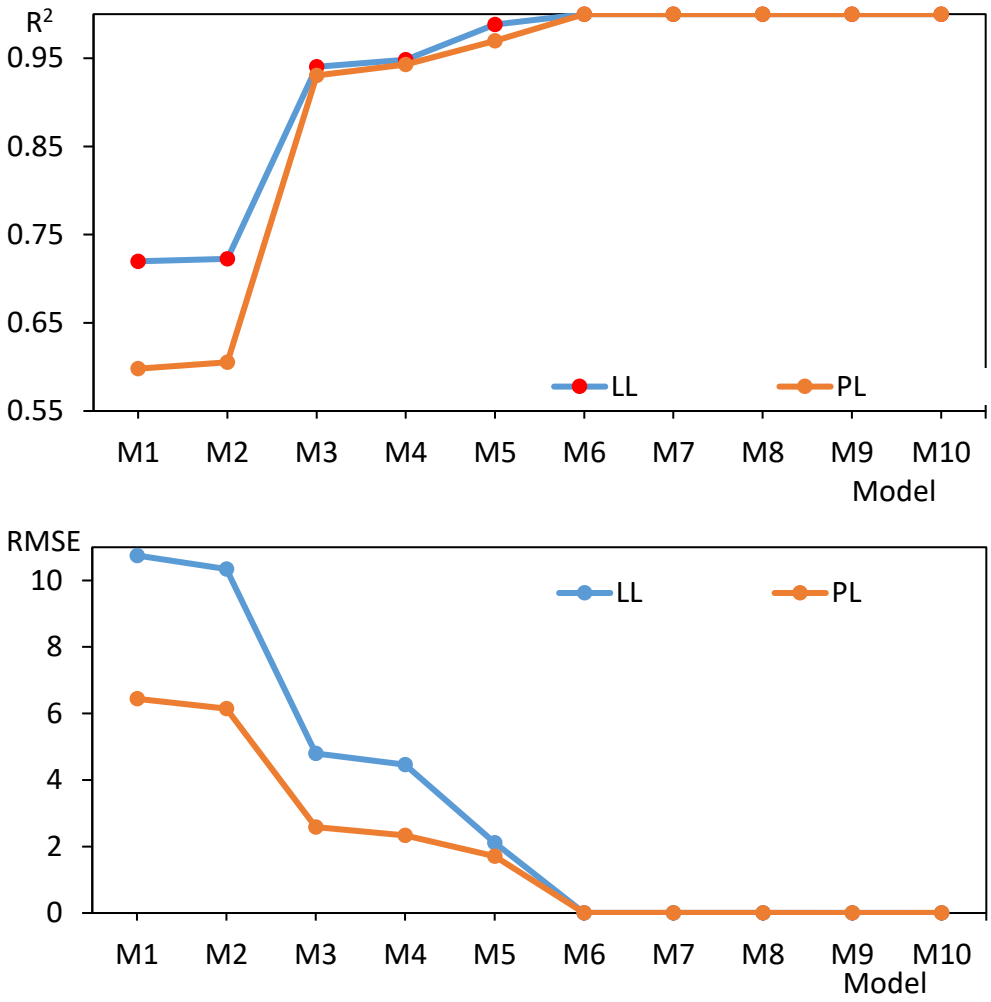

Figure 3. R^2 and RMSE values versus models.

From the 10 models in this study, it can be concluded that the LL and PL can be efficiently predicted from particle size analysis tests alone, without relying on any other chemical or physical property of the soil. This aligns exactly with the research findings **(Abdullah, 2024)**, which identified both MDD and OMC based solely on GSD data.

Fig. 5 illustrates the accuracy of Model M6 in predicting the LL directly from detailed GSD data using Eq. (5). By incorporating 20 input points, including the CF and retained percentages across 19 this model captures the complex influence of the entire soil gradation on plasticity, achieving a near-perfect correlation (R^2 = 1.000) and very low RMSE values (0.0064 for LL). This demonstrates that soil plasticity is governed not only by clay content but also by the full gradation profile, highlighting the importance of high-resolution GSD data for precise LL estimation. Compared to simpler models relying on fewer parameters as M1 (1 input parameters) and M2 (3 input parameters), Model M6 offers a significant improvement, enabling rapid and highly reliable LL predictions that can streamline geotechnical design processes, especially for projects requiring detailed soil characterization.

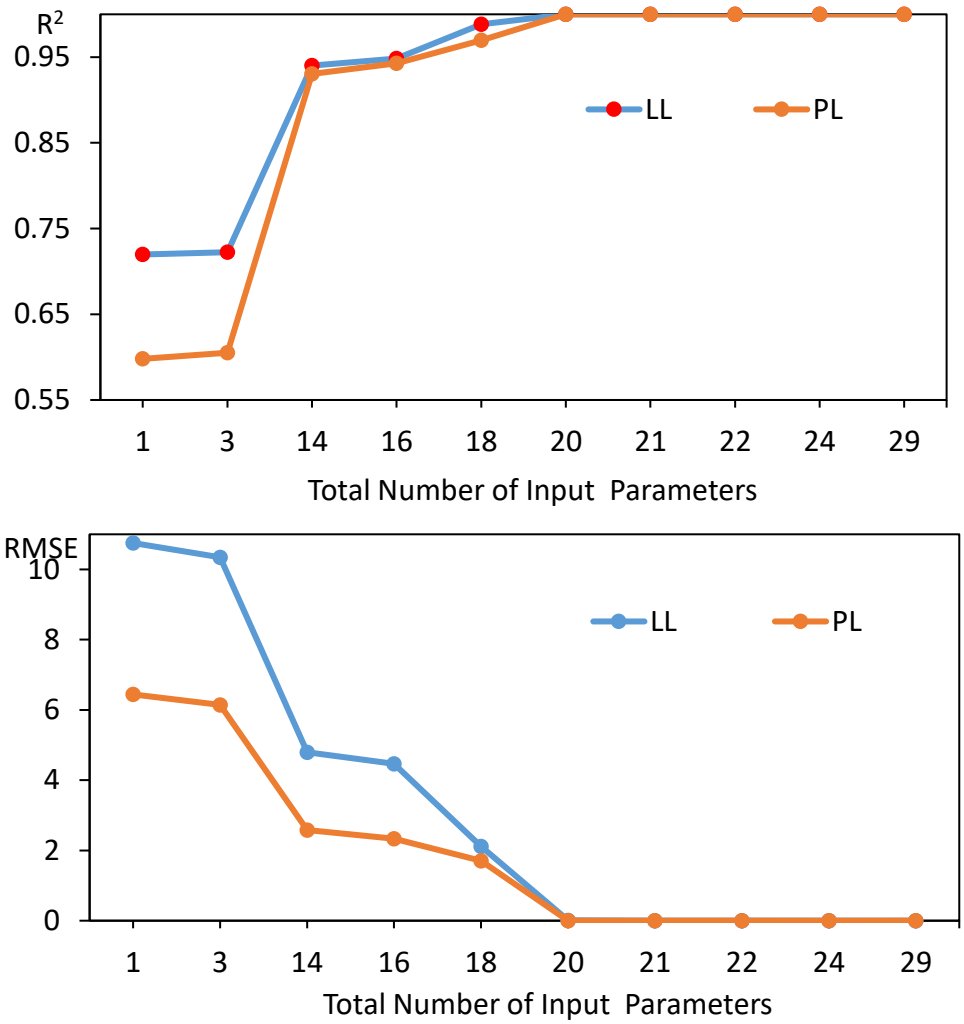

Figure 4. R^2 and RMSE values versus the total number of input Parameters.

The forecast outcomes of LL and PL from M6 by regression equations are compared to measured data from experimental tests in **Fig. 6**. The strength of Eq. (5) and (6) is indicated by the R^2 coefficient and RMSE, showing that the obtained results are very similar to the original data (experimental results).

As shown in **Fig. 6**, the proposed M6 model demonstrates an exceptional agreement between the measured and predicted values for both Liquid Limit (LL) and Plastic Limit (PL). The regression equations (Eq. 5 and Eq. 6) capture the functional relationship between soil grain size distribution (GSD) and soil consistency limits. Specifically, the LL and PL are modeled as functions of the clay fraction (CF2) and the retention percentages (Rtj) across various grain sizes, reflecting the comprehensive influence of the entire soil gradation on consistency limits

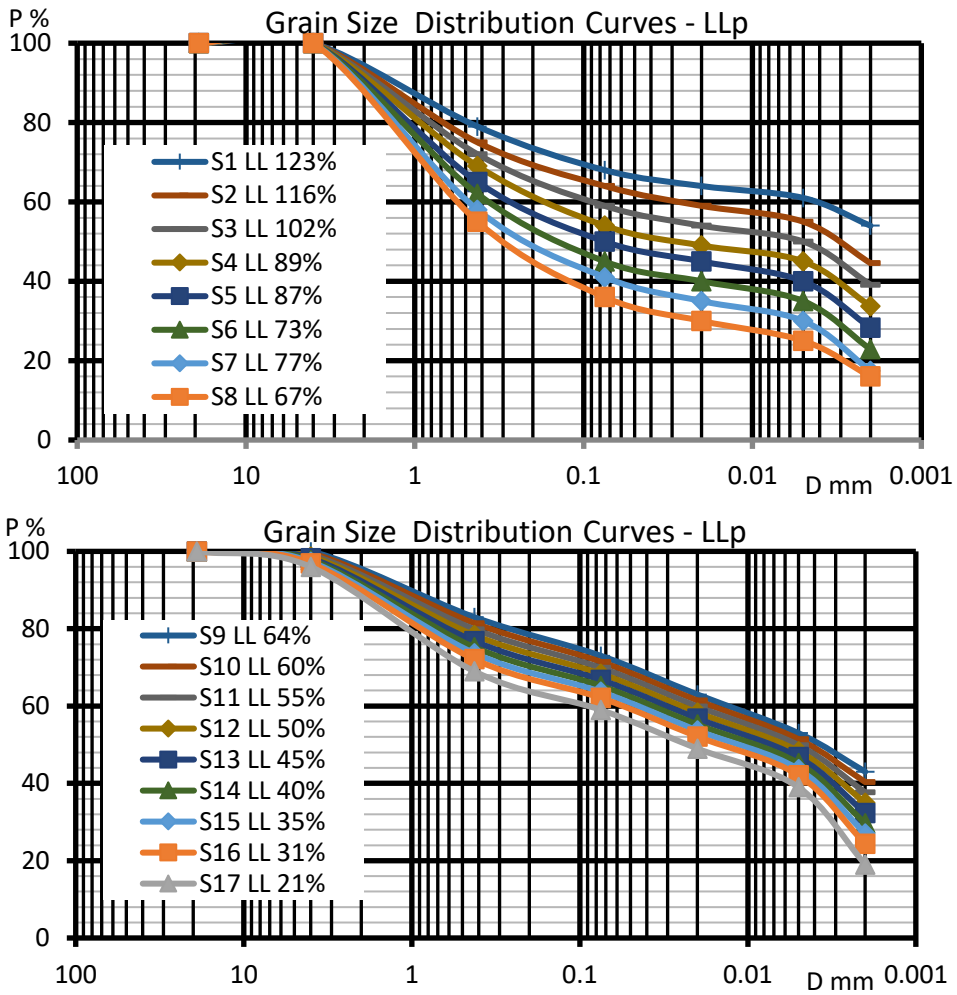

Figure 5. LLp relative to typical GSD curves for various soils, Eq. (5).

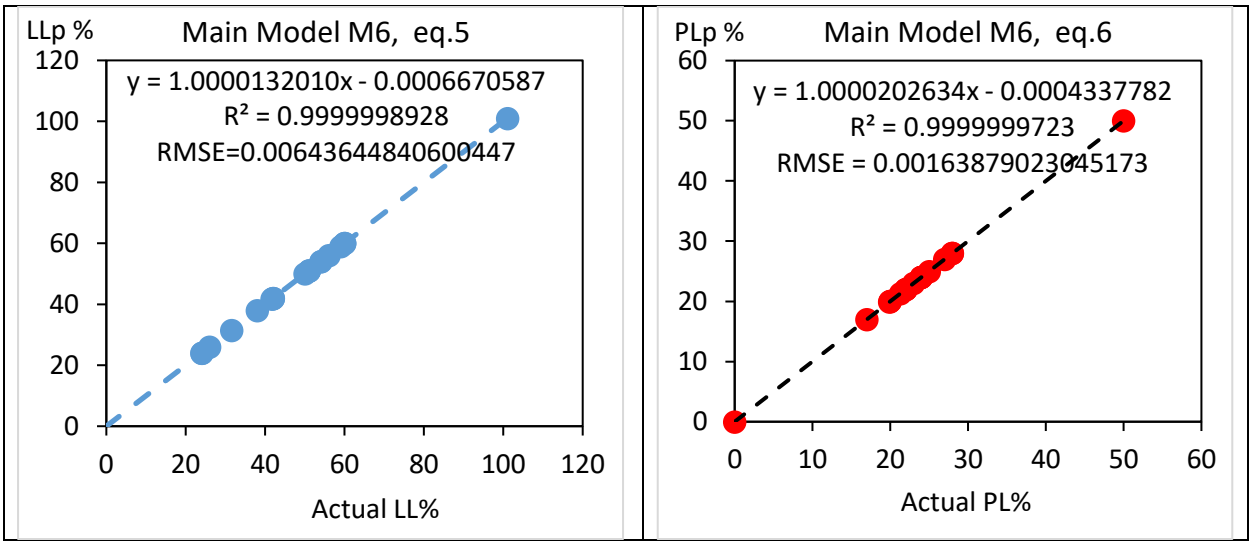

Figure 6. A comparison between the measured values and predicted values using M6 (Eq. 5 and Eq. 6).

.

Table 3. (i)th element of Eq. (5) and Eq. (6) from model M6.

Index J	Diameter (D)mm	Index i	LLp a-factor	PLp b-factor
#4	4.75	1	-3687.4	-4003.6
#8	2.36	2	4519.0	6136.5
#16	1.18	3	304.4	1020.2
#30	0.60	4	-1970.5	-1839.4
#40	0.425	5	4979.7	7447.1
#50	0.300	6	-4958.6	-7211.5
#100	0.150	7	-3.7	-1200.4
#200	0.075	8	975.4	1633.8
0.06	0.06	9	-3142.4	-3492.4
0.05	0.05	10	697.9	588.3
0.04	0.04	11	1303.0	1644.7
0.02	0.02	12	568.5	813.1
0.01	0.01	13	-571.1	-898.9
0.009	0.009	14	3393.7	2960.9
0.008	0.008	15	-2577.1	-1954.4
0.006	0.006	16	-2404.7	-1920.6
0.005	0.005	17	1904.1	1811.0
0.004	0.004	18	856.7	555.5
0.002	0.002	19	-115.0	-96.9

Sieve No. 40 (0.425 mm) is used in preparing the soil for LL and PL tests and is referenced in major standard soil testing methods, including AASHTO T89 and T90 **(AASHTO T89, 2022; ASTM D4318, 2017; BS 1377-2, 1990; IS 2720-5, 1985).** This sieve fraction excludes coarser sand-sized particles that do not exhibit plasticity, thereby focusing the tests on the fines that govern soil behavior related to plasticity. To investigate the effect of using only the material passing through the No. 40 sieve on model M6, this study modified model M6 to create a new model, named M#40, and analyzed by using MES and a GRG solver.

Models M#40LL and M#40PL were prepared and examined to predict LLp and PLp, respectively, and both R^2 and RMSE were determined. **Fig. 7** below shows the regression correlation between the actual and predicted values of LL and PL.

The RMSE of M#40LL and M#40PL was equal to 49 and 23, respectively, before running GRG in MES After running the solver tool, the RMSE of LL was equal to 6.1630. And this value comparison with RMSE (0.0064) in M6 is considered a very high value also, the R^2 decreases to 0.9016, less than R^2=1.0000 in M6, and to be more clear and more organized, see **Table 4.** for a comparison between model M6 vs M#40LL and M#40PL.

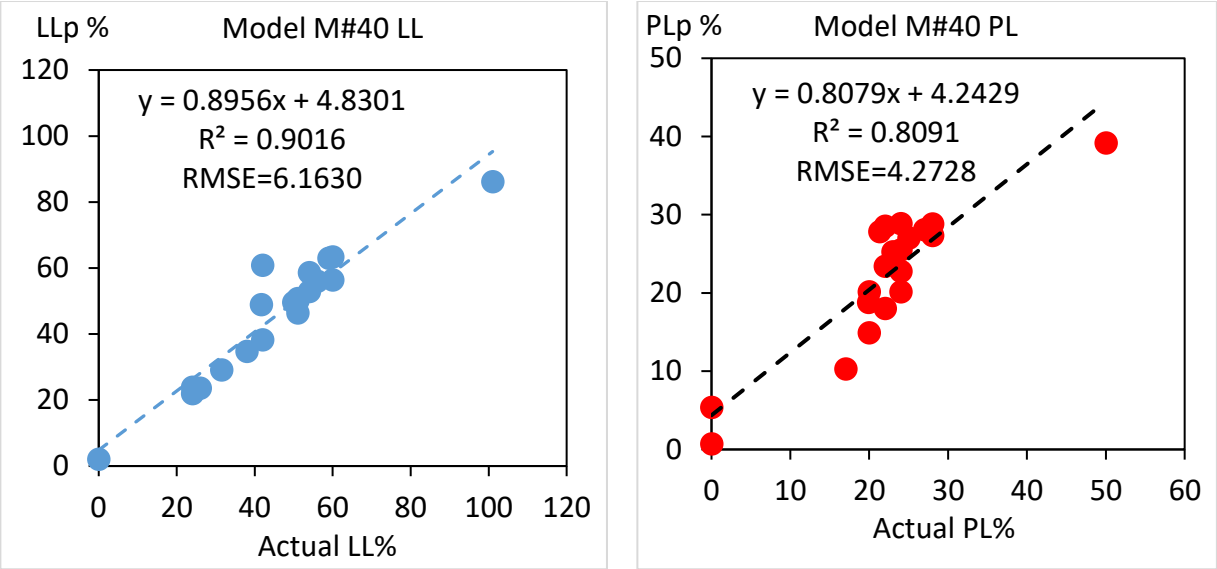

Figure 7. A comparison between the measured and predicted LL and PL values M#40.

Table 4. Comparison between two models.

Limit	LLp			Limit	PLp	
Model	**M6**	**M#40LL**		**Model**	**M6**	**M#40PL**
R^2	1.0000	0.9016		**R^2**	1.0000	0.8036
RMSE	0.0064	6.1630		**RMSE**	0.0016	4.3336
Minimum DLL	-0.011	-14.853		**Minimum DPL**	-0.003	-6.565
Maximum DLL	+0.020	+18.905		**Maximum DPL**	+0.003	+10.826

The statistical performance of the M6 model is outstanding, with R^2 values of 1.0000 for both LL and PL, and extremely low RMSE values (0.0064 for LL and 0.0016 for PL). The minimum and maximum differences between the measured and predicted values (DLL and DPL) are also very small, indicating high prediction accuracy and minimal bias.

From **Table 4**, it is clear that model M#40LL gives R^2 =0.90, which is a relatively high value but M#40LL cannot be relied upon due to the high value of RMSE=6.1630, the model gives a difference between actual and predicted LL ($DLL=LL_{actual}-LL_{predict}$) about (-14.8% to +18.9%) and this is the very high difference which cannot accept, and this conclusion fully applies to the model M#40PL (for predicting PL), also see **Table 4,** ($DPL=PL_{actual}-PL_{predict}$)**.** The M#40PL models perform worse because they show a lower R^2 (0.8036 vs. 1.0000) and much higher RMSE (4.3336 vs. 0.0016), indicating less accuracy and greater error. The wider range of DPL values (from -6.565 to +10.826) also suggests more variability and less reliable predictions compared to the M6 model. The M#40LL model performs worse due to its lower R^2 (0.9016 vs. 1.0000) and much higher RMSE (6.1630 vs. 0.0064), indicating less accuracy. The large range of DLL values (from -14.853 to +18.905) shows greater prediction variability compared to the M6 model. This likely results from the inclusion of coarser particles in the No. 40 fraction, which reduces the model's precision.

From the literature, several methods and models reporting R^2 and/or RMSE values are used for predicting LL and PL based on inputs such as clay fraction (CF), sand (S), and silt (M). The methods include:

a) Simple Linear Regression (SLR), used by **(Jong et al., 1990),** which showed a strong correlation for LL prediction with an R^2 of 0.84 but excluded sand and silt due to low correlation.
b) Adaptive Neuro-Fuzzy Inference System (ANFIS) and Support Vector Machine (SVM), applied by **(Sherzoy, 2017),** where ANFIS achieved very high accuracy (R^2 up to 0.987 for LL), outperforming SVM in both LL and PL predictions.
c) Visible and Near-Infrared Spectroscopy (VisNIR), used by **(Rehman et al., 2019)**, which reported strong correlations for both LL and PL, though exact R^2 and RMSE values were not specified.
d) Correlation and Regression analysis by **(Seybold et al., 2008)**, which yielded an R^2 of approximately 0.79 for LL.

Other studies in the literature did not report R^2 or RMSE values for their methods and models, often due to the specific focus or nature of their research. For example, Borowiec and Wilk **(Borowiec and Wilk, 2017)** concentrated on an artificial neural network (ANN) model architecture rather than detailed statistical metrics, which is common in some ANN studies. Similarly, the researchers in **(Deng et al., 2017; Knadel et al., 2021)** performed correlation and spectroscopy-based analyses that emphasized correlation coefficients or qualitative relationships without explicitly reporting R^2 or RMSE. Empirical studies such as those by **(Nini, 2014; Sen and Pal, 2014; Moradi and Ebrahimi, 2013; Khalaf and Issa, 2021)** often focused on observed trends and recommendations rather than formal predictive model performance metrics. Additionally, pedotransfer function studies **(Van Tol et al., 2016; Zolfaghari et al., 2015)** and regression analyses **(Polidori, 2003; 2007)** sometimes omitted detailed fit statistics in summaries or abstracts. Other works, including **(Kayabali, 2011; Dolinar and Škrabl, 2013)**, prioritized experimental comparisons or qualitative soil composition relationships over formal regression statistics. Lastly, **(Nawaz et al., 2022)** developed AI computational models but did not explicitly report R^2 and RMSE in their study summary. Overall, the absence of R^2 and RMSE reporting often reflects a focus on model development, qualitative analysis, or empirical observations rather than formal predictive model validation.
From previous results and **Table 4**, if the model depends on full GSD data only without taking any other physical or chemical properties as the independent variable, and the GSD data should be all taken and not missing any part of the curve data, then we can say the GSD curve looks like the DNA of a human because it is loaded by all soil properties.
The outcomes of this study confirmed a significant and strong correlation between both LL/PL and GSD curve data.

6. CONCLUSIONS

There is a functional relationship between soil grain size distribution (GSD) and soil consistency limits, such as the Liquid Limit (LL) and Plastic Limit (PL), which largely depend on the full soil gradation. Equations in model M6 effectively capture this relationship by expressing LL and PL as functions of the clay fraction (CF2) and the retention percentages (Rtj) across different grain sizes, highlighting the importance of incorporating full soil gradation in predictive modeling of soil consistency limits. These equations model the dependence of LL and PL on both the finest soil fraction and the distribution of particle sizes, thereby providing a comprehensive representation of soil consistency based on grain size data.

1. The model equations express the relationship between the Atterberg limits and the soil's grain size distribution GSD in excellent agreement because the R^2 Value is high and the RMSE is very low as the number of independent parameters increases.
2. The research models indicate that 20 parameters from sieve analysis and hydrometer analysis are enough to represent the optimal model, achieving a high R^2 value and a very low RMSE.
3. Using the complete grain size data of any soil, we can accurately predict the liquid limit (LL) and plastic limit (PL) without requiring any extra chemical or physical details.
4. Soil characteristics are greatly influenced by the distribution of grain sizes. The grain size distribution GSD curve is akin to the DNA of humans, storing all information about soil properties.
5. The models based solely on clay fraction (CF) or combined with sand (S) and silt (M) fractions (Models M1 and M2) show limited predictive accuracy for liquid limit (LL) and plastic limit (PL), with R^2 values below 0.73 and relatively high RMSE values. In comparison, models incorporating the full grain size distribution (GSD) provide significantly better predictions, as they capture the influence of all particle size fractions rather than just three broad categories. This demonstrates that relying only on CF, S, and M fractions restricts the model's ability to accurately forecast LL and PL, highlighting the importance of detailed GSD data for improved soil plasticity prediction.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

AASHTO M 145, 2024. Standard Specification for Classification of Soils and Soil-Aggregate Mixtures for Highway Construction Purposes. American Association of State Highway and Transportation Officials, Washington, D.C.: AASHTO.

AASHTO T89, 2022. Standard Method of Test for Determining the Liquid Limit of Soils, American Association of State Highway and Transportation Officials. Washington, D.C.: AASHTO.

Abdou Abd El-Naiem, M., 2008. Prediction of the angle of internal friction of sand using the grain-size distribution curve. *JES: Journal of Engineering Sciences*, 36(3), pp. 569-579. https://dx.doi.org/10.21608/jesaun.2008.116135

Abdullah, M.Y., 2024. June. Correlation of grain size distribution with the compaction parameters for fine-grained soils. In AIP Conference Proceedings. 2944(1). AIP Publishing. https://doi.org/10.1063/5.0205152

Ahmad, W., and Uchimura, T., 2023. The effect of moisture content at compaction and grain size distribution on the shear strength of unsaturated soils. *Sustainability*, 15(6), P. 5123. https://doi.org/10.3390/su15065123

ASTM D2487, 2017. Standard Practice for Classification of Soils for Engineering Purposes (Unified Soil Classification System). West Conshohocken, PA: ASTM International.

ASTM D4318, 2017. Standard Test Methods for Liquid Limit, Plastic Limit, and Plasticity Index of Soils, West Conshohocken.PA: ASTM International.

Basson, M.S., 2023. January. Effect of particle size distribution on monotonic shear strength and stress-dilatancy of coarse-grained soils. In *Geo-Congress*. https://doi.org/10.1061/9780784484678.031

Borowiec, A. and Wilk, K., 2017. Prediction of consistency parameters of fen soils by neural networks. *Computer Assisted Methods in Engineering and Science,* 21(1), pp. 67-75.

BS 1377-2, 1990. Methods of test for soils for civil engineering purposes – Part 2: Classification tests. British Standards Institution (BSI).

Das, B.M., 2013. *Principles of geotechnical engineering*. 7th ed. Boston, MA: Cengage Learning.

Deng, Y., Cai, C., Xia, D., Ding, S., Chen, J., and Wang, T., 2017. Soil Atterberg limits of different weathering profiles of the collapsing gullies in the hilly granitic region of southern China. *Solid Earth*, 8(2), pp. 499-513. https://doi.org/10.5194/se-8-499-2017

Díaz-Curiel, J., Miguel, M.J., Biosca, B. and Arévalo-Lomas, L., 2022. New granulometric expressions for estimating permeability of granular drainages. Bulletin of Engineering *Geology and the Environment*, 81(10), P. 397. https://doi.org/10.1007/s10064-022-02897-4.

Dolinar, B. and Škrabl, S., 2013. Atterberg limits in relation to other properties of fine-grained soils. *Acta Geotechnica Slovenica*, 10(2), pp. 4-13.

Duque, J., Fuentes, W., Rey, S. and Molina, E., 2020. Effect of grain size distribution on California bearing ratio (CBR) and modified proctor parameters for granular materials. *Arabian Journal for Science and Engineering*, 45(10), pp. 8231-8239. http://dx.doi.org/10.1007/s13369-020-04673-6

He, J., Liu, F., Deng, G., and Fu, P., 2021. Relationships between gradation and deformation behavior of dense granular materials: Role of high-order gradation characteristics. *International Journal for Numerical and Analytical Methods in Geomechanics*, 45(12), pp. 1791-1808. https://doi.org/10.1002/nag.3224

IS 2720-5, 1985. Methods of test for soils, Part 5: Determination of liquid and plastic limit. Bureau of Indian Standards.

Ishibashi, I. and Hazarika, H., 2015. Soil mechanics fundamentals and applications. 2nd ed. Boca Raton, FL:CRC Press, P. 9.

Jong, E.D., Acton, D.F., and Stonehouse, H.B., 1990. Estimating the Atterberg limits of southern Saskatchewan soils from texture and carbon contents. *Canadian Journal of Soil Science*, 70(4), pp. 543-554. http://dx.doi.org/10.4141/cjss90-057

Kayabali, K., 2011. Determination of consistency limits: A comparison between–# 40 and–# 200 materials. *Electronic Journal of Geotechnical Engineering*, 16(T), pp. 1547-1561.

Khalaf, M.O. and Issa, B.M., 2021. Study of the relationship between some physical properties of selected soils in Babylon, Central Iraq. *The Iraqi Geological Journal*, pp. 155-164. https://doi.org/10.46717/igj.54.2D.12Ms-2021-10-31

Knadel, M., Rehman, H.U., Pouladi, N., de Jonge, L.W., Moldrup, P., and Arthur, E., 2021. Estimating Atterberg limits of soils from reflectance spectroscopy and pedotransfer functions. *Geoderma*, 402, P. 115300. https://doi.org/10.1016/j.geoderma.2021.115300

Likhith, K.M., Bherde, V., and Balunaini, R.B.U., 2022. Prediction of california bearing ratio (CBR) of soils using AI-based techniques. *Indian Geotechnical Conference*, IGC 2022, 15th– 17th December 2022, Kochi.

Moradi, S. and Ebrahimi, E., 2013. Relationship between the percentage of clay with liquid limit, plastic limit, and plastic index in four different soil texture classes. *TJEAS Journal*, 3(8), pp. 697-702.

Nawaz, M.N., Qamar, S.U., Alshameri, B., Karam, S., Çodur, M.K., Nawaz, M.M., Riaz, M.S., and Azab, M., 2022. Study using a machine learning approach for a novel prediction model of liquid limit. *Buildings*, 12(10), P. 1551. https://doi.org/10.3390/buildings12101551

Nini, R., 2014. Effect of the silt and clay fractions on the liquid limit measurements by Atterberg cup and fall cone penetrometer. *International Journal of Geotechnical Engineering*, 8(2), pp. 239-241. https://doi.org/10.1179/1939787913Y.0000000018

Polidori E. 2007. Relationship between the Atterberg limits and clay content. *Soils and foundations*. Oct 1;47(5), pp. 887-96. https://doi.org/10.3208/sandf.47.887.

Polidori, E., 2003. Proposal for a new plasticity chart. *Geotechnique*, 53(4), pp. 397-406. https://doi.org/10.1680/geot.2003.53.4.397 .

Rehman, H.U., Knadel, M., Kayabali, K. and Arthur, E., 2019. Estimating Atterberg Limits of Fine-Grained Soils by Visible–Near-Infrared Spectroscopy. *Vadose Zone Journal*, 18(1), P. 190039. https://doi.org/10.2136/vzj2019.04.0039

Sen, B. and Pal, S.K., 2014. Index properties of soils collected from different locations and correlations of parameters. *EJGE*, 19, pp. 3443-3452.

Seybold, C.A., Elrashidi, M.A., and Engel, R.J., 2008. Linear regression models to estimate soil liquid limit and plasticity index from basic soil properties. *Soil science*, 173(1), pp. 25-34. http://dx.doi.org/10.1097/ss.0b013e318159a5e1

Sherzoy, M.M., 2017. Atterberg limits prediction comparing SVM with ANFIS model. *Journal of Geoscience, Engineering, Environment, and Technology*, 2(1), pp. 20-30. https://doi.org/10.24273/jgeet.2017.2.1.16

Smith, G.N., 1986. Probability and statistics in civil engineering. Collins professional and technical books, 244, P. 185.

Stephenson, H.K., Karrh, J.B., and Koplon, N.A., 1967. Compacted soil strength estimated from grain size distribution and soil binder analysis. *Highway Record*, 190, pp. 1-7.

Sun, Y., Xiao, Y. and Hanif, K.F., 2015. Compressibility dependence on grain size distribution and relative density in sands. *Science China Technological Sciences*, 58, pp. 443-448. https://doi.org/10.1007/S11431-015-5768-5

Taskiran, T., 2010. Prediction of California bearing ratio (CBR) of fine-grained soils by AI methods. *Advances in Engineering Software*, 41(6), pp. 886-892. https://doi.org/10.1016/j.advengsoft.2010.01.003

Ubani, C.E., Ani, G.O., and Womiloju, T.T., 2018. Permeability estimation model from grain size Sieve analysis: Data of onshore central Niger Delta. *European Journal of Engineering and Technology Research*, 3(12), pp. 119-125. https://doi.org/10.24018/ejeng.2018.3.12.503.

Van Hoa, C., Tuan, V.A., Sang, N.T., Lai, N.T., and Tiep, P.D., 2021. Effect of grain size on shear strength of coral gravel sand. In *Proceedings of the 3rd International Conference on Sustainability in Civil Engineering: ICSCE 2020*, 26-27 November, Hanoi, Vietnam, pp. 87-95. https://doi.org/10.1007/978-981-16-0053-1_11

Van Tol, J.J., Dzvene, A.R., Le Roux, P.A.L. and Schall, R., 2016. Pedotransfer functions to predict Atterberg limits for South African soils using measured and morphological properties. *Soil Use and Management*, 32(4), pp. 635-643. https://doi.org/10.1111/sum.12303

Yong, L., Chengmin, H., Baoliang, W., Xiafei, T., and Jingjing, L., 2017. A unified expression for grain size distribution of soils. *Geoderma*, 288, pp. 105-119. https://doi.org/10.1016/j.geoderma.2016.11.011

Zolfaghari, Z., Mosaddeghi, M.R., and Ayoubi, S., 2015. ANN-based pedotransfer and soil spatial prediction functions for predicting Atterberg consistency limits and indices from easily available properties at the watershed scale in western Iran. *Soil Use and Management*, 31(1), pp. 142-154. https://doi.org/10.1111/sum.12167

التنبؤ الإحصائي لِـ حدود أتربرغ، بالاعتماد فقط على توزيع الحجمي لحبيبات التربة، باستخدام أداة (سولفر) في برنامج الإكسل

محمد ياسين عبدالله

الكلية التقنية الهندسية، جامعة دهوك التقنية، دهوك، العراق

الخلاصة

تلعب حدود أتر بيرج دورًا مركزيًا في دراسات التربة نظرًا لقدرتها على تقديم رؤى أساسية حول السلوك الميكانيكي للتربة. الهدف الرئيسي من هذه الدراسة هو صياغة معادلات تجريبية لتحديد حد السائل (LL) وحد البلاستيك (PL) للتربة اعتمادًا فقط على بيانات توزيع حجم حبيبات التربة. تم استخدام برنامج مايكروسوفت إكسل للتحليل الإحصائي، مع تطبيق 10 مناهج مختلفة. شملت المنهجية دمج جزء الطين (CF) (<0.002 مم) كعامل مستقل رئيسي في جميع النماذج. في النموذج M1، كان جزء الطين هو المتغير المستقل الوحيد، بينما شمل النموذج M2 الرمل والغرين وجزء الطين كعوامل مستقلة. في العملية التحليلية، تم زيادة عدد المتغيرات المستقلة تدريجيًا من النموذج M1 إلى النموذج M10، مما أدى إلى تحسين القدرة التنبؤية للنماذج بشكل منهجي. باستخدام هذه المنهجية، تم تطوير معادلتين قويّتين في النموذج M6، قادرتين على تحديد حد السائل وحد البلاستيك بقيم معامل تحديد تصل إلى 1.0 وقيمة خطأ الجذر التربيعي الوسيط (RMSE) تقارب الصفر؛ ومع ذلك، من المهم الإشارة إلى أنه بينما يشير RMSE بقيمة صفر إلى معامل تحديد 1.0، فإن معامل التحديد 1.0 لا يعني بالضرورة أن RMSE يساوي صفرًا. تظهر نتائج هذه الدراسة أن بيانات تحليل حجم الحبيبات الكامل وحدها كافية وفعالة في التنبؤ بحد السائل وحد البلاستيك، مما يلغي الحاجة إلى خصائص فيزيائية أو كيميائية إضافية للتربة.

الكلمات المفتاحية: حدود أتربيرج، توزيع حجم الحبيبات، حد السيولة، حد اللدونة، الطين، الرمل، الغرين.