Prediction of Target LWD Modulus of Unbound Geomaterials by Lab LWD Test

Chuanjun Liu[1], Xiong Zhang[1], Jenny Liu[2]

[1]Department of Civil, Architectural and Environmental Engineering, Missouri University of Science and Technology, 1401 North Pine Street, Rolla, MO 65409

[2]Department of Civil, Architectural and Environmental Engineering, Missouri University of Science and Technology, 1401 North Pine Street, Rolla, MO 65409 (Corresponding author) e-mail: jennyliu@mst.edu

ABSTRACT

Modulus-based compaction acceptance evaluation of geomaterials using lightweight deflectometers (LWDs) had gained increasing attention in recent decades, as the traditional density-based method using nuclear density gauges (NDGs) was related to safety, regulatory, and cost concerns. Missouri Department of Transportation was interested in adopting LWDs in place of NDGs and creating related standards. In this method, determination of target LWD modulus was needed, because compaction acceptance was evaluated by comparing the in-situ LWD modulus with a predetermined target one. This study aimed to apply lab LWD test (LLWDT) method to characterize the relationship between lab LWD modulus, moisture content (MC), and applied stress for target LWD modulus prediction. With data obtained from a matrix of LLWDTs conducted using various soils, MCs, and applied stresses, the characterization was first performed for each soil using the model given in the American Society for Testing and Materials (ASTM) standard. However, this model illustrated a parabolic shape and failed to capture the inverse-sigmoid pattern shown in the data. Given the identified limitation, two machine learning (ML) models including random forest and multilayer perceptron (MLP) were employed for the characterization and compared with the ASTM model. The comparison suggested that the two ML models exhibited comparable prediction performance and considerably outperformed the ASTM model, with MLP showing the best generalization capacity.

Keywords: Unbound geomaterials, lightweight deflectometer modulus, target modulus, compaction acceptance, machine learning

1 Introduction

Non-nuclear compaction acceptance evaluation methods for unbound materials had been studied for decades, as the current density-based practice using nuclear density gauges (NDGs) had become less desirable due to safety, regulatory, and cost concerns (Nazarian et al. 2015; McLain and Gransberg 2019). Among these methods, the modulus-based method using lightweight deflectometers (LWDs) had gained more attention in recent decades, because of its ability to measure modulus, simple operation, and lower cost of ownership (Riad et al. 2023; Simon et al. 2023; Yousif et al. 2024; Wang et al. 2024). Missouri Department of Transportation (DOT) was interested in exploring the use of LWDs as replacement of NDGs and creating corresponding standards. This modulus-based method using LWDs assessed compaction quality by comparing LWD moduli (i.e., in-situ modulus and target modulus) and thus involved not only the in-situ LWD modulus measurement but also the establishment of target LWD modulus prior to the compaction assessment (Simon et al. 2023; Schwartz et al. 2017; Baker and Meehan 2020). Hence, determining target LWD moduli for typical Missouri unbound materials were needed for Missouri DOT.

A soil-specific target LWD modulus can be established using three standardized methods: empirical target values, control strip tests, and lab LWD tests (LLWDT) (Simon et al. 2023). The first two methods (i.e., empirical target values and control strip tests) were currently in practice for Minnesota DOT (MnDOT 5-692.256 2023). The empirical target values method derived target LWD moduli from a large database of existing LWD testing data for various geomaterials. This method necessitated not only an extensive dataset but also stringent consistency in LWD devices and testing methods. However, such a nationally applicable database remained unavailable, given Minnesota DOT was the only leading agency using this method (Riad et al.

2023; Simon et al. 2023). The control strip method determined target LWD moduli through continuous field LWD tests on a lift of soil that were repeatedly compacted by a roller in a designated area. The final modulus was determined when a negligible variation was observed between two continuously measured modulus values (MnDOT 5-692.256 2023). However, this field control strip method typically provided a limited number of target LWD moduli for various moisture content (MC) levels, as the MC control within the selected control strip was random and difficult to cover the full spectrum of MCs that possibly occurred in the field. Moreover, constructing a control strip was limited for small-area construction cases (Zhao et al. 2018).

LLWDT, also known as LWD test on Proctor mold, was first proposed by Schwartz et al. and later standardized in American Society for Testing and Materials (ASTM) E3331, *Standard Test Method for Measuring Target Modulus Using Light Weight Deflectometer on Compacted Proctor Mold Samples* (Schwartz et al. 2017; ASTM E3331 2022). In this method, pulse loads were applied onto the specimen prepared in a six-inch mold, and the elastic modulus under lateral confinement determined with known loads and collected deflection values was referred to as lab LWD modulus (E_{lab}). By systematically varying specimen MC and applied stress levels, a comprehensive dataset of E_{lab} can be acquired to characterize the MC-stress-modulus relationship using a two-way quadratic polynomial model. This calibrated model subsequently enabled the interpolation of target LWD modulus for any MC and applied stress. LLWDTs were found effective in capturing the impact of stress and MC on E_{lab} in previous studies (Schwartz et al. 2017; Baker and Meehan 2020). Overall, obtaining a target LWD modulus through LLWDTs was more comprehensive and informative than the other two methods, because it can produce predictions over a wide range of stress and MC, given both variables were found significantly influential to in-situ LWD modulus (Schwartz et al. 2017; Vennapusa 2008; Meehan et al. 2012).

Meanwhile, machine learning (ML) technique had been increasingly implemented in diverse industries in recent decades to characterize latent patterns in observed data and subsequently generate predictions with new input, due to the development of related algorithms, its proven reliability, and significant advancements in computational capacity (Alpaydın 2016). Among these ML techniques, random forest (RF) and multilayer perceptron (MLP) had been increasingly employed and exhibited superior prediction capacity in previous studies. For example, Hao and Pabst found RF and MLP outperformed other three ML models [i.e., neuroevolution of augmenting topologies, k-nearest neighbors, and decision tree (DT)] in predicting California bearing ratio and resilient modulus for crushed rocks (Hao and Pabst 2022). Saha et al. reported high prediction accuracy of MLP in estimating resilient modulus for both non-plastic and plastic soils (Saha et al. 2018).

Therefore, the objective of this study was to collect broad LLWDT data for typical Missouri geomaterials and characterize their MC-stress-modulus relationships for target LWD modulus predictions. In this study, eight unbound materials were sampled from different areas in Missouri for a series of lab tests. Following soil characterization tests, LLWDTs were performed on specimens prepared using at least five MC levels with seven stresses. Characterization of relationship between E_{lab}, stress, and MC was performed using the model suggested in ASTM E3331. In addition, two ML techniques including RF and MLP were employed and compared with the ASTM model.

2 Experiment Details and Data Results

2.1 Materials

Eight geomaterials were sampled from six areas in Missouri for this study. These soils were classified with their gradations shown in figure 1 and soil properties [i.e., liquid limit (LL) and

plastic limit (PL)] summarized in Table 1, according to the unified soil classification system (USCS). Soil types mainly included silty gravel (GM), well-graded gravel (GW), silty sand (SM), well-graded sand (SW), and lean clay (CL). Each soil was designated by its source area and USCS classification. In addition, specific gravity (G_s), and optimum MC (OMC) and maximum dry unit weight (MDUW) determined with modified Proctor tests (ASTM D1557 2021) are presented in Table 1.

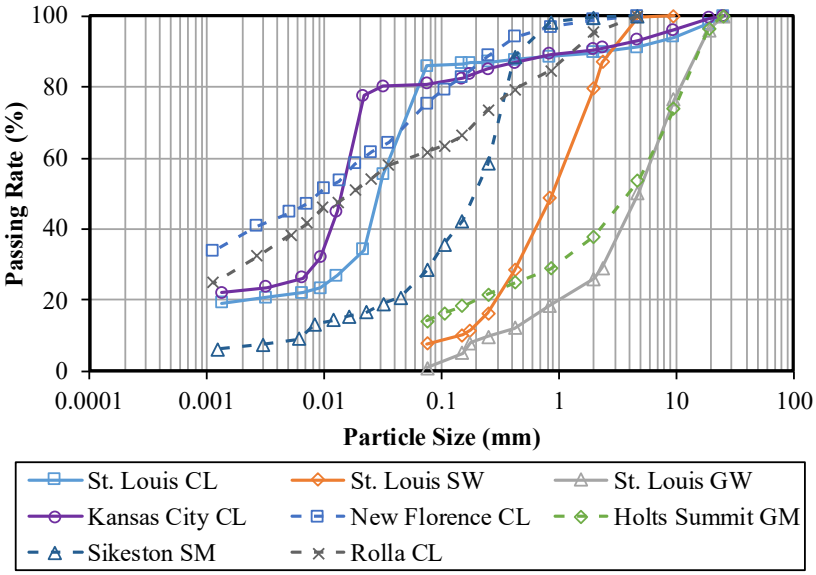


FIG. 1 Gradation curves of eight geomaterials

TABLE 1 Physical properties and classifications of eight geomaterials

Soils	LL	PL	G_s	USCS	OMC, %	MDUW, kN/m^3
Holts Summit GM	-	-	2.660	GM	7.2	21.6
St. Louis GW	-	-	2.663	GW	7.5	21.6
Sikeston SM	-	-	2.623	SM	8.8	19.6
St. Louis SW	-	-	2.645	SW	8.8	20.2
Rolla CL	31	15	2.697	CL	11	19.2
St. Louis CL	32.2	23.1	2.705	CL	14.5	18.6
Kansas City CL	30.7	21	2.613	CL	12.5	18.7
New Florence CL	40.7	15	2.742	CL	14.5	18.1

2.2 Testing Plan

A matrix of LLWDTs was planned to perform to collect comprehensive data. Figure 2 illustrates the general procedure of LLWDT. This test started with specimen preparation (step a), followed by positioning the LWD loading plate on the specimen (steps b and c), and ended with imposing six pulse loads on the specimen (step d). Data acquired from the test typically included applied stress, deflection, and MC. Assuming the soil was linearly elastic, homogeneous, and isotropic, the elastic modulus of soil with lateral constraint was calculated using equation (1) and referred to as E_{lab} (Nielson et al. 1969). In the calculation, typical Poisson's ratio values for different types of soils given in mechanistic-empirical pavement design guide (MEPDG) were employed (NCHRP 2004). Since three soil types (i.e., clayey, sandy, and gravelly soils) were involved in this study, Poisson's ratio values of 0.2, 0.3, and 0.35 were used for them, respectively.

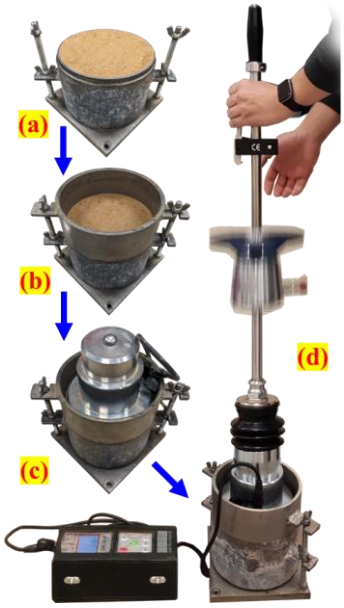


FIG. 2 LWD test on Proctor mold

$$E_{lab} = \left(1 - \frac{2v^2}{1-v}\right)\frac{H\sigma_{lab}}{s} \quad (1)$$

where

E_{lab} = lab LWD modulus, Pa,

v = Poisson's ratio,

H = specimen height, m,

σ_{lab} = applied stress, Pa, and

s = deflection, m.

A series of LLWDTs were conducted on specimens prepared at various stress and MC levels using the eight soils. Table 2 summarizes these stress and moisture levels for each soil. Except for Kansas City CL and St. Louis CL, the other six soils were tested at five different MCs. These five MCs included two lower than OMC, OMC (underlined), and two higher than OMC. Besides these five MCs, two additional MCs higher than OMC were tested for Kansas City CL and St. Louis CL. Moreover, intervals between MCs varied between gravelly soils and other soils, since the former had considerably lower fines content. For example, an interval of 2% was generally chosen for sandy and clayey soils, while MCs tested on gravelly soils increased by 1.5%. A relative compaction of 95% was controlled for all specimens by varying the number of blows applied per layer (Dutta and Kodikara 2022). In addition, all eight soils were tested with seven stress levels ranging from 75 kPa to 265 kPa. These stress levels were achieved by adjusting the falling height of LWD hammer. Each combination of stress and moisture levels was tested with three replicates.

TABLE 2 Stress and MC levels for each soil

Soil	MC, %	Stress, kPa
Holts Summit GM	5, 6.5, 7.2, 8.5, 10	75, 106, 130, 150, 168, 201, 265
St. Louis GW	4.5, 6, 7.5, 8.5, 9.5	
Sikeston SM	4.8, 6.8, 8.8, 10.8, 12.8	
St. Louis SW	5, 7, 9, 10, 11	
Rolla CL	7, 9, 11, 13, 15	
St. Louis CL	10.2, 12.2, 14.2, 16.2, 18.2, 20.2, 22.2	
Kansas City CL	8.5, 10.5, 12.5, 14.5, 16.5, 18.5, 20.5	
New Florence CL	10.5, 12.5, 14.5, 16.5, 18.5	

2.3 LLWDT Results

Upon the completion of all LLWDTs, a dataset containing 924 instances was acquired. Sequentially, noise reduction was conducted by comparing the values of three replicates for each combination of factors (i.e., soil, MC, and applied stress). In this process, z-score, defined as the ratio of the difference between measured value and mean to the standard deviation, was employed (Mendenhall and Sincich 2016). Instances that had E_{lab} values with z-score absolute values more than 1.1 were identified as outliers and excluded. 924 instances in the initial dataset were refined to 754 after the noise reduction.

Figure 3 illustrates the averaged results of E_{lab} measured at various MC and stress levels for the eight soils. Due to the large-scale premixing of samples, MCs for specimens did not exactly match those target values listed in Table 2, instead, comparable levels were maintained. Overall, noticeable interactions were identified between E_{lab} and factors including MC, soil type, and applied stress. As for MC, as shown in figure 3, E_{lab} of each soil generally decreased with the increase of MC at each stress level, exhibiting a trend that differed from the parabolic shape of a compaction curve. This agreed with the findings in previous studies (Zhao et al. 2018; Baker and

158 Meehan 2020). Specifically, E_{lab} overall either slightly fluctuated or remained relatively
159 unchanged as MC increased from the lowest tested MC to OMC and dramatically decreased to
160 lower levels as MC increased beyond OMC, showing a partial or complete inverse-sigmoid
161 trend. This was consistent with the findings in the previous study (Morales and Zuniga 2023). On
162 the other hand, New Florence CL was an exception, for which only fluctuations were observed in
163 its results throughout all tested MCs. This could be attributed to its significantly higher LL
164 compared to those of the other three CL soils (i.e., 41 vs. approximately 31). A higher LL
165 indicates a stronger pore water pressure buildup potential when being loaded at a status of
166 undrained and high MC, thus leading to lower deflections and higher moduli at tested high MCs
167 (Zhi-peng 2020; Casey and Germaine 2013).

168 With respect to the soil type, fine-grained soils generally exhibited higher E_{lab} than coarse-
169 grained soils at MCs no higher than OMCs. For example, as illustrated in figure 3, the four fine-
170 grained soils (i.e., Rolla CL, St. Louis CL, and Kansas City CL, and New Florence CL) overall
171 produced highest values, followed by the two gravelly soils (i.e., Holts Summit GM and St.
172 Louis GW), and the two sandy soils (i.e., Sikeston SM and St. Louis SW) yielded the lowest
173 values. In addition, the impact of stress level on E_{lab} was significant at MCs no more than OMC.
174 Within this MC range, E_{lab} of each soil overall increased with the increase of applied stress,
175 showing a typical stress dependency feature of unbound materials. This was in agreement with
176 the findings in the previous study (Baker and Meehan 2020).

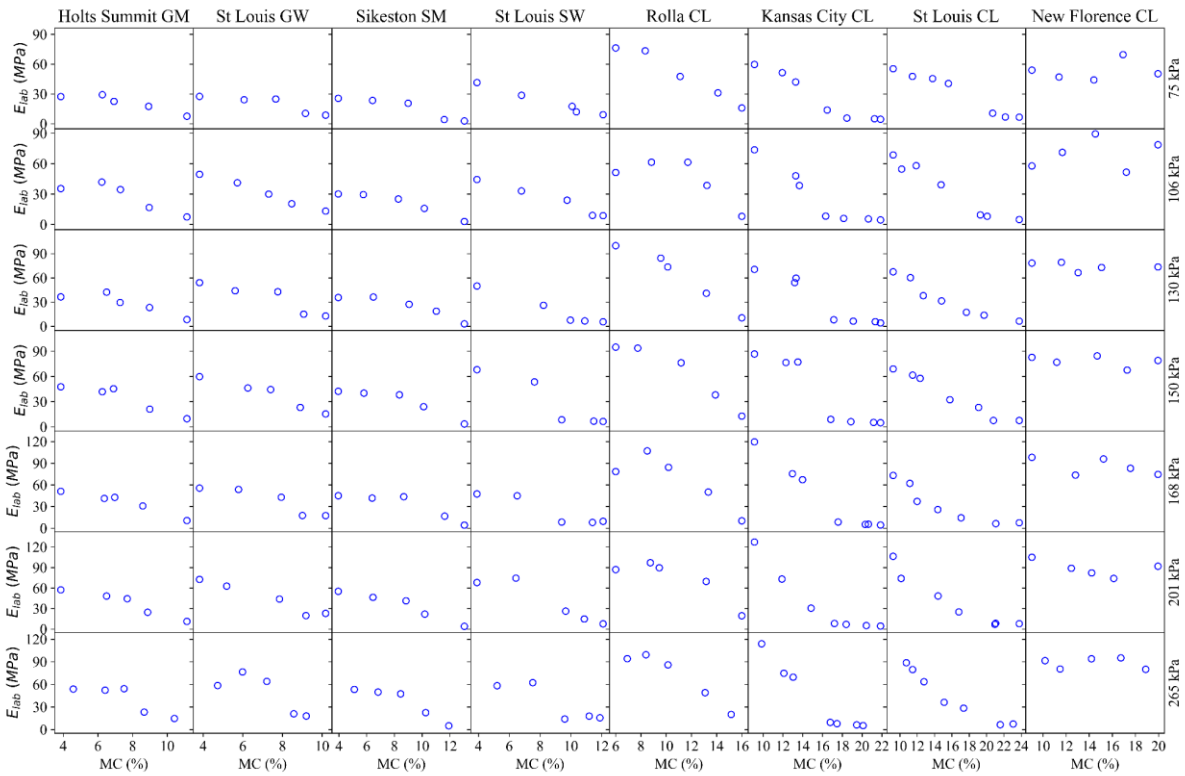


FIG. 3 E_{lab} of eight geomaterials

3 Characterization of E_{lab} Using ASTM Model

According to ASTM E3331, a soil-specific two-way quadratic polynomial model [equation (2)] was recommended for characterizing the relationship between E_{lab}, MC, and applied stress, and predicting target LWD modulus after calibration (ASTM E3331 2022). Therefore, a series of soil-specific nonlinear regressions were performed for the eight tested soils using the least-squares method implemented via the minimize method in SciPy (version 1.15.2) (Virtanen et al. 2020). Several performance metrics were employed for the regression analysis, including coefficient of determination (R^2), root mean square error (RMSE), and mean absolute percentage error (MAPE) (Zhang et al. 1998). R^2 [equation (3)] was a widely used metric that measured the linear correlation between two datasets. This metric was typically no more than one, with a

criterion that a value closer to one indicated a stronger correlation. RMSE, defined in equation (4), was an error metric that avoided the cancellation between under- and over-predictions and eliminated the impact of one or two large differences that could be heavily weighted (Wallach et al. 2019). This metric was non-negative, and a lower value typically indicated a superior model accuracy. MAPE, defined in equation (5), was a metric that focused more on the relative error between observations and predictions. MAPE ranged from zero to 100%, with a criterion that the closer the MAPE was to zero the better the prediction.

$$E_{lab} = a_0 + a_1 \times MC + a_2 \times MC^2 + a_3 \times \sigma_d + a_4 \times \sigma_d^2 \quad (2)$$

where

a_i $(i = 0,1,2,3,4)$ = regression coefficients, and

MC = specimen moisture content, %.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2}{\sum_{i=1}^{n}(Y_i - \overline{Y}_i)^2} \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2}{n}} \quad (4)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \widehat{Y}_i}{Y_i}\right| * 100\% \quad (5)$$

where

Y_i = observed value,

$\overline{Y}_i$ = mean observed value,

$\widehat{Y}_i$ = predicted value, and

n = number of observations.

Table 3 lists performance metrics of regression for the ASTM model. R^2 values ranged from 0.60 to 0.91, indicating moderate to good correlations for different soils. For example, except for

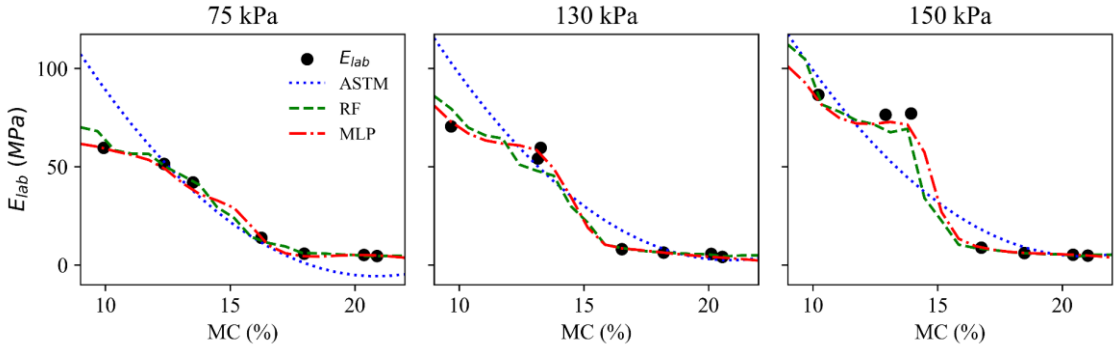
New Florence CL (i.e., R^2 of 0.60), the remaining seven soils overall yielded R^2 values greater than 0.8. On the other hand, the error metrics, RMSE and MAPE, displayed different trends. For example, RMSE values exhibited a span from 5.22 MPa (Holts Summit GM) to 10.97 MPa (Kansas City CL). Values of MAPE ranged from 10.3% to 58.4%, with Kansas City CL being the highest and New Florence CL being the lowest. Overall, these metrics indicated that the ASTM model exhibited moderate to good goodness of fit for all tested soils.

TABLE 3 Performance metrics of ASTM model

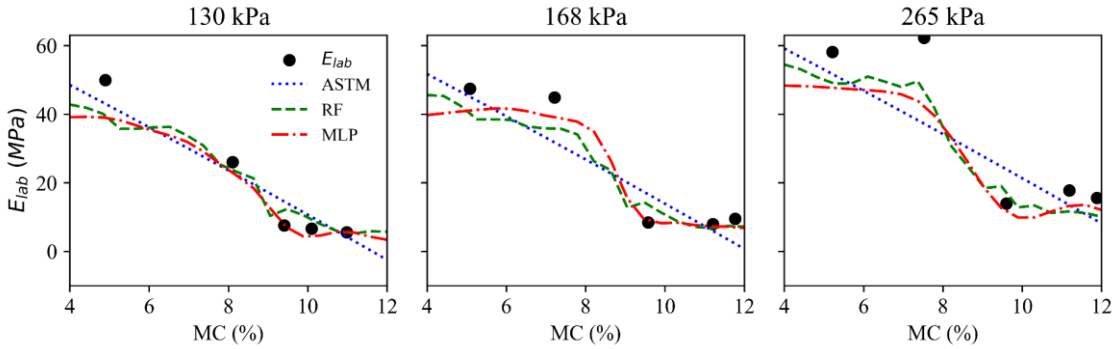
Soil	R^2	RMSE, MPa	MAPE, %
Holts Summit GM	0.89	5.22	18.0
St. Louis GW	0.79	9.07	30.8
Sikeston SM	0.89	5.53	50.3
St. Louis SW	0.80	7.66	44.8
Rolla CL	0.91	9.29	19.5
Kansas City CL	0.90	10.97	58.4
St. Louis CL	0.91	8.38	31.6
New Florence CL	0.60	9.45	10.3

On the other hand, two limitations were identified when inspecting the fitted trendlines of the ASTM model. First, the parabolic curves described by the ASTM model could not fit the inverse logistic trends displayed in most E_{lab} results. Figures 4A and 4B show some examples of poor fits of the ASTM model observed on Kansas City CL and St. Louis SW, respectively. As shown in figure 4A, the prediction of ASTM model for Kansas City CL only fitted a segment of the inverse-sigmoid trend. For example, the ASTM model missed three points (i.e., MCs of 10%, 21%, and 22%) at 75 kPa, two points (i.e., MC of 10% and 16%) at 130 kPa, and three points (i.e., MCs of 13%, 14%, and 17%) at 150 kPa. As illustrated in figure 4B, fittings of the ASTM model for St. Louis SW deteriorated further, producing an essentially linear line approximation

and consequently exhibiting greater discrepancies. Notable deviations occurred at MCs of 5% and 9% for 130 kPa, at 7% and 10% MCs for 168 kPa, and at MCs of 8% and 10% for 265 kPa.



(*A*)

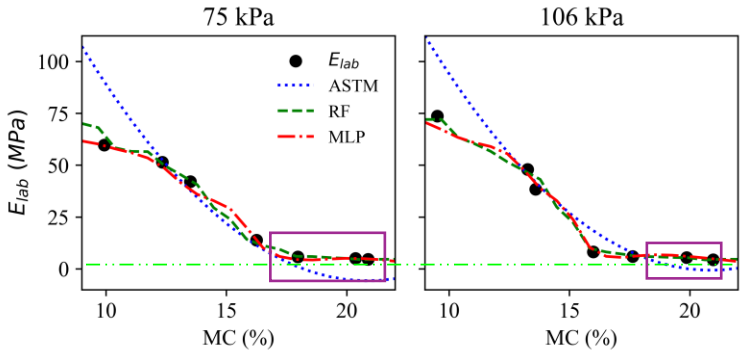


(*B*)

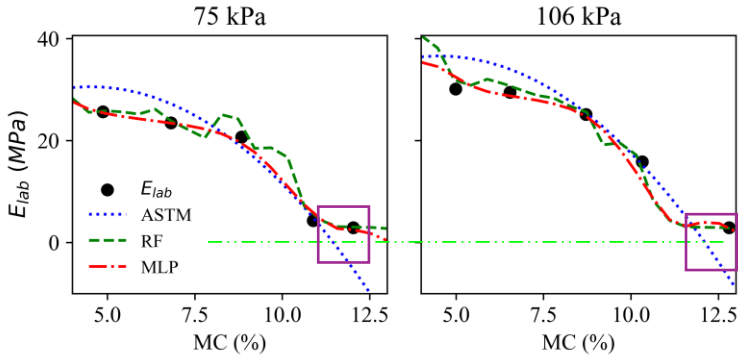
FIG. 4 Examples of poor fits of ASTM model on sigmoid shape: (*A*) Kansas City CL and (*B*) St. Louis SW

Moreover, the ASTM model generated nonpositive modulus predictions due to its inadequate fitting capacity for the inverse sigmoid pattern. These non-positive predictions occurred at combinations of high MCs and low stress levels (i.e., 75 kPa and 106 kPa) for some soils, which agreed with the findings of the previous study (Simon et al. 2023). Figures 5a through 5c illustrate some examples of nonpositive modulus predictions of the ASTM model for Kansas City CL, Sikeston SM, and St. Louis GW, respectively. In these figures, horizontal dash double-dot lines representing $E_{lab} = 0$ were drawn to assist in recognizing the positivity of

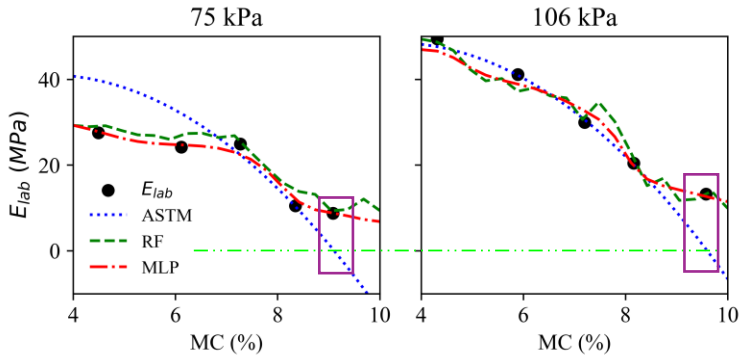
predicted moduli. As shown in figure 5A, for Kansas City CL, the model predicted negative modulus values as MC increased beyond 18% and 20% at stresses 75 kPa and 106 kPa, respectively. For Sikeston SM illustrated in figure 5B, MCs beyond which the modulus prediction yielded negative values were 11% and 12% for stresses 75 kPa and 106 kPa, respectively. The modulus predictions for St. Louis GW (displayed in fig. 5C) did not give negative values but zero at MCs around 9% for both stresses of 75 kPa and 106 kPa.



(*A*)



(*B*)



(*C*)

FIG. 5 Examples of nonpositive modulus predictions of ASTM model: (*A*) Kansas City CL, (*B*) Sikeston SM, and (*C*) St. Louis GW

Based on fitting results of the ASTM model, its limitations were identified in fitting an inverse-S data trend and providing reasonable modulus prediction for combinations of high MCs and low stresses for some soils. Therefore, a new model that better characterized the relationship between E_{lab}, MC, and applied stress was needed.

4 Characterization of E_{lab} Using ML Models

4.1 Methodology

4.1.1 RF and MLP Models

RF was an ensemble model that consists of multiple DTs. DTs in an RF were incorporated with randomization implemented in data sampling and feature selection (Rigatti 2017; Breiman 2001). With a new subset randomly sampled from the original dataset using bagging or pasting, a hierarchical structure of nodes (i.e., root, split or internal, and leaf nodes) was generated to form a DT (illustrated in fig. 6). The node splitting was based on impurity measures (e.g., Gini and Entropy) for classification, or mean square error (MSE) for regression (Géron 2022). Meanwhile, during the feature selection at each node, a subset of features was randomly selected for DT

training. The outputs of all DTs in an RF were aggregated into a single prediction by voting for classification tasks, or by averaging for regression tasks.

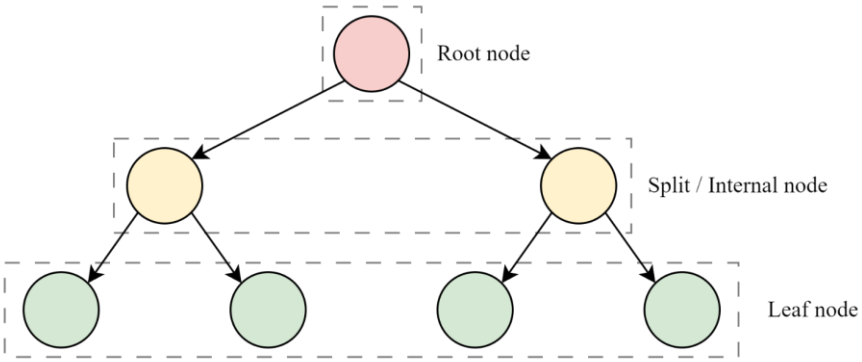
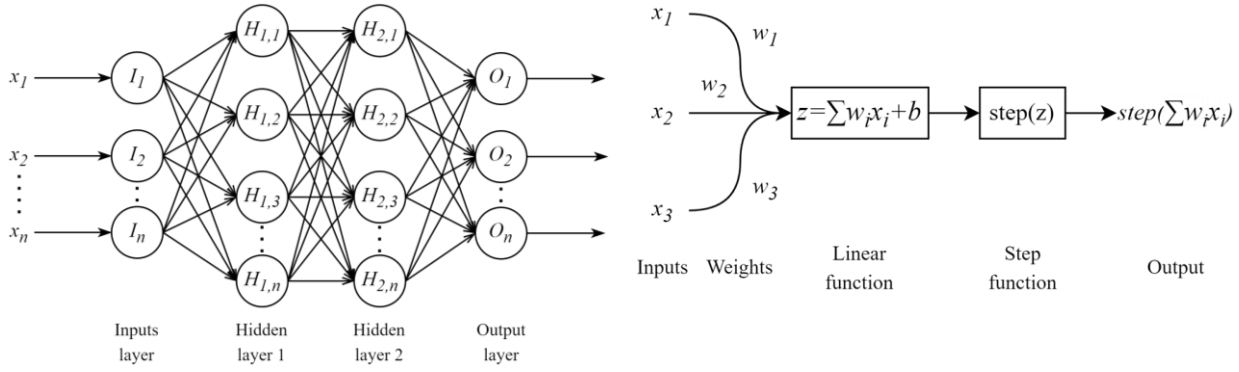


FIG. 6 A general structure of DT

MLP was a type of artificial neural network that consisted of multiple layers of perceptron. As illustrated in figure 7A, layers in an MLP generally included one input layer, one or more hidden layers, and one output layer. Each layer of perception contained one or more threshold units (TLU), also known as neurons. Neurons in two neighboring layers were usually fully connected (Géron 2022; Krogh 2008; Rosenblatt 1958; Zou et al. 2008; Zhang et al. 1998). Figure 7B illustrates the process of computation in a TLU. With inputs (x_i) fed to the TLU, it first calculated the sum of inputs multiplied with weights (w_i) and a bias term b. Then, a step function (also known as activation function) was applied to this sum to introduce a degree of nonlinearity before outputting.

Both RF and MLP were well supported by scikit-learn (Hao and Ho 2019; Pedregosa et al. 2011). Hence, scikit-learn (version 1.6.1) along with Python 3.12.0 was employed in this study. In addition, hyperparameter terminologies used in this study mostly adhered to the conventions of scikit-learn.



(*A*) (*B*)

FIG. 7 A general structure of (*A*) MLP and (*B*) TLU

4.1.2 Data Preparation

Data preparation mainly included selecting features and target from the data and splitting the data into training and test sets. Predictor and response variables employed in ML models were denoted as features and targets, respectively. In this study, variables that were related to soil properties and LLWDT were chosen as predictive features for ML model development. For example, incorporated soil properties comprised proportions of gravel, sand, and fine, G_s, OMC, and MDUW. LLWDT related parameters, such as applied stress and MC, were also included as features. The target variable was E_{lab}. In addition, a ratio of 75% to 25% was employed when splitting the data into training and test sets, with a random state of 42. This random state was employed to ensure consistent data split throughout the model development.

4.1.3 Hyperparameters Tuning Methods

Optimizing hyperparameters for an ML model eliminated overfitting and underfitting, leading to better model accuracy and efficiency. An overfitted model exhibited strong performance on the training data but generalized poorly to unseen data, while an underfitted one had poor performance on both training and novel datasets. Grid search and k-fold cross-validations were used for achieving a balance between overfitting and underfitting. In grid search, a predefined set

of hyperparameter values were tested and their respective performance were compared to identify the optimal configuration, thus eliminating underfitting (Ben-Hur and Weston 2010). K-fold cross-validation method partitioned the training dataset into k parts of equal size and iteratively used each for model evaluation. This method can help eliminate overfitting (Lever et al. 2016; Sweet et al. 2023). Grid search, combined with k-fold cross-validation, was usually performed at least twice to refine the hyperparameter selection. The first search identified promising hyperparameter ranges, and subsequent ones zoomed in those ranges further to find optimal values without overfitting.

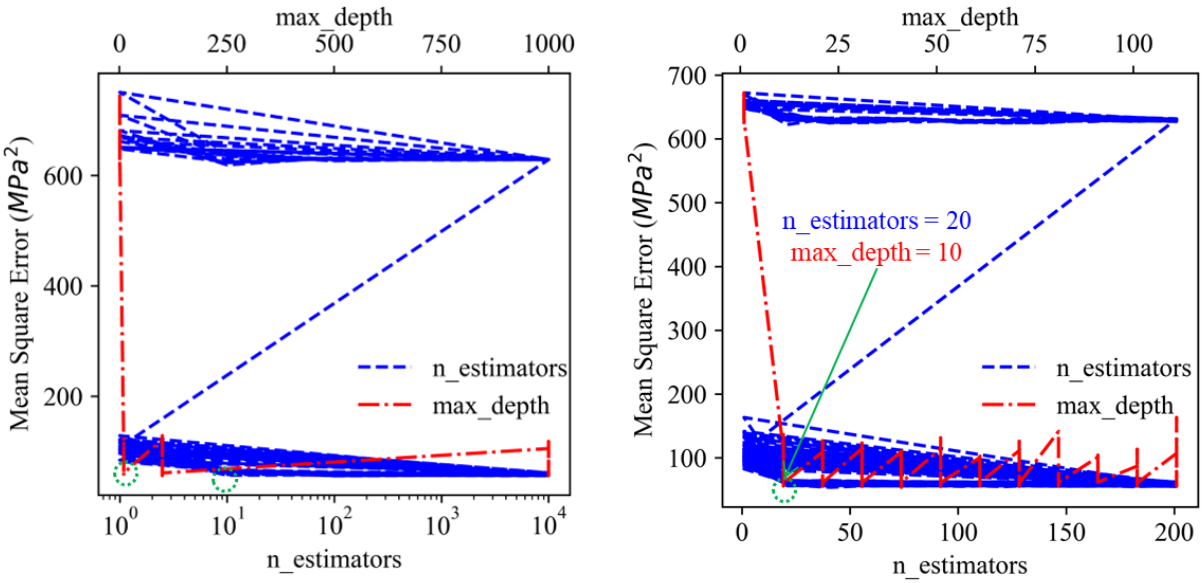
4.2 Results and Discussion

4.2.1 Optimal Hyperparameters

Table 4 summarizes initial hyperparameter candidates and optimal ones for both RF and MLP. These initial candidates covered a wide range of options for model training, which were narrowed down gradually with the tuning process going on. For RF, two optimization trials (shown in fig. 8) were performed to determine its optimal hyperparameters. The first trial (fig. 8A) suggested that optimal values of n_estimators and max_depth fell into the same range of 1 to 100. A further inspection (fig. 8B) revealed optimal values of 20 and 10 for n_estimators and max_depth, respectively. In these trials, initial options for min_samples_split and min_samples_leaf remained unchanged, and their optimal values were determined in the second trial (listed in Table 4). For MLP, three hyperparameter tuning trials were conducted with only candidates of hidden_layer_sizes being tuned. Candidates and outputs of these trials are summarized in Table 5. An optimal hidden layer configuration of (25, 50) occurred in the second trial and was further confirmed in the following one. Similarly, optimal values for other hyperparameters were concurrently determined at the end of model training (listed in Table 4).

TABLE 4 Hyperparameters and their ranges in grid search

Model	Hyperparameters	Search range	Optimal
RF	n_estimators	1, 10, 100, 1000, 10000	20
	max_depth	1, 10, 100, 1000	10
	min_samples_split	2, 3, 4	2
	min_samples_leaf	1, 2, 3	1
MLP	hidden_layer_sizes	(50,), (200,), (50, 50), (200, 200)	(25, 50)
	activation	logistic, relu, tanh	tanh
	learning_rate_init	0.0001, 0.001, 0.01, 0.1	0.01
	solver	adam, sgd, lbfgs	adam



(*A*) (*B*)

FIG. 8 Hyperparameter tuning of RF: (*A*) first trial and (*B*) second trial

TABLE 5 Three hyperparameter tuning trials of MLP

Trial	hidden_layer_sizes candidates	Output
1	(50,), (200,), (50, 50), (200, 200)	(50, 50)
2	(25, 50), (50, 50), (50, 75)	(25, 50)
3	(25,), (25, 25), (25, 50)	(25, 50)

4.2.2 Model Performance

Three identical performance metrics (i.e., R^2, RMSE, and MAPE) that were used to assess the ASTM model were employed to evaluate the two trained ML models (i.e., RF and MLP). Table 6 lists performance metrics of the ASTM model and training and test processes for each ML model. In terms of the correlation between predicted and measured values, both models demonstrated robust prediction capabilities, achieving R^2 values greater than 0.96 for both training and test processes. Specifically, compared with MLP, RF yielded a slightly higher R^2 value (i.e., 0.99 > 0.97) in the training phase and an identical value (i.e., 0.96) in the test phase. RMSE and MAPE generally revealed a similar trend to R^2. RF generally exhibited superior performance to MLP. On the other hand, with respect to the balance between overfitting and underfitting, MLP overall showed closer values in training and test phases for all metrics compared with RF, indicating a better balance between overfitting and underfitting. Nonetheless, both ML models outperformed the ASTM model in terms of prediction accuracy (i.e., R^2: 0.96 > 0.91).

TABLE 6 Performance metrics of RF, MLP, and ASTM Model

Metric	RF		MLP		ASTM Model
-	*Training*	*Test*	*Training*	*Test*	*Range*
R^2	0.99	0.96	0.97	0.96	0.61 ~ 0.91
RMSE, MPa	3.07	6.20	4.7	6.2	5.22 ~ 10.97
MAPE, %	7.2	16.0	10.0	17.5	10.3 ~ 58.4

The rigorous fitting quality of each model was further proven by their prediction trendlines against observations and distributions of residuals. Figures 9a and 9b illustrate the correlation between predicted and measured E_{lab} for RF and MLP, respectively. As shown in figure 9A, a strong correlation between predictions and observations was indicated for RF by the close

proximity of data points to the line of equality (LOE). For MLP, as depicted in figure 9B, the distribution of correlation points along the LOE was slightly less concentrated, indicating a slightly deteriorated prediction performance. This agreed with their R^2 values for the training phase (i.e., 0.99 and 0.97 for RF and MLP, respectively).

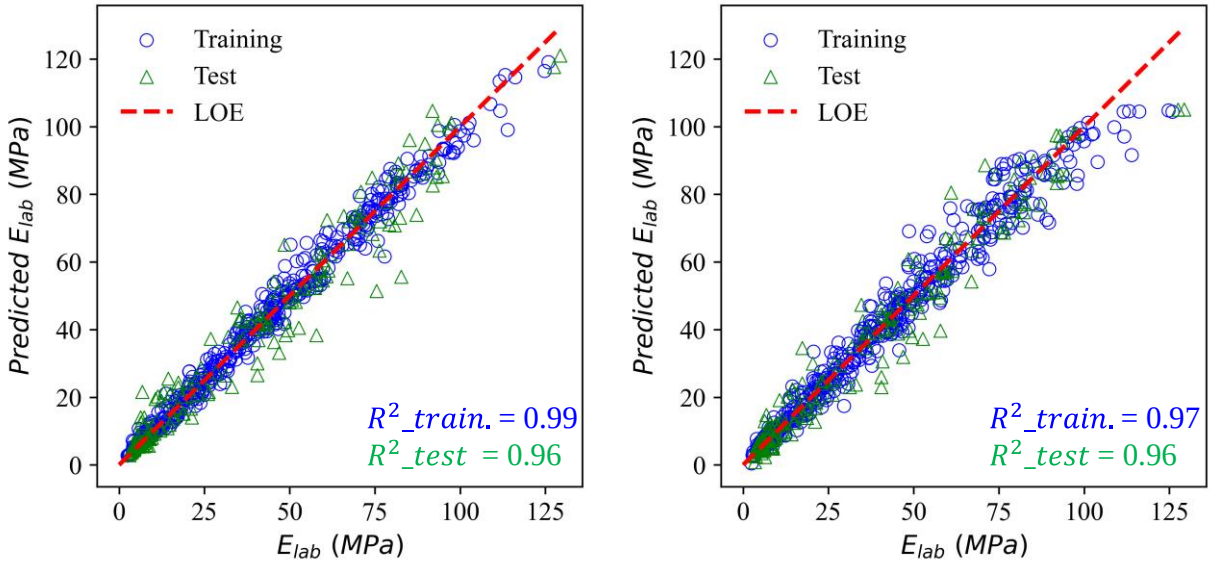
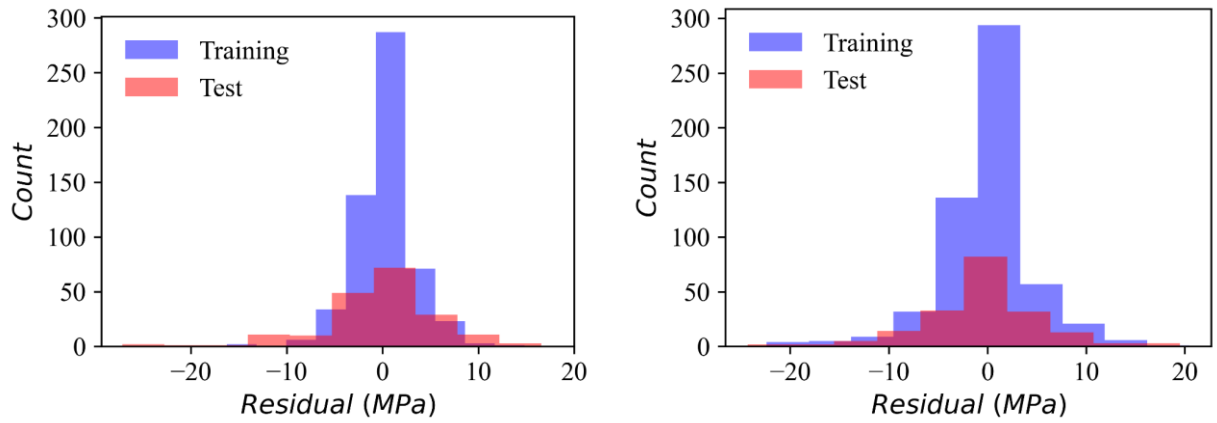


FIG. 9 Predicted and measured moduli: (*A*) RF and (*B*) MLP

Figures 10a and 10b show prediction residuals for RF and MLP, respectively. As illustrated in figure 10A, prediction residuals of RF overall followed a bell-shape normal distribution pattern, with the residual of zero having the highest count. Most residuals lay within a relatively smaller span (i.e., ±5 MPa) for both training and test processes, while a small number of residuals had absolute values more than 10 MPa. The distribution of residuals of MLP (shown in fig. 10B) was generally similar to that of RF, following a bell-shape distribution. The residual of zero had the most instances, and the majority of residuals fell within the range of ±10 MPa, with a limited number of residuals out of this range. Exhibiting bell-shape normal distribution patterns for both models indicated that they were trained with overfitting and underfitting being well balanced.



(*A*) (*B*)

FIG. 10 Prediction residuals: (*A*) RF and (*B*) MLP

Besides having a better applicability and excellent prediction performance, the two ML models overcame limitations identified in fittings of the ASTM model. In terms of fitting inverse-sigmoid data patterns, as shown in figures 4a, both RF and MLP accurately captured data patterns for Kansas City CL, while the ASTM model only fitted partially. Likewise, as illustrated in figure 4B, fittings on St. Louis SW revealed a similar conclusion, though both RF and MLP slightly deviated from measured E_{lab} values at the two MCs lower than OMC (i.e., 8.2%) for stresses of 168 kPa and 265 kPa.

In cases where the ASTM model produced nonpositive E_{lab} predictions, the two ML models yielded more rational ones. As illustrated in figure 5A, for Kansas City CL, while the ASTM predicted negative E_{lab} values at MCs higher than 20% for both stresses of 75 kPa and 106 kPa, both trendlines of RF and MLP precisely passed measured data points, indicating accurate predictions. As shown in figure 5B, fittings on Sikeston SM suggested a similar conclusion that both ML models overcame the negative predictions issue at high MCs, so as fittings on St. Louis GW depicted in figure 5C.

In addition, there was a subtle difference between trendlines of RF and MLP. As shown in figures 4 and 5, the trendlines of RF, compared with those for MLP, exhibited some spikes on

curve segments where there were no measured data points, such as spikes at approximately 8% MC and 75 kPa stress for Sikeston SM (fig. 5B) and approximately 7.5% MC and 106 kPa for St. Louis GW (fig. 5C). This indicated a lower level of generalization for RF (Barbiero et al. 2020). This agreed with the previous observation that RF showed a slightly worse balance between overfitting and underfitting than MLP. Therefore, MLP was recommended for predicting target LWD modulus, given its outstanding prediction performance and generalization ability.

5 Conclusions

In this study, the relationship between E_{lab}, MC, and applied stress was characterized using LLWDT data from eight soils to predict target LWD modulus for modulus-based compaction acceptance evaluation using LWDs. The study started with a series of tests for geomaterial characterizations, followed by a matrix of LLWDTs conducted using at least five MCs and seven applied stresses. Experiment data from LLWDTs were first denoised and analyzed. The analysis revealed that E_{lab} of each soil, except for New Florence CL, generally decreased with the increase of MC at each stress level. E_{lab} overall increased as the applied stress increased at each MC for all soils. Clayey soils overall had higher E_{lab} values than gravelly soils, and those of sandy soils were the lowest. With all denoised data, the characterization of relationship between E_{lab}, MC, and applied stress was first performed for each soil using the model given in the ASTM standard. However, limitations were observed: the ASTM model illustrated a parabolic shape and failed to capture the inverse-sigmoid pattern shown in data. In addition, the ASTM model predicted nonpositive E_{lab} values at high MCs for certain soils. Given the identified limitations of the ASTM model, two ML models including RF and MLP were employed for the characterization. Instead of being soil-specific, the two ML models were trained and tested for all

tested soils. Besides MC and applied stress, some soil properties were included in features, including proportions of gravel, sand, and fine, G_s, OMC, and MDUW. The two trained ML models were then compared with the ASTM model using three performance metrics. The comparison suggested that the two ML models exhibited comparable prediction performance and considerably outperformed the ASTM model. For example, the two ML models produced R^2 values greater than 0.96, while those for the ASTM model ranged from 0.6 to 0.91. Moreover, the two ML models overcame the limitations of the ASTM model in characterizing the relationship between E_{lab}, MC, and applied stress. Lastly, MLP was recommended for predicting target LWD modulus, given its outstanding prediction performance and generalization ability.

While the developed MLP model demonstrated excellent prediction capacity within the scope of this study, future work including training and validating the model with more data from different project sites was recommended to make it more universally applicable. In addition, developing an easy-to-use graphic user interface program was recommended, as it would greatly facilitate the implementation of the developed MLP model in highway agencies.

Acknowledgements

The authors would like to gratefully acknowledge the funding provided by Missouri DOT. Special thanks are also extended to Beshoy Riad for his help in data collection.

References

Alpaydın, Ethem. 2016. *Machine Learning: The New AI*. The MIT Press Essential Knowledge Series. The MIT Press.

ASTM D1557. 2021. *Standard Test Methods for Laboratory Compaction Characteristics of Soil Using Modified Effort (56,000 Ft-Lbf/Ft3 (2,700 kN-m/M3))*. Test Method D1557. Version 12. ASTM International, July. https://doi.org/10.1520%2FD1557-12R21.

ASTM E3331. 2022. *Standard Test Method for Measuring Target Modulus Using Light Weight Deflectometer (LWD) on Compacted Proctor Mold Samples*. Test Method E3331. Version 22a. ASTM International, December. https://doi.org/10.1520%2FE3331-22A.

Baker, William J., and Christopher L. Meehan. 2020. *Continuous Compaction Control Measurements for Quality Assurance in Conjunction with Light Weight Deflectometer Target Modulus Values*. February 21, 368–76. https://doi.org/10.1061/9780784482803.040.

Barbiero, Pietro, Giovanni Squillero, and Alberto Tonda. 2020. "Modeling Generalization in Machine Learning: A Methodological and Computational Study." arXiv:2006.15680. Preprint, arXiv, June 28. https://doi.org/10.48550/arXiv.2006.15680.

Ben-Hur, Asa, and Jason Weston. 2010. "A User's Guide to Support Vector Machines." *Methods in Molecular Biology (Clifton, N.J.)* 609: 223–39. https://doi.org/10.1007/978-1-60327-241-4_13.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. https://doi.org/10.1023/A:1010933404324.

Casey, Brendan, and John T. Germaine. 2013. "Stress Dependence of Shear Strength in Fine-Grained Soils and Correlations with Liquid Limit." *Journal of Geotechnical and Geoenvironmental Engineering* 139 (10): 1709–17. https://doi.org/10.1061/(ASCE)GT.1943-5606.0000896.

Dutta, Troyee Tanu, and Jayantha Kodikara. 2022. "Evaluation of Unbound/Subgrade Material Rutting and Resilient Behaviour Based on Initial Density and Saturation Degree." *Transportation Geotechnics* 35 (July): 100782. https://doi.org/10.1016/j.trgeo.2022.100782.

Géron, Aurélien. 2022. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. 3rd edition. O'Reilly Media, Inc.

Hao, Jiangang, and Tin Kam Ho. 2019. "Machine Learning Made Easy: A Review of Scikit-Learn Package in Python Programming Language." *Journal of Educational and Behavioral Statistics* 44 (3): 348–61. https://doi.org/10.3102/1076998619832248.

458 Hao, Shengpeng, and Thomas Pabst. 2022. “Prediction of CBR and Resilient Modulus of
459 Crushed Waste Rocks Using Machine Learning Models.” *Acta Geotechnica* 17 (4):
460 1383–402. https://doi.org/10.1007/s11440-022-01472-1.

461 Krogh, Anders. 2008. “What Are Artificial Neural Networks?” *Nature Biotechnology* 26 (2):
462 195–97. https://doi.org/10.1038/nbt1386.

463 Lever, Jake, Martin Krzywinski, and Naomi Altman. 2016. “Model Selection and Overfitting.”
464 *Nature Methods* 13 (9): 703–4. https://doi.org/10.1038/nmeth.3968.

465 McLain, K. W., and D. D. Gransberg. 2019. “Nuclear Density Gauge Compaction Testing
466 Alternatives: Synthesis and Critical Analysis.” *Journal of Structural Integrity and*
467 *Maintenance* 4 (2): 86–96. https://doi.org/10.1080/24705314.2019.1603192.

468 Meehan, Christopher L., Faraz S. Tehrani, and Farshid Vahedifard. 2012. “A Comparison of
469 Density-Based and Modulus-Based in Situ Test Measurements for Compaction Control.”
470 *Geotechnical Testing Journal* 35 (3): 387–99.

471 Mendenhall, William M., and Terry L. Sincich. 2016. *Statistics for Engineering and the Sciences*.
472 6th ed. Chapman and Hall/CRC.

473 MnDOT 5-692.256. 2023. *Light Weight Deflectometer – LWD Procedure & Target Value*
474 *Determination*. 5-692.256. Minnesota Department of Transportation, April 1.

475 Morales, Sebastian, and Isaac E. Zuniga. 2023. “Impact of Fines on Various Base Material
476 Properties.” *Transportation Research Record* 2677 (11): 367–74.
477 https://doi.org/10.1177/03611981231166389.

478 Nazarian, Soheil, M. Mazari, I. N. Abdallah, A. J. Puppala, L. N. Mohammad, and M. Y. Abu-
479 Farsakh. 2015. *Modulus-Based Construction Specification for Compaction of Earthwork*
480 *and Unbound Aggregate*. NCHRP 10-84. Transportation Research Board Washington,
481 DC, USA.

482 NCHRP. 2004. *Guide for Mechanistic-Empirical Design of New and Rehabilitated Pavement*
483 *Structures*. NCHRP 01-37A. National Cooperative Highway Research Program.

484 Nielson, F. Dwayne, Choosin Bhandhausavee, and Ko-Shing Yeb. 1969. “Determination of
485 Modulus of Soil Reaction from Standard Soil Tests.” *Highway Research Record*, no. 284.

486 Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, et al. 2011. “Scikit-Learn: Machine
487 Learning in Python.” *Journal of Machine Learning Research* 12: 2825–30.

488 Riad, Beshoy, Xiong Zhang, Jenny Liu, and Yizhuang David Wang. 2023. “Compaction Quality
489 Assurance Specifications of Unbound Materials.” *Journal of Transportation Engineering,*
490 *Part B: Pavements* 149 (1): 03122003. https://doi.org/10.1061/JPEODX.0000403.

491 Rigatti, Steven J. 2017. “Random Forest.” *Journal of Insurance Medicine* 47 (1): 31–39.
492 https://doi.org/10.17849/insm-47-01-31-39.1.

Rosenblatt, F. 1958. “The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain.” *Psychological Review* 65 (6): 386–408. https://doi.org/10.1037/h0042519.

Saha, Sajib, Fan Gu, Xue Luo, and Robert L. Lytton. 2018. “Use of an Artificial Neural Network Approach for the Prediction of Resilient Modulus for Unbound Granular Material.” *Transportation Research Record* 2672 (52): 23–33. https://doi.org/10.1177/0361198118756881.

Schwartz, Charles W., Zahra Afsharikia, and Sadaf Khosravifar. 2017. *Standardizing Lightweight Deflectometer Modulus Measurements for Compaction Quality Assurance*. MD-17-SHA-UM-3-20.

Simon, Doug P, Jacqueline LaBelle, and LLC HDL Engineering Consultants. 2023. *Lightweight Deflectometer for Quality Assurance of Compacted Sublayers and Earthwork*. HFHWY00274/000S(964).

Sweet, Lily-belle, Christoph Müller, Mohit Anand, and Jakob Zscheischler. 2023. “Cross-Validation Strategy Impacts the Performance and Interpretation of Machine Learning Models.” Artificial Intelligence for the Earth Systems. *Artificial Intelligence for the Earth Systems* 2 (4). https://doi.org/10.1175/AIES-D-23-0026.1.

Vennapusa, Pavana Kumar Reddy. 2008. “Investigation of Roller-Integrated Compaction Monitoring and In-Situ Testing Technologies for Characterization of Pavement Foundation Layers.” In *Dissertation Abstracts International. Vol. 70, No. 01, Suppl. B, 151 p. 2008.* Doctoral dissertation, Iowa State University.

Virtanen, Pauli, Ralf Gommers, Travis E. Oliphant, et al. 2020. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python.” *Nature Methods* 17 (3): 261–72. https://doi.org/10.1038/s41592-019-0686-2.

Wallach, Daniel, David Makowski, James W. Jones, and François Brun. 2019. “Chapter 9 - Model Evaluation.” In *Working with Dynamic Crop Models: Methods, Tools and Examples for Agriculture and Environment*, 3rd Edition, edited by Daniel Wallach, David Makowski, James W. Jones, and François Brun. Academic Press. https://doi.org/10.1016/B978-0-12-811756-9.00009-5.

Wang, Hailin, Jinsong Qian, Yang Liu, and Junlin Zhang. 2024. “Evaluating Subgrade Compaction for Different Soils Using Nondestructive Lightweight Deflectometer.” *Journal of Materials in Civil Engineering* 36 (5): 04024080. https://doi.org/10.1061/JMCEE7.MTENG-16691.

Yousif, Nada A., Alaa M. Shaban, and Raid R. Almuhanna. 2024. “Use of Light Weight Deflectometer in Characterizing Compaction Quality of Sand Soils.” *AIP Conference Proceedings* 2864 (1): 030003. https://doi.org/10.1063/5.0186559.

Zhang, Guoqiang, B. Eddy Patuwo, and Michael Y. Hu. 1998. “Forecasting with Artificial Neural Networks: The State of the Art.” *International Journal of Forecasting* 14 (1): 35–62. https://doi.org/10.1016/S0169-2070(97)00044-7.

Zhao, Guangyuan, Yao Yao, Shuo Li, Yi Jiang, and Purdue University. Joint Transportation Research Program. 2018. *Maximum Allowable Deflection by Light Weight Deflectometer and Its Calibration and Verification*. FHWA/IN/JTRP-2018/21. https://doi.org/10.5703/1288284316866.

Zhi-peng, LI. 2020. “The Analysis of The Impact of Liquid Limit on Mechanical Properties of Clayey Soil.” *IOP Conference Series: Earth and Environmental Science* 525 (1): 012036. https://doi.org/10.1088/1755-1315/525/1/012036.

Zou, Jinming, Yi Han, and Sung-Sau So. 2008. “Overview of Artificial Neural Networks.” In *Artificial Neural Networks*, edited by John M. Walker, vol. 458, edited by David J. Livingstone. Humana Press.