



Predicting the maximum dry density and optimum moisture content from soil index properties using efficient soft computing techniques

Hunar Farid Hama Ali¹ · Bashdar Omer² · Ahmed Salih Mohammed³ · Rabar H. Faraj¹

Received: 13 May 2023 / Accepted: 25 March 2024 / Published online: 15 April 2024
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2024

Abstract

This study aims to achieve the densest possible state of soil for constructing dams and roads. This requires assessing the compaction characteristics, Optimum Moisture Content (OMC), and Maximum Dry Density (MDD) to determine the soil's suitability for earthworks. However, this process is resource-intensive and time-consuming. To streamline the assessment, the study incorporates six parameters: gravel (G), sand (S), fine (F) contents, plastic limit (PL), liquid limit (LL), and plasticity index (PI). Four different models are used to predict compaction characteristics: artificial neural network (ANN), nonlinear regression (NLR), linear regression (LR), and multilinear regression (MLR). The study utilized a substantial dataset of 2162 entries, considering various soil gradation and plasticity properties as input variables. To evaluate the models' effectiveness, several statistical measures, including coefficient of determination (R^2), scatter index (SI), root mean squared error (RMSE), mean absolute error (MAE), a20-index, and Objective (OBJ) value, were employed. The ANN model outperformed other models in predicting OMC, with RMSE, MAE, OBJ, SI, a20-index, and R^2 values of 3.51, 2.31, 4.26, 0.202, 0.7, and 0.92%, respectively. However, for predicting MDD, the ANN model had the highest R^2 value ($R^2 = 0.87$), but the minimum RMSE (1.01), MAE (0.8), a20-index (0.998), and OBJ (1.07) were obtained from the MLR and LR models. Furthermore, sensitivity analyses revealed that the plastic limit significantly influences the OMC, while the gravel content plays a dominant role in predicting MDD.

Keywords OMC · MDD · Standard proctor · Plasticity · Gradation · Models

1 Introduction

Soil compaction is critical in geotechnical engineering, particularly in construction projects like roads and embankments. Accurate prediction of the Maximum Dry Density (MDD) and Optimum Moisture Content (OMC) is essential to achieve the desired level of compaction. In recent years, researchers have increasingly turned to soft computing models and artificial neural networks (ANN) to improve the accuracy of such predictions. This literature review provides an overview of the existing research in this field.

Soil compaction plays a vital role in geotechnical engineering by enhancing the mechanical properties of soil. Proper compaction ensures structural integrity and prevents settlement and instability [18].

Historically, compaction characteristics were predicted using empirical and semi-empirical methods, such as the Proctor test. These methods, while practical, often lack the

✉ Hunar Farid Hama Ali
hunar.hamaali@uoh.edu.iq

✉ Ahmed Salih Mohammed
ahmed.mohammed@univsul.edu.iq

Bashdar Omer
bashdar.omer@uor.edu.krd

Rabar H. Faraj
rabar.faraj@uoh.edu.iq

¹ Department of Civil Engineering, University of Halabja, Halabja, Iraq

² Civil Engineering Department, University of Raparin, Rania, Sulaymaniyah, Iraq

³ Civil Engineering Department, College of Engineering, University of Sulaimani, Kurdistan, Iraq

accuracy and flexibility required for various soil types and conditions [57].

Soft computing models, including artificial neural networks (ANN), fuzzy logic, and genetic algorithms, have gained popularity for predicting soil compaction characteristics. These models can handle complex, non-linear relationships within large datasets [116].

ANN is a popular choice for predicting soil properties due to its ability to learn from data and adapt to various soil types. Studies have successfully applied ANN to predict MDD and OMC from soil index properties [83].

Previous research has shown that the accuracy of predictions is highly dependent on the selection of input parameters. Parameters like gravel (G), sand (S), plastic limit (PL), liquid limit (LL), and plasticity index (PI) have been identified as significant predictors of compaction characteristics.

Researchers often evaluate the performance of predictive models using various statistical metrics, including R-squared (R²), root mean squared error (RMSE), mean absolute error (MAE), and scatter index (SI). These metrics help assess the accuracy and reliability of the models.

Recent studies have focused on enhancing the efficiency and accuracy of predictive models. Some have incorporated advanced optimization techniques and ensemble modeling to improve results.

The choice of an appropriate model is essential and varies based on the specific application. ANN may perform well in some cases, while linear or multilinear regression models may be more suitable in others. Understanding each model's strengths and limitations is crucial.

Accurate predictions of MDD and OMC from soil index properties have practical implications in geotechnical engineering, enabling engineers to optimize construction processes, reduce costs, and enhance the quality and durability of infrastructure projects. There are several methods to stabilize soils. One of the most significant methods to densify the soil is compaction. Soil compaction results in the decrease of the voids between soil particles. Consequently, soil shear strength observably increases along with the declining compressibility and permeability of the soil [33]. To determine both compaction characteristics, Maximum Dry Density (MDD) and Optimum Moisture Content (OMC), in the laboratory, either the Standard Proctor (SP) or Modified Proctor (MP) methods are used [26].

To assess the suitability of a soil for an earthwork, the compaction characteristics of the soil must be identified. For an earthwork project, large quantities of soil are necessitated. Obtaining the massive volume with a desired compaction characteristic from a single borrow source

might be tedious and time-consuming. Compaction characteristics must be obtained from a laboratory compaction test to determine the suitability of soils collected from various borrow sources. Nevertheless, for laboratory compaction tests, sufficient effort and time are required. Consequently, for any such project, the suitability of the required soils is preliminarily assessed by establishing correlations between the compaction characteristics and simple physical properties obtained through simple index tests [104]. There have been several attempts to determine the compaction characteristics indirectly. For this reason, several correlations have been developed to estimate the compaction characteristics through soil index properties [7, 23, 28, 31, 44, 51, 52, 89, 104, 117] and because of soil fractions [45, 99]. Some studies indicated that an individual input parameter could not sufficiently predict the compaction characteristics. Therefore, multiple linear regression (MLR) models were utilized to predict the compaction characteristics considering different basic soil parameters [9, 24, 26, 45, 49, 67, 98]. In addition to this, the machine learning technique was used in some other studies to develop more accurate correlations [48, 79, 101, 110]. An early study employing artificial neural networks (ANN) by Shook and Fang [97] to predict the soil compaction characteristics indicated that the ANN technique outperforms the traditional statistical models, and a reliable prediction can be attained. Further, Sivrikaya and Soykan [100] utilized the ANNs for estimating the compaction characteristics of 108 fine-grained soil samples.

Sivrikaya [98] performed the multilinear regression model (MLR) to predict the compaction characteristics of fine-grained soils from soil index properties and soil gradations. The considered input parameters were combinations of gravel content (G), sand content (S), fine-grained content (F), plasticity index (PI), liquid limit (LL), and plastic limit (PL). The study concluded that the compaction characteristics could correlate with Plastic limit well compared to other index properties. [31] employed different techniques: simple—multiple analysis and artificial neural networks to forecast the compaction characteristics based on the soil gradations. The study highlights that reliable correlations ($R^2 = 0.70\text{--}0.95$) for preliminary design can be achieved from both techniques. Furthermore, [67] developed multiple regression analysis models for experimental data collected from 110 sandy soils to predict the compaction characteristics in terms of the uniformity coefficient (Cu) and Compaction Energy (CE). Considering the liquid limit (LL), the plasticity index (PI), and compaction energy (CE). Tenpe and Kaur [108] Investigated artificial neural network (ANN) modeling performance for forecasting compaction parameters based on soil index properties. Moreover, Omar et al. [77] utilized advanced mathematical models and an innovative solution to predict

the compaction characteristics of fine-grained soils from numerous physical properties. In this study, multiple linear regression (MLR), artificial neural networks (ANNs), and support vector regression (SVR) are employed. In addition, Farooq et al. [26] utilized a multiple regression model to forecast the optimum moisture content. Also, Saikia et al. [89] developed a set of regression models for predicting the compaction characteristics concerning the consistency limits.

In their study, Karimpour-Fard et al. [53] utilized artificial neural networks (ANNs) and multilinear regression (MLR) techniques across 728 datasets to develop predictive models for compaction characteristics. These models were designed to estimate compaction characteristics based on soil type, grain size distribution (including LL, PL), and specific gravity at varying energy levels. Their findings emphasized the significant impact of fine content on compaction characteristics compared to other factors. Additionally, they underscored the potential advantages of MLR models in predicting compaction characteristics despite the greater effectiveness of ANN models, attributing this to the inherent “black box” nature of ANN models. Taking into account input parameters such as compaction energy and plastic limit, these models were evaluated.

A novel application of artificial neural network (ANN) was utilized in a recent study Verma and Kumar [110] to estimate fine-grained soil’s modified Proctor compaction parameters. From a highway construction work site, from 532 in situ soil samples, several geotechnical parameters were obtained from the laboratory testing. Besides the index properties test, modified Proctor compaction tests were conducted on the collected soil samples. Python V3.7.9 platform was adopted to write the ANN algorithm code for the analysis. Such useful geotechnical parameters as gravel (%), sand (%), fine content (FC), and percent material retained on 2.0 mm (R2.0 mm), 0.425 mm (R0.425 mm), and 0.075 mm (R0.075 mm), coarse sand (CS), medium sand (MS), fine sand (FS), liquid limit (LL), plastic limit (PL), and plasticity index (PI) were considered as input parameters. Verma and Kumar [111] focused on developing a multi-layer perceptron neural network model to predict the modified compaction characteristics of coarse and fine-grained soils. Similarly, the Python V3.7.9 platform was adopted to write artificial neural network (ANN) algorithm code. One hundred seventy-nine datasets of coarse-grained and 69 datasets of fine-grained soils are examined, considering gravel (%), sand (%), FC (%), LL (%), PL (%), PI (%) as input parameters to predict the modified compaction characteristics. In the developed models, the high correlation coefficient (R) was obtained; as the value was more than 0.80 and 0.90 for coarse-grained and fine-grained soil respectively.

In the context of existing literature, various statistical and machine learning models were employed to forecast compaction characteristics based on diverse soil properties, primarily for assessing soil suitability in earthworks. Statistical models offer ease of use, delivering output predictions in equations that can prove valuable for practical field applications. On the other hand, machine learning models excel in processing substantial datasets and identifying nuanced trends and patterns that might elude human perception. Machine learning algorithms exhibit exceptional capability in handling complex, multi-dimensional, and multifaceted data, even in dynamically changing and uncertain conditions, as demonstrated by studies conducted by Mahmood et al. [61], Piro et al. [82], Abdalla and Mohammed [2], and Hama Ali [34]. These models have been employed in a few research studies to predict compaction characteristics.

Furthermore, in rare studies, large quantities of data covering soil index properties and particle—size were considered. In this work, a large volume of the dataset (2162 datasets) has been compiled from previous studies. Four different models of Linear Regression, Nonlinear Regression (NL), Multilinear Regression (ML), and Artificial Neural Network (ANN) have also been utilized to predict the compaction characteristics (OMC, MDD) concerning soil index properties and soil particle—sizes. In addition to predicting the compaction characteristics, this work evaluates the performance of the utilized models.

1.1 Related studies

Using various modeling techniques, the literature review provides valuable insights into predicting soil compaction characteristics, specifically the Maximum Dry Density (MDD) and Optimum Moisture Content (OMC). Below are the essential findings and trends observed in the related studies: soil compaction is a critical aspect of geotechnical engineering, ensuring the structural integrity of construction projects such as roads and embankments. Accurate predictions of MDD and OMC are essential for achieving the desired level of compaction. Researchers have increasingly turned to soft computing models, including Artificial Neural Networks (ANN), to improve the accuracy of predictions. These models are valued for handling complex, non-linear relationships within large datasets. Traditional methods, such as the Proctor test, were historically used for predicting compaction characteristics. However, these methods often lack accuracy and flexibility across different soil types and conditions. The accuracy of predictions depends significantly on the selection of input parameters. Parameters like gravel content, sand content, plastic limit, liquid limit, and plasticity index have been identified as significant predictors of compaction

characteristics. Researchers employ various statistical metrics, including R-squared (R^2), root mean squared error (RMSE), mean absolute error (MAE), and scatter index (SI), to evaluate the performance of predictive models. These metrics help assess the accuracy and reliability of the models. Recent studies have focused on enhancing the efficiency and accuracy of predictive models. Some have incorporated advanced optimization techniques and ensemble modeling to improve results. Choosing the appropriate model is essential and depends on the specific application. While ANN performs well in some cases, linear or multilinear regression models may be more suitable in others. Understanding the strengths and limitations of each model is crucial. Accurate predictions of MDD and OMC from soil index properties have practical implications in geotechnical engineering, allowing engineers to optimize construction processes, reduce costs, and enhance the quality and durability of infrastructure projects. Some studies have employed machine learning techniques, such as ANNs, to predict soil compaction characteristics, with promising results regarding model performance and accuracy. Large datasets and model evaluation: the work under review stands out for its use of a large dataset (2162 datasets) and for applying four different models, including ANN, to predict compaction characteristics. The study not only predicts these characteristics but also rigorously evaluates the performance of the utilized models. Comparison of soft computing models and statistical models: the literature reveals a trend of comparing soft computing models (e.g., ANN) with traditional statistical models (e.g., linear regression) to determine which one offers better predictive capabilities for soil compaction characteristics. Several studies have indicated that relying on a single input parameter is insufficient for accurate predictions. Hence, models that consider multiple basic soil properties, especially those related to particle size distribution and consistency limits, have been developed and found to provide more reliable predictions. In summary, the reviewed literature emphasizes the importance of accurate predictions of MDD and OMC for soil compaction in geotechnical engineering. Soft computing models, particularly ANNs, have shown promise in improving the accuracy of predictions. The choice of input parameters, model evaluation metrics, and the practical implications of these predictions are significant considerations. Additionally, recent studies have incorporated advanced techniques to enhance the efficiency and accuracy of predictive models. The field is continually evolving, with an increasing focus on harnessing the potential of machine learning and large datasets for more precise predictions.

2 Scope of the work

In light of the literature, an individual input parameter is unlikely to be applicable to predict the compaction characteristics of different soil types. Therefore, numerous datasets from the literature were compiled to develop models to predict the compaction characteristics of soil by including simple soil properties such as the G, S, F, LL, PL%, and PI%. Subsequently, four different models were developed. This enables evaluating the models' performance and the identification of the effects of various soil properties. Consequently, the datasets in two groups of training and testing are examined in the models to reach the main objectives of the study:

- (i) The influence of the physical soil properties on the compaction characteristics will be investigated. The input parameter with the most influential role in measuring the value of the OMC and MDD will be assigned from a sensitivity analysis.
- (ii) Concerning the statistical evaluation tools, the model outperforming the prediction of the compaction characteristics of soil from basic soil properties will be determined and compared with the other models.

The major contributions and novelty in this work can be summarized as follows:

Multi-parameter approach: the study's innovation lies in adopting a multi-parameter approach to predict the compaction characteristics of soils. Instead of relying on a single input parameter, it incorporates six different parameters: gravel (G), sand (S), fine content (F), plastic limit (PL), liquid limit (LL), and plasticity index (PI). This approach recognizes that a single parameter may not be universally applicable across different soil types, and therefore, a more comprehensive set of input variables is considered.

Model Diversity: The work employs four distinct models: artificial neural network (ANN), nonlinear regression (NLR), linear regression (LR), and multilinear regression (MLR) models. This diversity in modeling techniques allows for a robust assessment of each model's predictive capabilities and performance, thus ensuring a comprehensive evaluation of the data.

Extensive dataset compilation: to support the development of these models, a comprehensive dataset of 2162 entries is compiled, encompassing various soil gradation and plasticity properties as input variables. This dataset's size and comprehensiveness contribute to the reliability and generalizability of the models.

Statistical evaluation tools: the study employs a range of statistical evaluation tools, including the coefficient of

determination (R^2), scatter index (SI), root mean squared error (RMSE), mean absolute error (MAE), a20-index, and Objective (OBJ) value. These tools provide a thorough assessment of the models' performance and ability to predict compaction characteristics accurately. Identification of influential parameters: through sensitivity analyses, the study identifies the specific input parameter with the most significant influence on measuring the values of Optimum Moisture Content (OMC) and Maximum Dry Density (MDD). This insight can be crucial for understanding the key factors affecting soil compaction characteristics.

Comparative analysis: the research develops these models and rigorously compares their performance. It concludes that the ANN model excels in forecasting OMC, while for MDD prediction, the MLR and LR models outperform the ANN model. This comparative analysis aids in selecting the most appropriate model for specific applications also, Given the study's objective to streamline the assessment of soil compaction characteristics for constructing dams and roads, along with the use of multiple parameters and models, a novel method could involve the following approaches:

Hybrid model integration: combine the strengths of different models to improve prediction accuracy. For instance, a hybrid model that integrates the advantages of artificial neural networks (ANN) in capturing complex relationships with the interpretability of linear regression (LR) or multilinear regression (MLR) could be developed. This hybrid model could potentially enhance predictions by leveraging the ANN's ability to capture nonlinear patterns and the interpretability of LR/MLR for a better understanding of the impact of individual parameters.

Feature selection techniques: utilize advanced feature selection methods to identify the most influential parameters. Techniques like recursive feature elimination, feature importance ranking from tree-based models, or Lasso regularization can help narrow down the most critical parameters affecting soil compaction characteristics. This could streamline the assessment process by focusing on key parameters and reducing resource consumption.

Ensemble learning: implement ensemble techniques such as bagging, boosting, or stacking to combine predictions from multiple models. Ensemble methods often lead to better generalization and robustness by leveraging the strengths of different models. This approach could enhance prediction accuracy for Optimum Moisture Content (OMC) and Maximum Dry Density (MDD).

Sensitivity analysis optimization: utilize sensitivity analysis to identify influential parameters and optimize the assessment process. By understanding the impact of

different parameters on OMC and MDD, the study can propose specific guidelines or thresholds that indicate when certain parameters have a significant effect. This could help streamline decision-making in soil compaction assessments.

Automated Data Processing: Implement automation in data processing and model selection. Leveraging machine learning pipelines or frameworks can streamline data preprocessing, feature engineering, model training, and evaluation, reducing the time and resources needed for analysis.

Combining the strengths of different modeling techniques, optimizing parameter selection, leveraging ensemble methods, and automating processes can streamline the assessment of soil compaction characteristics, making it more efficient and less resource-intensive while maintaining or improving prediction accuracy.

2.1 Components in the block diagram

- Input data: this represents the dataset containing soil properties such as Gravel (G), Sand (S), Fine content (F), plastic limit (PL), liquid limit (LL), and plasticity index (PI).
- Preprocessing module: this stage involves data cleaning, normalization, and feature scaling to prepare the input data for modeling.
- Modeling phase:
 - LR model: linear regression model based on soil physical properties.
 - NLR model: nonlinear regression model.
 - MLR model: multilinear regression model.
 - ANN model: artificial neural network model.
- Evaluation and comparison: this part includes the assessment of model performance using various metrics like R-squared (R^2), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Scatter Index (SI), and Objective (OBJ) values.
- Results and analysis: this section interprets and analyzes the model outputs, compares predictions, and identifies influential parameters for OMC and MDD.
- Sensitivity analysis module: this component pinpoints the significant parameters affecting OMC and MDD predictions.
- Conclusion and recommendations: this stage presents conclusions drawn from the analysis, recommendations for model selection, and potential areas for further research.

3 Methodology

Overall, 2162 datasets from the literature were gathered. The data were randomly mixed and divided into two groups: training datasets and testing datasets. The training datasets comprised 70% of the dataset, while 30% was for the testing ones. The training data groups were used to develop the models. The foremost objective of the models was to predict the standard Proctor compaction characteristics (OMC and MDD). Then, using the testing data, the models were evaluated. Table 1 includes the number of used data from different studies and ranges of the input parameters: gravel content (G %), sand content (S %), fine content (F %), liquid limit (LL %), plastic limit (PL %) and plasticity index (PI %).

Further, in the table, the ranges of the measured values of optimum moisture content (OMC %) and maximum dry density (MDD kN/m^3) are contained within, compared with the predicted values obtained from the models. These input parameters are used to develop the models and the actual values of output parameters. The procedure of the work is illustrated in a flowchart in Fig. 1 and summarized as:

In total, 2162 datasets were compiled from the available literature. These datasets were randomly combined and then divided into two sets: one for training the models, consisting of 70% of the data, and the other for testing the models, which accounted for the remaining 30%. The primary aim of these models was to forecast the standard Proctor compaction characteristics, specifically the Optimum Moisture Content (OMC) and Maximum Dry Density (MDD). Subsequently, the models' performance was assessed using the testing data. Table 1 presents the quantity of data sourced from various studies and the ranges of the input parameters, including gravel content (G %), sand content (S %), fine content (F %), liquid limit (LL %), plastic limit (PL %), and plasticity index (PI %).

Furthermore, Table 1 provides a ranged of measured values of OMC (%) and MDD (kN/m^3), with which the obtained values from the models will be compared later. These input parameters are the foundation for model development, while the actual output parameter values are also included. The workflow of the study is visually represented in Fig. 1 and can be summarized as follows:

Stage 1: Collecting data.

Stage2: Correlating input and output parameters.

Stage 3: Splitting data into two groups: 70% training and 30% testing.

Stage 4: Developing different models.

Stage 5: Evaluating the performance of the models.

4 Correlation between soil index properties and compaction parameters

The illustration demonstrates the validity of the connection between input parameters and compaction characteristics. Each input parameter was graphically depicted and scrutinized concerning OMC and MDD. Figure 2 presents a matrix plot elucidates the link between input parameters and compaction characteristics. The figures unmistakably reveal that, except for the correlation between PL and OMC, with an R-value of 0.82 (as shown in Fig. 3), there are generally insufficient correlations between compaction characteristics and the input parameters. This particular correlation can be considered robust but may apply primarily to fine-grained soil. In other words, PL may not accurately represent the soil's true properties for soils where the percentages of G (gravel) and S (sand) dominate. Moreover, it is worth noting that some types of purely coarse-grained soil may not possess a plastic limit (PL) at all.

Additionally, Fig. 4 illustrates the typical distribution of OMC and MDD. Moreover, Table 2 presents various statistics, including minimum, maximum, mean, standard deviation, skewness, kurtosis, and variance. In the context of kurtosis, a strongly negative value suggests shorter tails compared to a normal distribution, while a positive value indicates longer tails. As for the skewness parameter, a negative value signifies an extended left tail, while a positive value signifies an extended right tail.

5 Modeling

In the above figures, each parameter of G, S, F, LL PL, and PI is connected to the OMC and MDD so that a direct relationship between these parameters can be developed to predict the compaction characteristics from one of these parameters. However, because the R^2 value is extremely low, a reliable direct correlation between the compaction characteristics and the input parameters cannot be attained, following the statistical analysis and Figs. 2, 3 and 4. Therefore, four separate models are developed considering the influence of soil gradation and index properties to overcome this shortcoming and establish a reliable correlation to forecast OMC and MDD from the input parameters.

In this work, the models are utilized to predict the OMC and MDD, and each model's performance is evaluated according to the original data. The following evaluation criteria are used to evaluate and compare the performance of the four models: For a model to be scientifically accurate, there should be a minor percentage error between

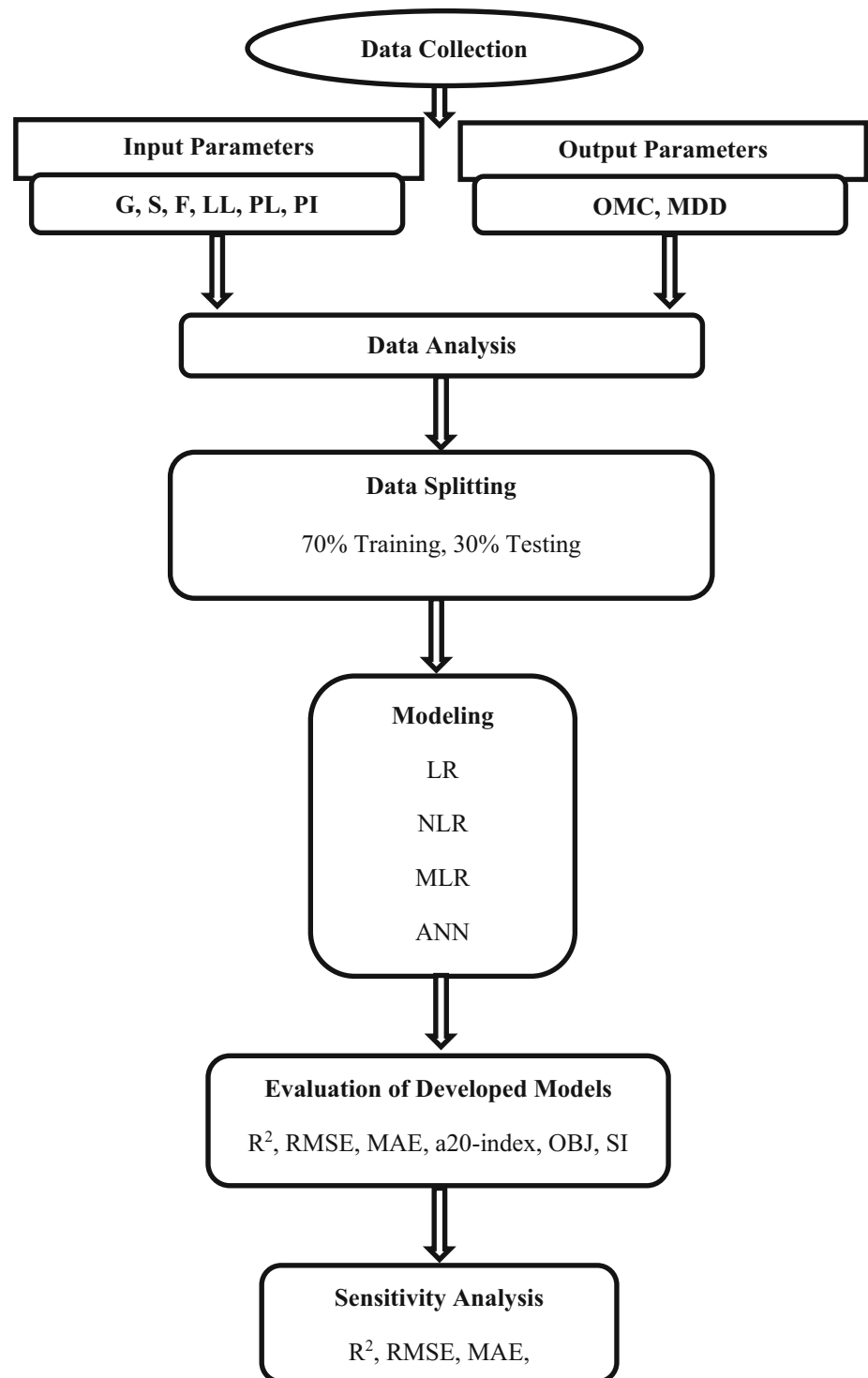
Table 1 The range and the number of the used datasets

Authors	Data ranges								
	No. data	Gravel (%)	Sand (%)	Fines (%)	LL (%)	PL (%)	PI (%)	OMC (%)	MDD (kN/m ³)
Shook and Fang [96]	59	0–11.8	17–90.5	8.0–81	23.5–34	10.3–18	9.1–18	9.5–15.5	17.3–20.8
Ring et al. [88]	10	0–0	0–50	50–100	24–72	15–37	2.0–40	14–31	14–18.7
Johnson and Sallberg [50]	51	0–0	5.0–94	6.0–100	16–76	16–107	0–45	6.6–32	9.9–21.3
Felt [27]	6	0–0	8.0–81	19–92	26–58	19–25	4.0–36	10–20.8	15.9–20.1
Harris [37]	296	0–26	0–62	38–100	17–111	7.0–42	1.0–80	8.9–41	12–20.4
Hopkins [41]	7	0–7.7	4.7–24	71–91	14.5–61	1.0–27	1.0–34	13.5–24	15.1–18.7
Ramiah et al. [86]	16	0–0	8.4–67.8	32.2–91.6	26.6–72.6	10.5–38.1	13.4–39.4	14.8–33	12.7–18.1
Wesley [114]	27	0–0	2.0–86	14–98	44–213	33–167	2.0–58	20–160	4.4–14
Horz [43]	5	0–27.5	2.5–66.1	6.4–97.5	18–61	13–21	5.0–41	8.7–23.2	15.5–19.8
Wang and Huang [112]	57	0–20	0–100	0–100	24–495	10.0–47	2–449	8–32.5	12.6–20.8
Foreman and Daniel [30]	3	0–0	0–18	82–100	48–56	14–31	23–42	18.1–31.5	13.6–16.8
Sridharan et al. [105]	15	0–0	5.0–60	40–95	43–123	22–50	21–100	20.4–49	10.2–16
Al-Khafaji [5]	88	0–0	1.0–52	48–99	18–66	12–29	6.0–39	9.0–26	14.6–20.3
Benson and Trast [14]	13	0–8	6.0–48	52–94	24–70	12.0–32	11.0–46	10–24	15.4–20.4
Mohammad et al. [65]	8	0–0	0–100	0–100	19–93	12.0–29	4.0–64	8.9–27.7	14.2–19.4
Burrourghs [17]	7	0–0	5.0–94	6.0–95	16–67	16–27	0–40	7–20.9	15.3–21.4
Albrecht and Benson [6]	8			43–94	18–67	7.0–35	11.0–46	9.0–21	15.5–20.6
Gurtug and Sridharan [33]	5	0–0	1–13	87–99	28.2–98	21.1–40	7.1–58	18.6–32	12.5–16.9
Miller et al. [64]	4	0–29	0–56	44–98	16–83	9–23	7.0–60	9.0–29.1	14.7–20
David Suits et al. [20]	3	0–0	2.0–56	44–98	16–81	9–26	7.0–55	9–25	15.3–20.5
Luczak-Wilamowska [59]	5	0–0	7.0–87	13–93	20.6–48.6	16.9–21.3	2–27.3	11.1–23.3	18.1–20.2
Sridharan and Nagaraj [104]	10	0–0	0–36.5	63.5–100	37–73.5	18–51.9	9.5–37.9	16.2–44.4	11.1–17.9
Daita et al. [19]	9	0–0	10.0–50	50–90	24–48	15–21	9.0–29	12.5–20.5	16.2–19.1
Bellezza and Fratolocchi [13]	15	0–0	0–79	21–100	31–102	18.5–39	7.0–63	11.0–29	13.5–18.8
Olmez [75]	107	0–97	1.0–79	1.0–49	17–83	8.0–43	5.0–40	6.0–32	12.7–22.6
Horpibulsuk et al. [42]	9	0–13.3	0–44.3	55.7–100	39.7–256.3	6.1–48.2	17.2–217.1	15.6–33.8	12.8–17.6
Sivrikaya [98]	44	0–41	1–49.0	41–99	28–74	12–39	12–46	13–40	11.3–18.8
Hong [40]	92	0–11	0–46	53–100	25–48	13–29	10.0–24	10.0–22	14.7–19.9
Liang et al. [58]	2	5.3–5.9	25.3–38.4	56.3–68.8	27.8–30.8	8–12.4	18.4–19.8	14.2–16.5	17.69–17.74
Gunaydin [31]	126	0.1–67.1	15.5–71.3	13–83.3	23.1–56.8	13.7–29.8	7.4–33.3	7.6–26	14–20.5
Sawangsurinya et al. [92]	7	0–3.9	3.1–59	41–96.4	26–85	14–33	9.0–52	13.5–27.5	14.4–18.5
Demiralay and Guresinli [21]	43	0–0	15–67	33–85	20.8–58.8	14–38.2	4.7–36.2	12.8–32.4	13–18.8
Yildirim and Gunaydin (2011)	124	0–78	0.9–49	10–99.1	20–89	11.0–43	3.0–52	7.2–40.2	11.9–21.4
Bera and Ghosh [15]	5	0–0	0.9–19.6	80.4–99.1	30.8–213.3	17.1–44.5	10.3–168.8	15.4–30.1	12.9–17.5
Berney et al. [16]	7	0–52.8	4.9–92	3.1–95.1	15–73	12–24	3.0–49	6.8–24.6	13.5–21.4
Nerea [70]	130	0–35	0.9–47	34–99.1	24–106	15–46	6.0–71	14–42	10.9–19.4
Sotelo [103]	5	0–65.5	2.7–71.9	4.3–97.3	22–86	11.0–33	9.0–53	5.7–25.9	15–23.1
Isik and Ozden [47]	200	0–80	10.0–96	1.0–85	24.8–58.3	15.4–26.6	10.7–37.4	6.2–19.7	16.4–22.1
Ramasubbarao and Sankar [85]	25	0–24	0–40.1	52–100	24.6–94	11.9–36	8.3–58	12.3–35.4	12.3–18.1
Ibrahim [46]	4	0–0	17–24	76–83	38.3–46.4	18.7–21.2	17.9–27	17–17.8	15.8–16.8
Sathawara and Patel [91]	40	0–24	8.0–81	13–92	18–47	12–30	10–29	8.0–19	15.8–21.1
Mejias-Santiago et al. [63]	16	0–41.3	2.5–92	3.1–97.5	21–109	10.0–72	8.0–49	8.0–62	8.7–20.2
Abasi [1]	26	0–0	0–32	68–100	22.5–70	18–31	1.5–42	15–25.7	14.5–18.1

Table 1 (continued)

Authors	Data ranges								
	No. data	Gravel (%)	Sand (%)	Fines (%)	LL (%)	PL (%)	PI (%)	OMC (%)	MDD (kN/m ³)
Kolay and Baser [55]	8	0–0	24.2–80.2	19.8–75.8	30–61.6	16.7–32.5	6.5–32.6	9.8–23.7	14.9–21.5
Ören [78]	9	0–0	3.4–35.1	64.9–96.6	33.4–92.4	17.5–42.3	15.9–50.1	20–36.3	11.8–16.2
Raju et al. [84]	5	0.2–18.5	8.2–63.2	18.3–91.6	30–56	15–20	10.0–41	9–16	17.2–18.4
Talukdar [107]	16	0–4.7	13.4–36.5	58.8–83.2	28.5–36.8	20.2–29	6.1–8.5	14.2–15.8	15.9–17.4
Shirur and Hiremath [95]	20	0–9	22–89	4.0–77	22–66	13–40	10.0–35	10.5–26	14.2–22.6
Wells [113]	43	0–64	2.5–94.4	3.1–97.5	15–109	10.0–72	3.0–53	7.6–27.3	13.5–22.2
Dway and Thant [25]	6	0–0	7.7–22.2	77.8–92.3	41–70.1	16–21.5	25–48.8	18.5–20.6	16.3–17.1
Ng et al. [72]	9	0–26	8–46	34–91	41–53	26–34	15–19	13.5–24	14.7–17.4
Nagaraj et al. [68]	42	0–2	2–80	20–98	24–115	17.4–45.3	3.7–75.6	9.5–36.8	12.6–18.3
Mujtaba [66]	120	0–5	49–100	0–47	18–22	15–18	4–4	11–18	14.8–19.5
Korde and Yadav [56]	8	0–0	0.8–50.4	49.6–99.2	28.9–73	18.1–32	6.2–41	10.9–24.4	16–20
Nesamatha and Arumairaj [71]	20	0–20.3	4.5–43.9	35.9–95	29–77.8	18.2–30	10–55.5	12.0–30	14–19.1
Kaushal and Guleria [54]	3	0–0	9–17.7	41.5–82.3	38.2–48.7	18.2–26.6	20–22.1	17.2–19.6	16.4–18.8
Parkoh [80]	168	0.6–52	11–86.5	8–83.4	19–64	9.5–39.2	0.7–41	8.8–24.5	16.1–24.3
Hussain [45]	52	0–0	7.6–71.6	28.4–92.4	33.8–87.5	15.4–32.3	16.5–55.2	14–24	15.2–20
Winta [115]	124	0.4–5.2	11–34.1	41–97.7	31–50	19–37	3–25.4	15–35	11.9–18.3
Ahmed et al. [4]	3	0–0	3.0–6.0	94–97	38–40	13–14	24–26	13.1–13.5	14.1–15.8
Ali and Fakhraldin [8]	5	0–6.3	76.8–93.6	4.8–23.2	21–29	11.0–15	10.0–14	8.6–13.2	17.5–19
Setiawan [94]	30	0–73.2	1.2–100	0–98.8	20.5–61.4	15.5–41.2	1.9–28.1	12.8–31.8	6.6–18.4
Han and Vanapalli [36]	3	0–0	3–28	72–97	19.6–48	13.6–22	6–26	13.5–23	16.2–19.2
[89]	40	0–1.1	1.5–47	53–98.5	20.4–56.2	10–29.9	3.9–28.4	14.5–31	13.8–18.9
Zhang and Frederick [117]	13	0–0	0–0	100–100	30.5–328.2	19.7–33.5	10.8–294.7	17.5–38.4	11.7–16.6
Tsegaye et al. [109]	56	0.3–9.3	0.6–27.4	68.7–96.7	45–97	22–41	22–58	15.5–37	12.3–17.3
Bekele [12]	31	0–75.8	1.4–8.5	16–99.6	51–106	24–43	21–73	26.3–41	11.7–14.8
Mazari and Nazarian [62]	5	0–8	3.0–73	27–97	23–86	11.0–33	12.0–53	9.4–25.9	15.1–19.6
Sen and Pal [93]	7	0–0	6.6–56.5	43.5–93.4	23–41.7	14.9–24.7	5.4–17	14–22.1	13.9–18
Karakan and Demir [52]	20	0–0	0–66.4	33.6–100	27.5–158.2	18.3–66.5	4.1–91.8	17.5–47.2	11–16.1
Adeoye et al. [3]	4	1–7	41–58	40–52	35–50	19.2–27.5	15.8–26.6	16.5–23.5	15.8–16.9
Taha et al. [106]	6	0–0	32.5–56.6	43.4–67.5	37.8–87.2	17.7–28	17–69	14.8–24.8	14.4–18.1
Saran [90]	13	0–0	3.6–18.9	81.1–96.4	27.3–61.4	19.6–29	7.3–33	16.5–24.5	14.6–17.7
Soltani et al. [102]	4	0–0	1.0–31	69–99	44–78	18–28	22–57	16–25	14.6–16.4
Ratnam and Prasad [87]	8	1.2–9.2	14.9–63.5	35.4–78.7	33–81	17–25	8.0–60	14–17	17.9–19.5
Pillai and Vinod [81]	25	0–0	0–60	40–100	23–227	14.7–39.7	4.0–47	12–29	13.2–18.5
Ogbuchukwu et al. [74]	8	0–2.9	5.2–37	62.1–94.8	63.1–77.4	29.3–37.7	33.6–46.5	17.7–27	13.7–17.2
Hussain and Atalar [44]	8	0–0	17.7–67.8	32.3–82.3	43.2–76.8	18–25.7	22.9–51.1	15.5–22.5	15.6–19.3
Arshid and Kamal [10]	15	0–4.8	2–25.3	72.7–96.4	25.5–33.6	16.7–20.6	7.4–16.1	11.8–15.6	17.9–18.9
Machado et al. [60]	17	0–9	28–82	17–72	25–91	15–41	10.0–50	10.4–29.6	13.8–19.2
Di Sante [22]	7	0–3	1–16	84–99	40–63	20–30	18–37	16–21	16–18.2
Fondjo et al. [29]	15	1.1–16.1	15.9–44	49.5–83	40.3–78.9	17.8–25.3	19.2–56.7	17.2–27.8	15.7–19.6
Hassan et al. [38]	30	0–7	7.0–26	72–93	27–32.3	17–22.6	4.9–12.4	11–13	18.9–19.6
Nagaraju et al. [69]	10	0–25	35–61	20–58	23–42	18–23	5–19	8.3–18	17.3–20
Nwaiwu and Mezie [73]	20	0–8	56.4–80.2	19.8–40.7	20.9–51.3	8.6–32.6	6.5–21	11.2–17.5	16.8–19.5

Fig. 1 The procedure of the study by a flow chart diagram



observed and predicted data (higher a20 – index) and a higher R^2 value with a lower RMSE, Objective (OBJ), MAE, and SI values (Figs. 5, 6, 7, 8, 9, 10, 11, 12).

Complexity analysis of Multiple Linear Regression (MLR) is relatively straightforward compared to more complex models like neural networks. MLR is a simple, linear modeling technique to establish relationships

between dependent and multiple independent variables. Here are the key aspects to consider when analyzing the complexity of an MLR model:

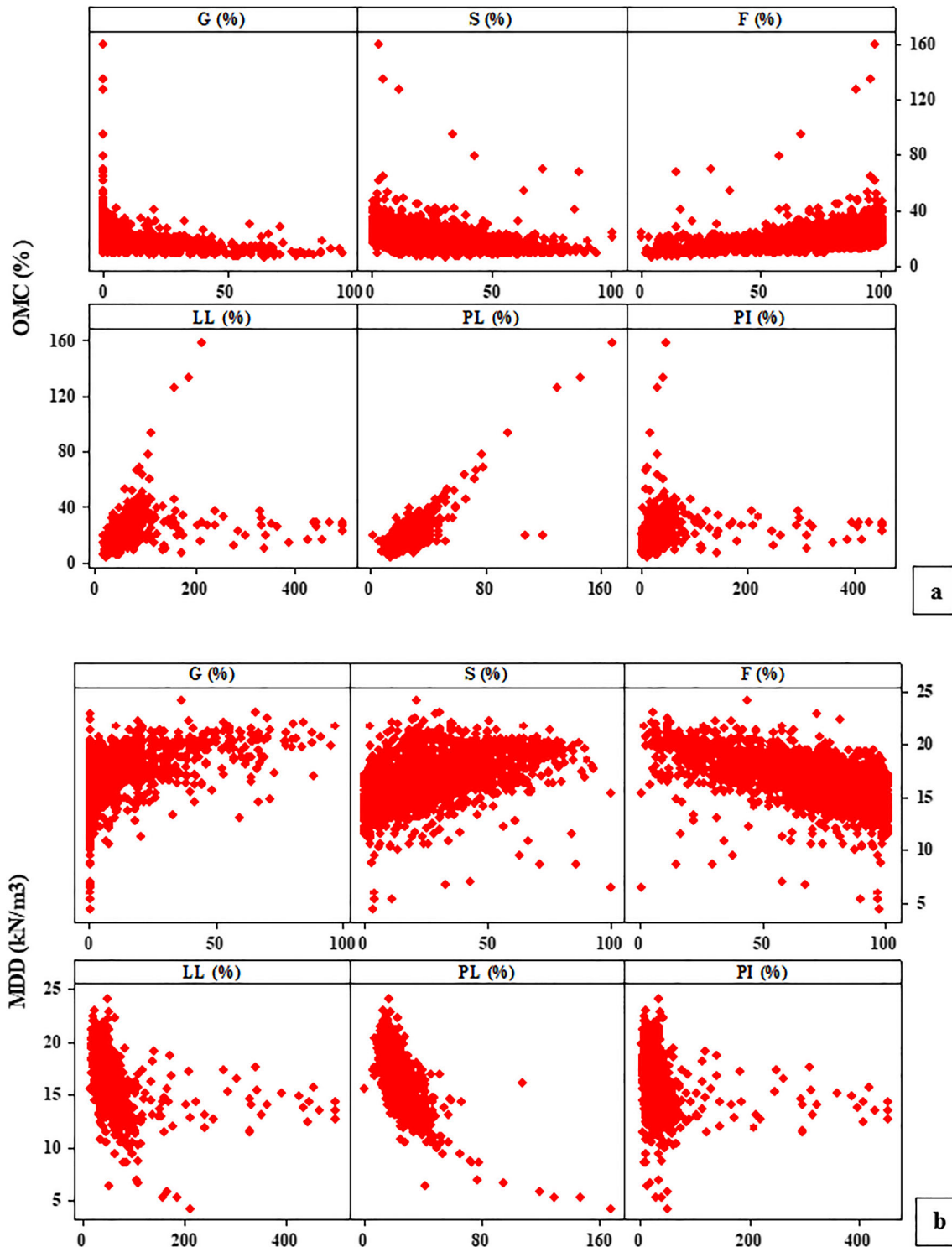


Fig. 2 Matrix Plot between the input parameters and the compaction characteristics

5.1 Model complexity

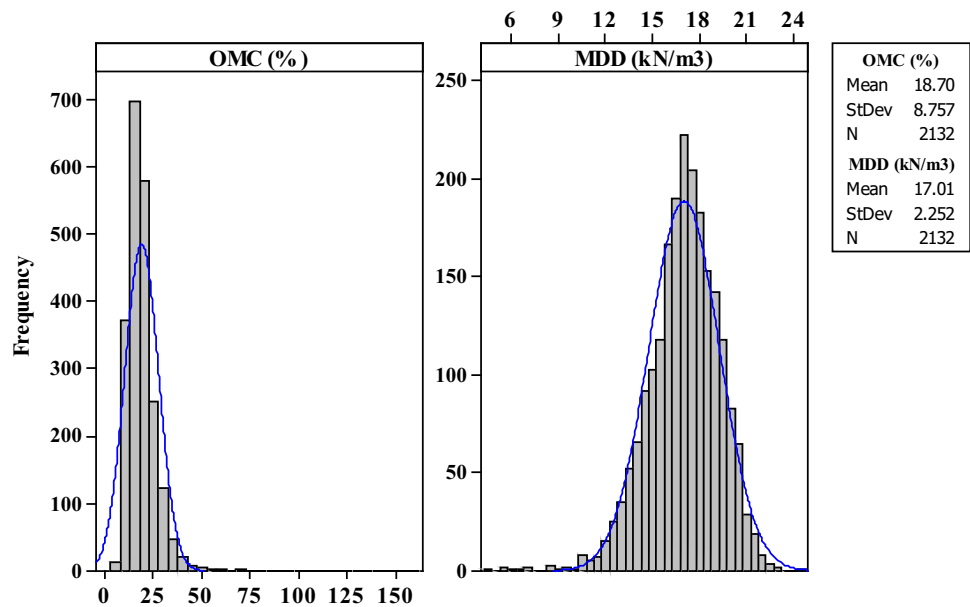
MLR models are linear, assuming a linear relationship between the dependent and independent variables. This simplicity contributes to low model complexity.

The complexity of MLR, LR, and NLR models increases with the number of independent variables (features) included in the model. Adding more components can lead to a more complex model.

Fig. 3 Correlation Matrix between input and output parameter

	G (%)	S (%)	F (%)	LL (%)	PL (%)	PI (%)	OMC (%)	MDD (kN/m3)
G (%)	1							
S (%)	0.08	1.00						
F (%)	-0.65	-0.81	1.00					
LL (%)	-0.08	-0.21	0.21	1.00				
PL (%)	-0.19	-0.24	0.30	0.49	1.00			
PI (%)	-0.04	-0.17	0.16	0.97	0.29	1.00		
OMC (%)	-0.31	-0.36	0.46	0.40	0.82	0.22	1.00	
MDD (kN/m3)	0.45	0.41	-0.58	-0.40	-0.71	-0.26	-0.85	1

Fig. 4 Normal distribution of the compaction characteristics



Each independent variable has a coefficient (weight) associated with it in an MLR model. The number of independent variables determines the total number of parameters in the model. As the number of independent variables grows, so does the number of model parameters.

MLR, LR, and NLR can become more complex if interactions (cross-product terms) or higher-order terms (quadratic or cubic terms) are included in the model. These additions increase the number of parameters and model complexity.

Increasing model complexity by adding more features or terms can lead to overfitting. Overfitting occurs when the model fits the noise in the data rather than the underlying patterns. Complexity should be carefully balanced to avoid overfitting. Proper variable selection is crucial to managing complexity. Including irrelevant or redundant variables can increase model complexity without improving predictive accuracy.

Table 2 Statistical parameters for input and output variables

Nos.	Variables	Min	Max	Mean	Variance	SD	Skewness	Kurtosis
1	G (%)	0	97	7.65	234.79	15.31	2.727	7.84
2	S (%)	0	100	27.13	387.96	19.7	0.745	0.02
3	F (%)	0	100	65.18	670.47	25.9	-0.59	-0.67
4	LL (%)	14.5	495	47.8	1647.65	40.89	6.89	59.12
5	PL (%)	1	119	22.52	70.8	9.732	2.59	16.78
6	PI (%)	0	449	25.35	1373.98	37.04	7.75	70.56
7	OMC (%)	5.7	79.3	18.54	53.02	7.28	1.83	7.29
8	MDD (kN/m ³)	4.41	31	16.99	5.14	2.269	-0.529	2.3

Techniques like L1 (Lasso) and L2 (Ridge) regularization can be applied to penalize large coefficient values to control model complexity and reduce overfitting.

Simplicity in an MLR model can enhance its interpretability. Complex models with numerous variables and interactions can be challenging to interpret.

MLR models are computationally efficient and require minimal memory and processing power resources, making them suitable for large datasets.

Model complexity should be balanced with model performance. Evaluation metrics like R-squared (R²), adjusted R-squared, and RMSE can help assess model fit and complexity.

Multiple Linear Regression is a relatively simple and interpretable modeling technique. Its complexity primarily depends on the number of features and the inclusion of interactions or higher-order terms. Proper variable selection and regularization techniques are essential for managing complexity and preventing overfitting. MLR suits situations where a straightforward linear relationship between variables is sufficient, and model interpretability is a priority.

MLR is straightforward, interpretable, and computationally efficient. It is well-suited when you clearly understand the relationships between variables and want a simple model that can provide insight into how each independent variable affects the dependent variable.

NLR is employed when the relationship between variables is nonlinear, meaning that the effect of an independent variable is not constant but varies across the range of the variable.

NLR allows more flexibility in modeling complex, nonlinear relationships. It can capture relationships that MLR cannot. NLR can provide more accurate predictions when the data suggests a nonlinear relationship.

LR is a simplified form of MLR with only one independent variable. It's used when you want to model the relationship between a dependent variable and a single independent variable linearly.

Why Use LR: LR is helpful when you're interested in understanding the impact of a single variable on the

dependent variable. It's a basic and interpretable model for simple relationships.

ANNs are used when dealing with complex, nonlinear relationships involving many independent variables. ANNs can capture intricate patterns and interactions within the data.

Researchers and data scientists often start with simpler models like MLR or LR to understand the data and establish a baseline. If the relationships are nonlinear or involve complex interactions, NLR or ANN might be more appropriate. Model selection is often an iterative process involving testing multiple models to determine which best fits the data and produces accurate predictions.

5.2 Linear regression model

The linear regression model (LR) [7, 31, 32, 34, 35, 44], as demonstrated in Eq. (1), is the most common technique for predicting the compaction characteristics of soils:

$$\text{OMC, MDD} = a + bX \quad (1)$$

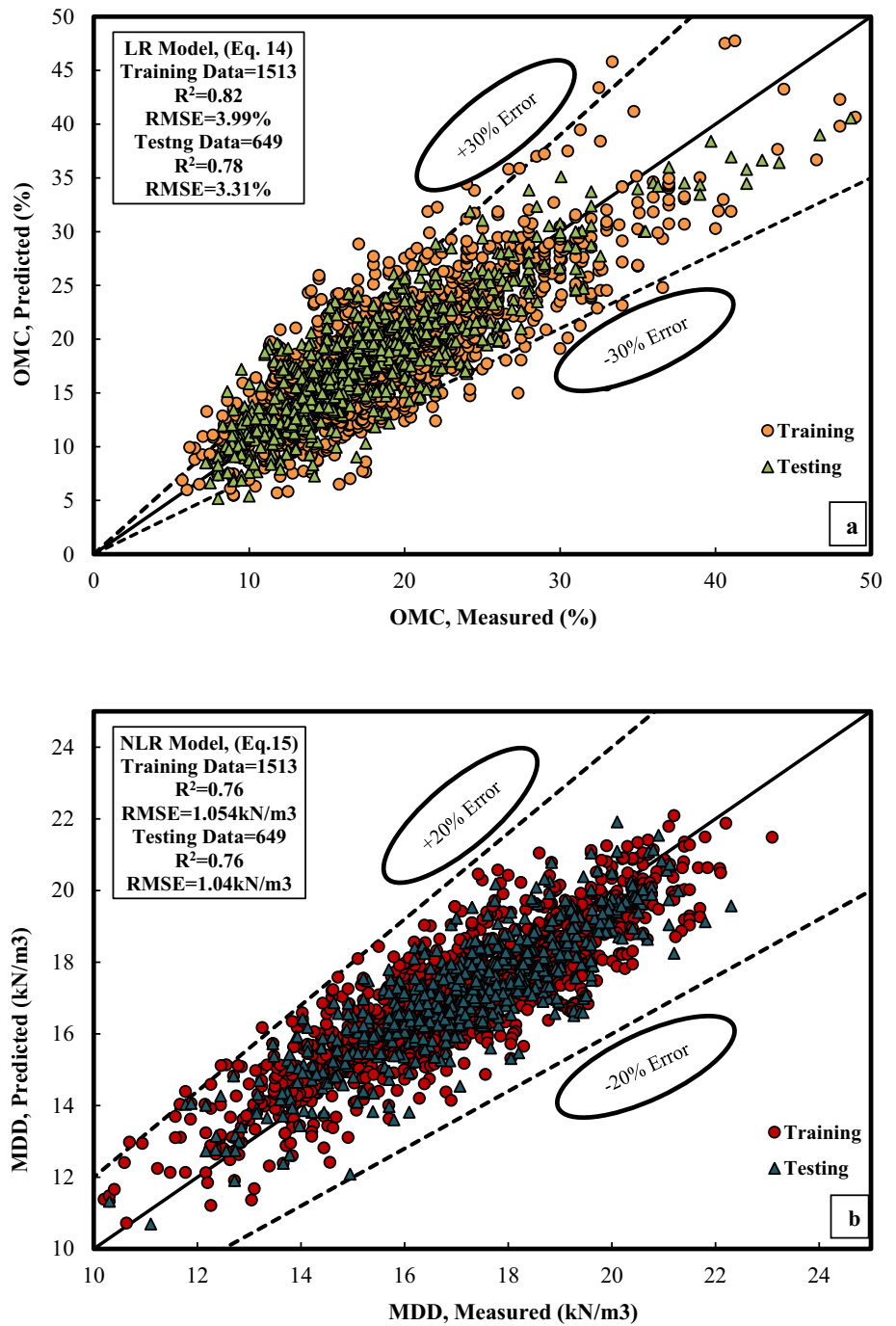
A and b are constants, and x might be one of the G, S, F, LL, PL, or PI. The previous formula does not include the combination of other variables that might affect OMC and MDD, including particle size and soil plasticity. Equation (2) is presented to combine all different parameters and factors that might influence OMC and MDD to obtain more reliable scientific findings.

$$\text{OMC, MDD} = a + bG + cS + dF + eLL + fPL + gPI \quad (2)$$

where G is gravel content (%), S is sand content (%), F is fine content (%), LL is the liquid limit, PL is the plastic limit, and PI is the plasticity index.

Moreover, the model parameters are a, b, c, d, e, f, and g. All variables can be changed linearly. Hence Eq. (2) can be considered an expansion of Eq. (1). This combination might not always be the case as all factors are unlikely to affect the compaction characteristics and interact with one another. Consequently, the model should be updated often to accurately estimate the OMC and MDD [45, 76, 99].

Fig. 5 Comparison between measured and predicted values of **a** OMC and **b** MDD for LR model: training data and testing data



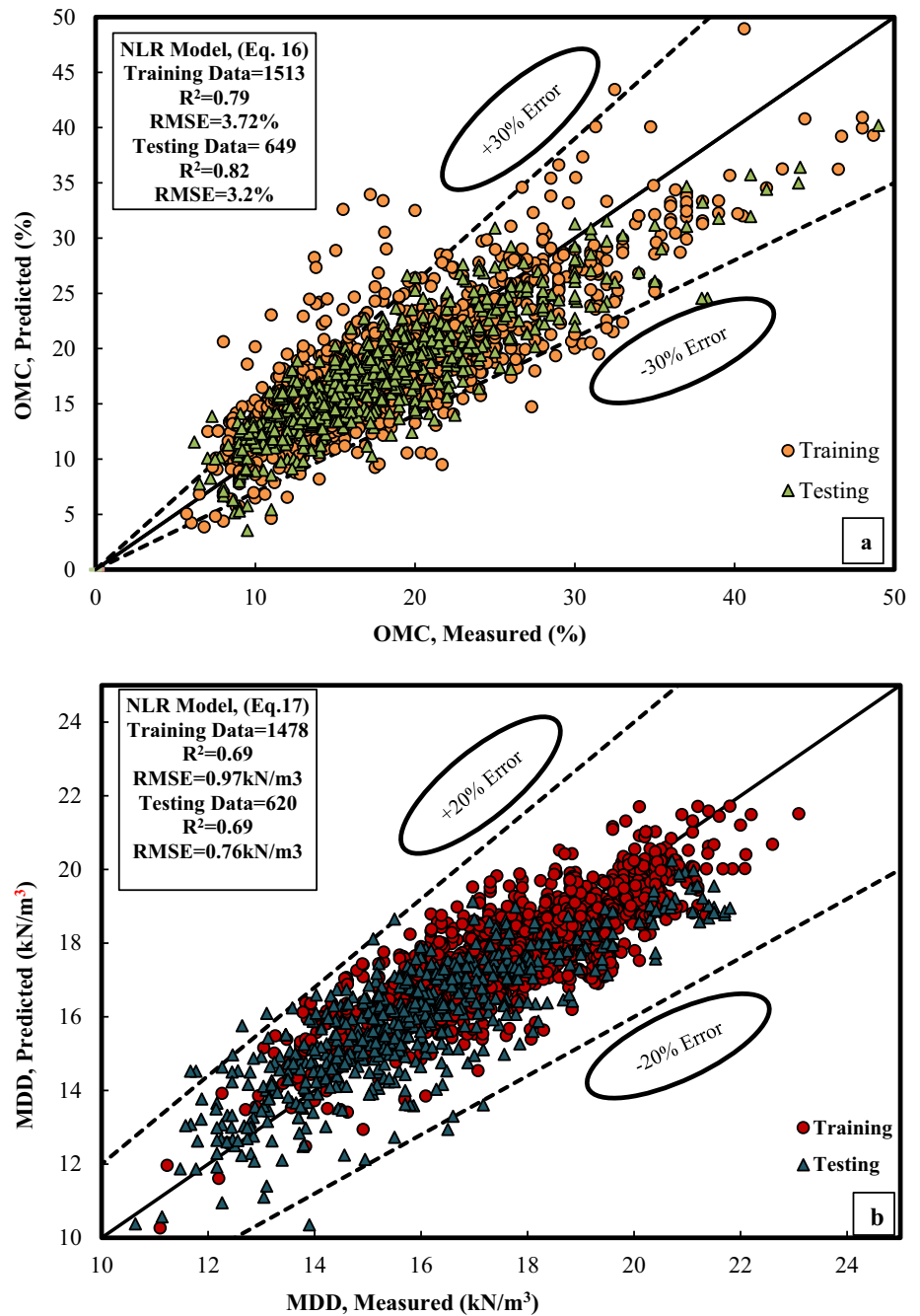
5.3 Nonlinear regression model

Generally, to develop a nonlinear model, the following Eq. (3) can be applied [39]. To forecast the OMC and MDD, the connection among various variables in Eqs. (1) and (2) can be expressed in Eq. (3):

$$OMC, MDD = a + b(G)^c + d(S)^e + f(F)^g + h(LL)^i + j(PL)^k + l(PI)^m \quad (3)$$

where G stands for gravel content (%), S stands for sand content (%), F stands for fine content (%), LL stands for liquid limit, PL stands for plastic limit, and PI stands for plasticity index. Moreover, the model parameters are a, b, c, d, e, f, g, h, I, j, k, l, and m, calculated based on the least square method.

Fig. 6 Comparison between measured and predicted values of **a** OMC and **b** MDD for NLR model: training data and testing data



5.4 Multiple linear regressions

The MLR, also a regression procedure, can be employed when the expected variable has a parameter greater than two stages. MLR is a statistical approach that is comparable to multiple linear regressions.

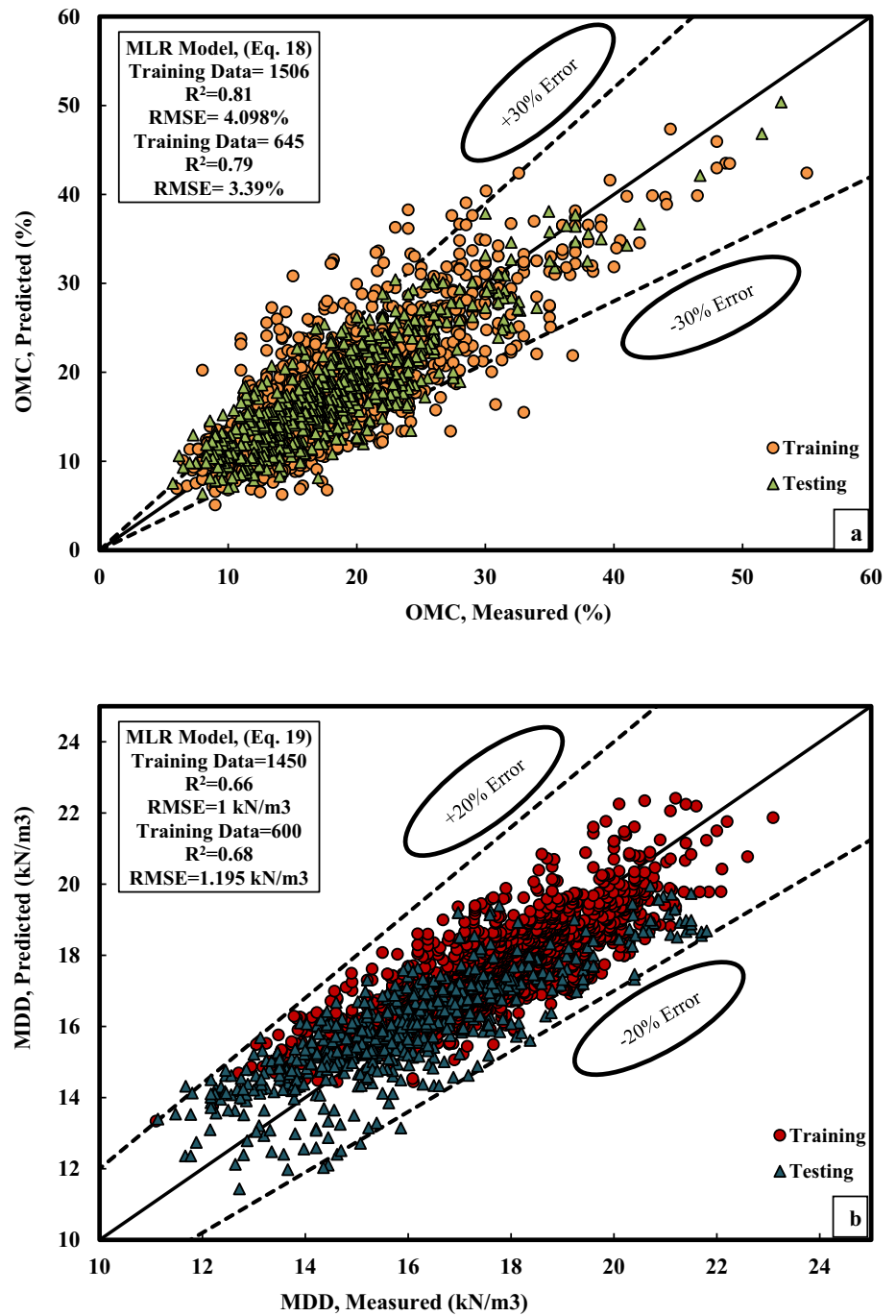
Equation (4) can be utilized to find the variance among predictable and independent variables.

$$OMC, MDD = a(G)^b(S)^c(F)^d(LL)^e(PL)^f(PI)^g \quad (4)$$

If soil is purely coarse-grained ($F\% = 0$) or fine-grained ($G\% + S\% = 0$), Eq. (4) cannot be applied to forecast the compaction characteristics. Therefore, there is a limitation in applying this model as it cannot represent the real formation of the soil. The least-square method was implemented to find the parameters (a, b, c, d, e, f, and g) and model variables.

The datasets are divided into two groups for the above models: Training and Testing. 70% of the datasets are used for training, while 30% go to the testing phase.

Fig. 7 Comparison between measured and predicted values of **a** OMC and **b** MDD for MLR model: training and testing data

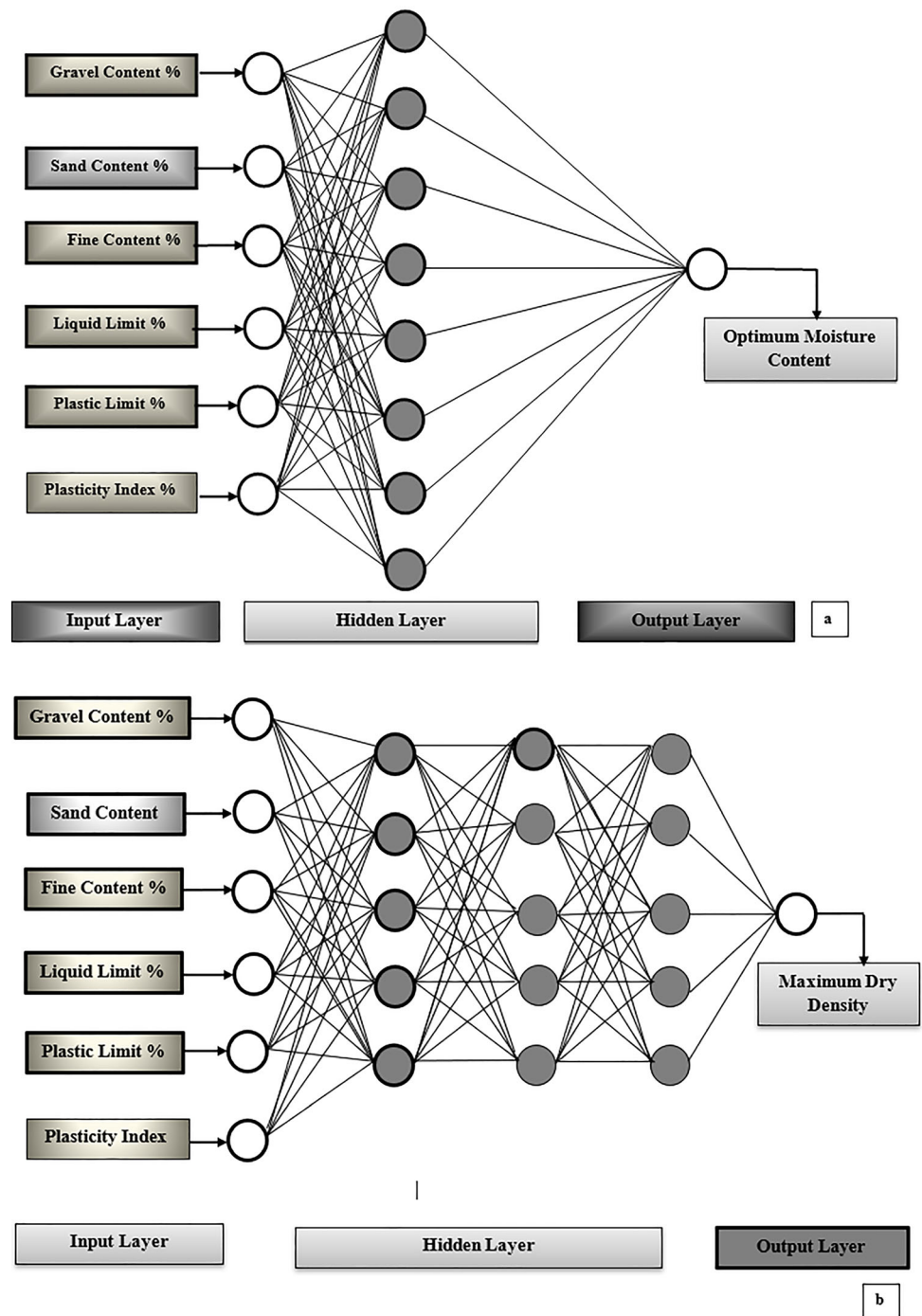


5.5 ANN model

ANN is a robust simulation software developed for data analysis and computing to process and analyze information similarly to a human brain. Construction engineering frequently uses this machine learning method to forecast the future behavior of numerous numerical issues [110, 111]. The ANN model has three primary layers: input, hidden, and output. Depending on the intended problem, each input and output layer may consist of one or more layers. The

hidden layer is often extended for two or more layers. The objective of the designed model and the data collected often decide the input and output layers, while the rating weight, transfer function, and the bias of each layer toward other layers determine the hidden layer. Based on a combination of proportions, weight/bias, and various parameters, including (G, S, F, LL, PL, and PI) as inputs, a multi-layer feed-forward network is constructed, and output ANN here is either the OMC or MDD. For designing the network architecture, no standard approach is available.

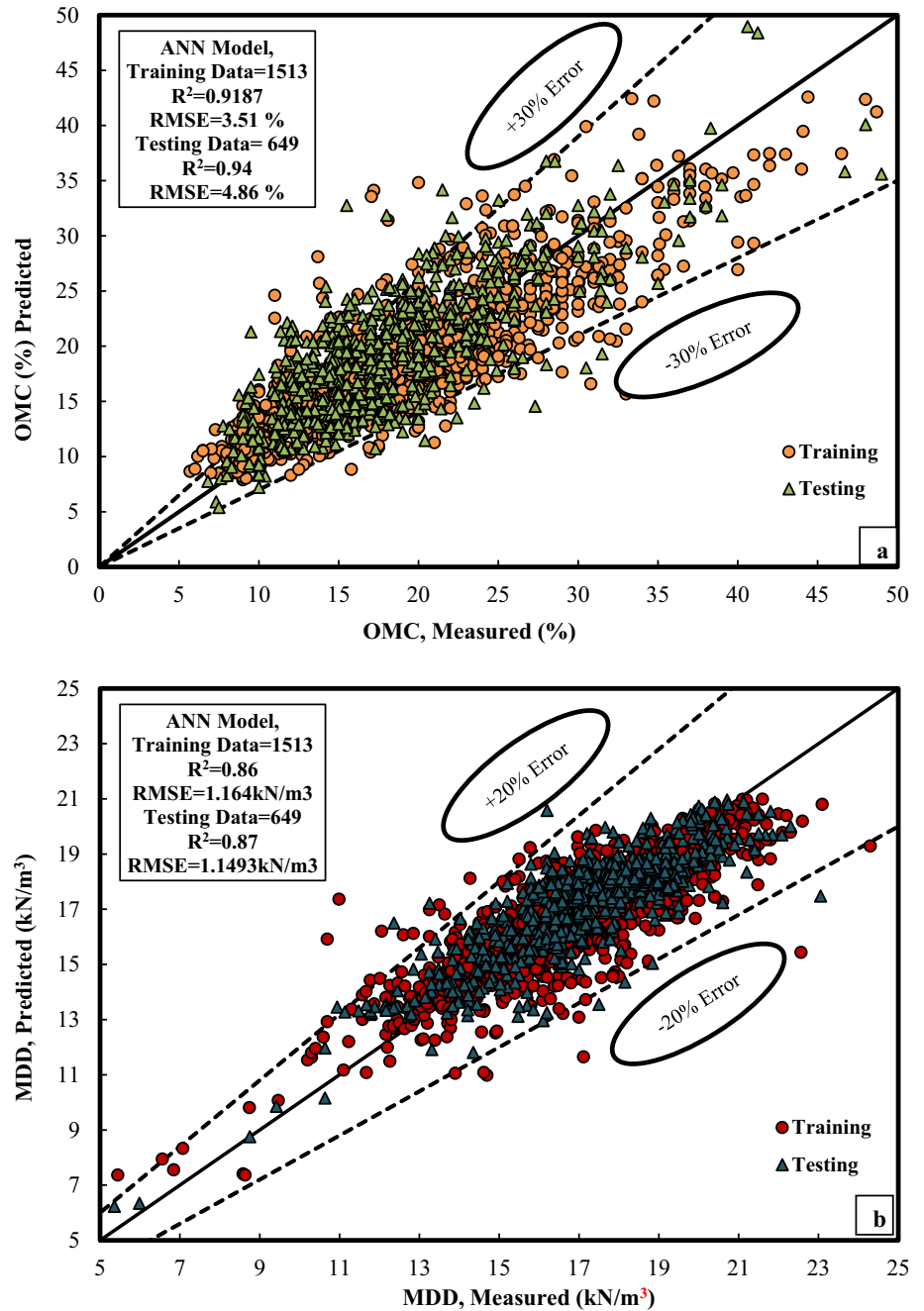
Fig. 8 The architecture of the used ANN models **a** OMC and **b** MDD



The number of hidden layers and neurons is determined based on the trial and error test. The prime goal of the training process of the network is to achieve the optimized number of iterations (epochs) from which the minimum mean absolute error (MAE), root means square error (RMSE), and the high R^2 -value are provided. Several studies have examined the influence of iteration on reducing the MAE and RMSE. In order to prepare for the designed ANN, the obtained dataset (a total of 2162 data

has been separated into two groups. Approximately 70% of datasets were utilized as trained data for training the network. For the testing data, 30% of the overall data was used for the trained network [34]. The constructed ANN was trained and tested for several hidden layers to select the best network structure based on the compatibility of the predicted compaction characteristics with the obtained data, for predicting OMC, the best-trained network that offers the highest R^2 and the lowest MAE and RMSE as the

Fig. 9 Comparison between measured and predicted values of **a** OMC and **b** MDD for ANN model: training data and testing data

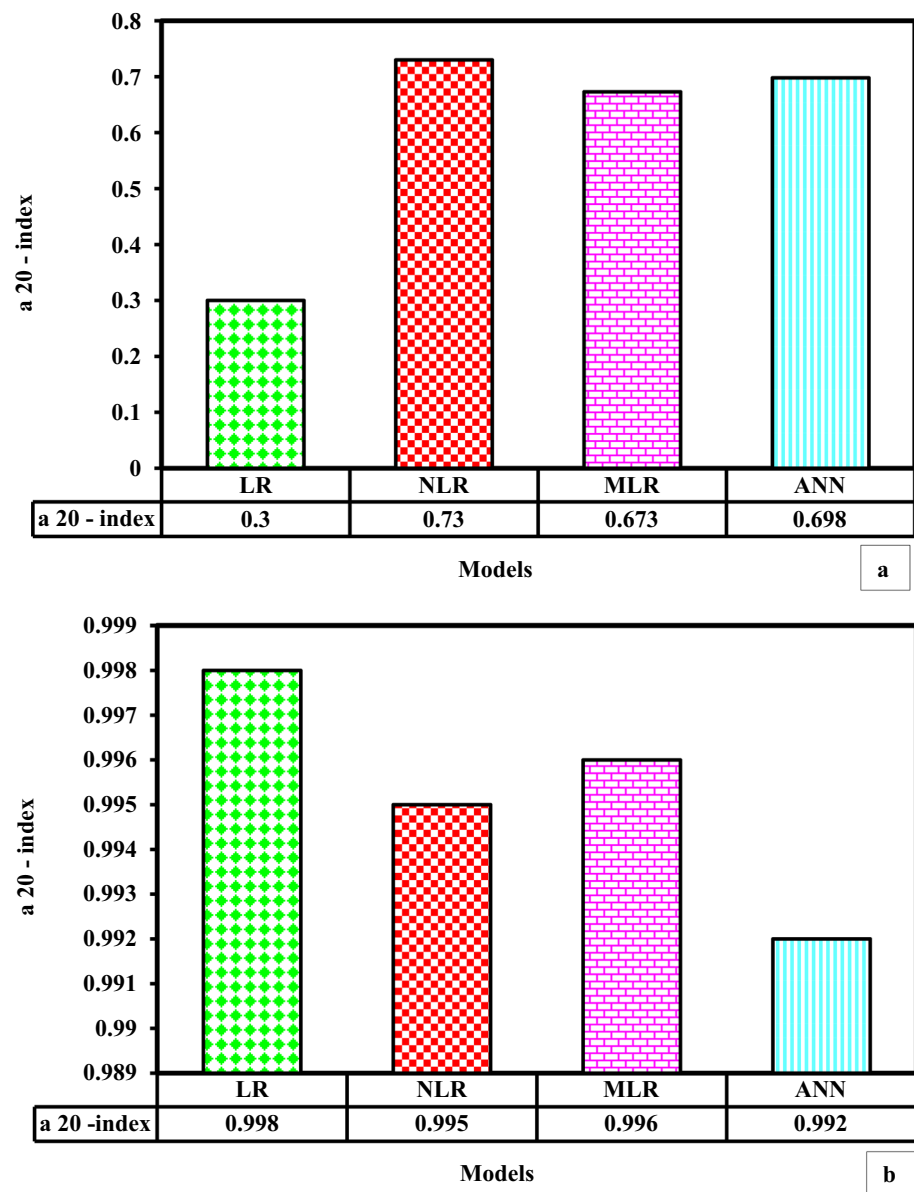


ANN structure with one hidden layer, eight neurons, and a hyperbolic tangent transfer function (as illustrated in Fig. 13, Table 3). As far as the prediction of MDD is concerned, the best-trained network was three hidden layers, five neurons, and a hyperbolic tangent transfer function.

Analyzing the complexity of an Artificial Neural Network (ANN) involves considering several factors related to its architecture and operation. Below the following is the breakdown of the key aspects to consider when conducting a complexity analysis of an ANN:

- i. Number of layers: the number of layers in the network, including input, hidden, and output layers, significantly impacts complexity. Deeper networks are generally more complex.
- ii. Layer size: the number of neurons or units in each layer, especially in the hidden layers, affects model complexity.
- iii. Connectivity: the extent of connections between neurons, also known as the network’s topology, plays a role in complexity.

Fig. 10 The a20-index values of **a** OMC and **b** MDD using all data



- iv. Number of parameters: the number of weights and biases in the network. The total parameters in an ANN increase with the number of neurons and layers.
- v. Input features: the dimensionality of the input data contributes to the complexity. Higher-dimensional input data require larger networks.
- vi. Activation functions: each layer's choice of activation functions can influence complexity. Non-linear activation functions can add to the complexity of the model.
- vii. Training algorithm: the algorithm used for training the ANN can impact complexity. Some training algorithms, like backpropagation, require more iterations to converge.
- viii. Hyperparameters: the settings of hyperparameters, such as learning rate, batch size, and dropout rate, can impact training time and model complexity.
- ix. Regularization techniques: techniques like L1 and L2 regularization can add complexity to the model by penalizing large weights.
- x. Pruning: pruning techniques can reduce complexity by eliminating unnecessary connections and neurons after training.
- xi. Parallelization: distributed training across multiple GPUs or CPUs can change the complexity of the training process.
- xii. Complexity metrics: several complexity metrics, like FLOPs (floating-point operations), are used to

quantitatively measure the computational complexity of ANN models.

- xiii. Overfitting: the potential for overfitting increases with the complexity of the model. Complexity can lead to the model fitting noise in the data rather than the underlying patterns.

Equations (5)–(7) display the ANN model's General Equation.

From linear node 0:

$$\text{OMC, MDD} = \text{Threshold} + \left(\frac{\text{Node 1}}{1 + e^{-B1}} \right) + \left(\frac{\text{Node 2}}{1 + e^{-B2}} \right) + \dots \quad (5)$$

From sigmoid node 1:

$$B1 = \text{Threshold} + \sum (\text{Attribute} * \text{Variables}) \quad (6)$$

From sigmoid node 2:

$$B2 = \text{Threshold} + \sum (\text{Attribute} * \text{Variables}) \quad (7)$$

5.6 Pseudocode/algorithm

```
# Function to train models and predict OMC and MDD
Function train_and_predict(input_data):
  # Split input_data into features (X) and target variables (y) for OMC and MDD
  X_omc, y_omc = extract_features_and_target_for_OMC(input_data)
  X_mdd, y_mdd = extract_features_and_target_for_MDD(input_data)

  # Split data into training and testing sets (e.g., 80-20 split)
  X_train_omc, X_test_omc, y_train_omc, y_test_omc = split_data(X_omc, y_omc)
  X_train_mdd, X_test_mdd, y_train_mdd, y_test_mdd = split_data(X_mdd, y_mdd)

  # Train models for OMC prediction
  lr_omc = train_linear_regression(X_train_omc, y_train_omc)
  nlr_omc = train_nonlinear_regression(X_train_omc, y_train_omc)
  mlr_omc = train_multilinear_regression(X_train_omc, y_train_omc)
  ann_omc = train_neural_network(X_train_omc, y_train_omc)

  # Train models for MDD prediction
  lr_mdd = train_linear_regression(X_train_mdd, y_train_mdd)
  nlr_mdd = train_nonlinear_regression(X_train_mdd, y_train_mdd)
  mlr_mdd = train_multilinear_regression(X_train_mdd, y_train_mdd)
  ann_mdd = train_neural_network(X_train_mdd, y_train_mdd)

  # Make predictions using trained models
  pred_omc_lr = lr_omc.predict(X_test_omc)
  pred_omc_nlr = nlr_omc.predict(X_test_omc)
  pred_omc_mlr = mlr_omc.predict(X_test_omc)
  pred_omc_ann = ann_omc.predict(X_test_omc)

  pred_mdd_lr = lr_mdd.predict(X_test_mdd)
  pred_mdd_nlr = nlr_mdd.predict(X_test_mdd)
  pred_mdd_mlr = mlr_mdd.predict(X_test_mdd)
  pred_mdd_ann = ann_mdd.predict(X_test_mdd)

  # Evaluate model performance using metrics
  evaluation_omc_lr = evaluate_performance(y_test_omc, pred_omc_lr)
  evaluation_omc_nlr = evaluate_performance(y_test_omc, pred_omc_nlr)
  evaluation_omc_mlr = evaluate_performance(y_test_omc, pred_omc_mlr)
  evaluation_omc_ann = evaluate_performance(y_test_omc, pred_omc_ann)

  evaluation_mdd_lr = evaluate_performance(y_test_mdd, pred_mdd_lr)
  evaluation_mdd_nlr = evaluate_performance(y_test_mdd, pred_mdd_nlr)
  evaluation_mdd_mlr = evaluate_performance(y_test_mdd, pred_mdd_mlr)
  evaluation_mdd_ann = evaluate_performance(y_test_mdd, pred_mdd_ann)

  # Output model evaluations and predictions
  output_results(evaluation_omc_lr, evaluation_omc_nlr, evaluation_omc_mlr, evaluation_omc_ann,
    evaluation_mdd_lr, evaluation_mdd_nlr, evaluation_mdd_mlr, evaluation_mdd_ann)
```

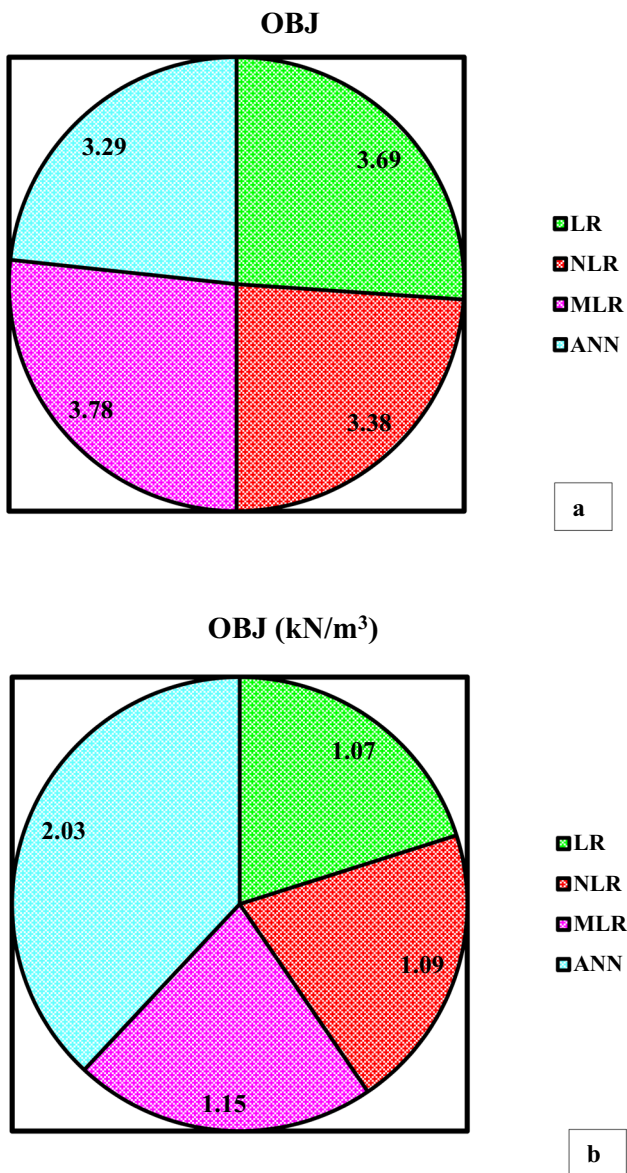



Fig. 11 The OBJ values of **a** OMC, **b** MDD for all the models

6 Model assessment tools

The suggested models were evaluated using a variety of metrics, including coefficient of determination (R^2), a20-index, scatter index (SI), Objective (OBJ), root mean squared error (RMSE), and mean absolute error (MAE), which may be calculated using the formulas below:

$$R^2 = \left(\frac{\sum_{p=1}^p (y_i - y_i')(y_p - y_p')}{\sqrt{[\sum_{p=1}^p (y_i - y_i')^2][\sum_{p=1}^p (y_p - y_p')^2]}} \right)^2 \tag{8}$$

$$RMSE = \sqrt{\frac{\sum_{p=1}^p (y_p - y_i)^2}{p}} \tag{9}$$

$$MAE = \frac{\sum_{p=1}^p |y_p - y_i|}{p} \tag{10}$$

$$a20 - index = \frac{m20}{N} \tag{11}$$

$$SI = \frac{RMSE}{y'} \tag{12}$$

$$OBJ = \left(\frac{n_{tr}}{n_{all}} * \frac{RMSE_{tr} + MAE_{tr}}{R_{tr}^2 + 1} \right) + \left(\frac{n_{tst}}{n_{all}} * \frac{RMSE_{tst} + MAE_{tst}}{R_{tst}^2 + 1} \right) + \left(\frac{n_{val}}{n_{all}} * \frac{RMSE_{val} + MAE_{val}}{R_{val}^2 + 1} \right) \tag{13}$$

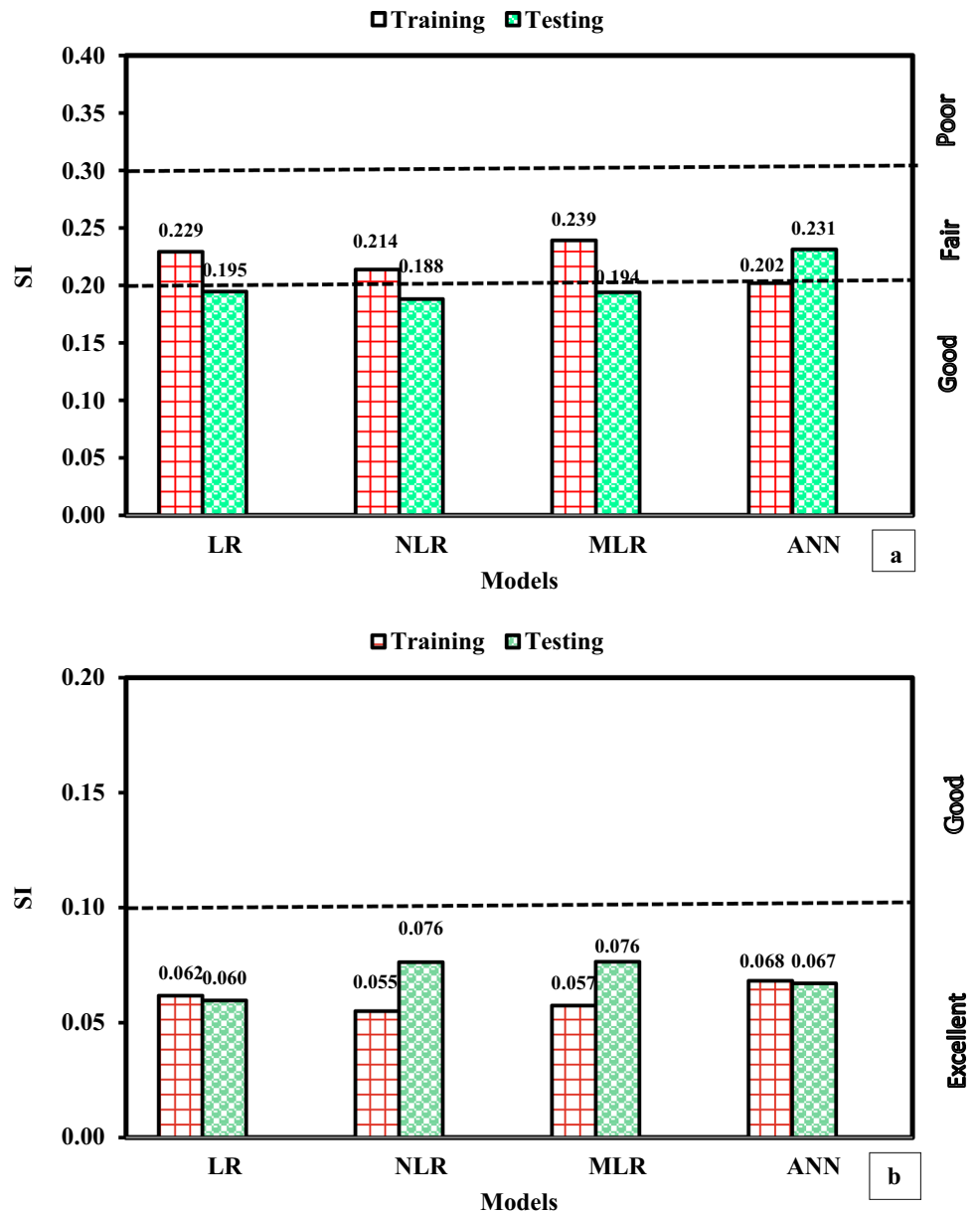
From the formulas above, y_p and y_i are the expected and actual values of the path pattern, and y_p' and y_i' are the averages of the actual and forecasted values. Training, testing, and validating datasets are tr , tst , and val , respectively; the number of patterns (collected data) in the associated dataset is denoted as n . In contrast to R^2 , which has an optimal value of one, the other evaluating factors have optimal values of zero. About the SI parameter, a model performs poorly when it is [0.3, fairly when it is between 0.2 and 0.3, well when it is between 0.1 and 0.2, and very well when it is < 0.1. To assess the effectiveness of the recommended models, the OBJ parameter was also used as an integrated performance parameter in Eq. (12). $m20$ is the ratio of experimental to the predicted value, which varies between 0.8 and 1.2, and H is the total number of data samples. The prime benefit of the a20-index is that the model predicts values with a deviation of $\pm 20\%$ compared to actual values. In order to graphically show how each model overestimates and underestimates the expected results of OMC and MDD compared to the actual values from the experiments, positive and negative error margin lines were added to the model findings. A positive value is represented for the overestimated percentage of OMC and MDD, while a negative value means the underestimated percentage of OMC and MDD.

7 Results and analysis

7.1 The LR model

Figure 5 exhibits the relation between actual and estimated OMC and MDD for all training and testing datasets. In order to develop the model, 2162 data sets have been utilized. Concerning the developed equations, the two plasticity parameters of LL and PI play a substantial role in

Fig. 12 The OBJ values of **a** OMC and **b** MDD for all developed models



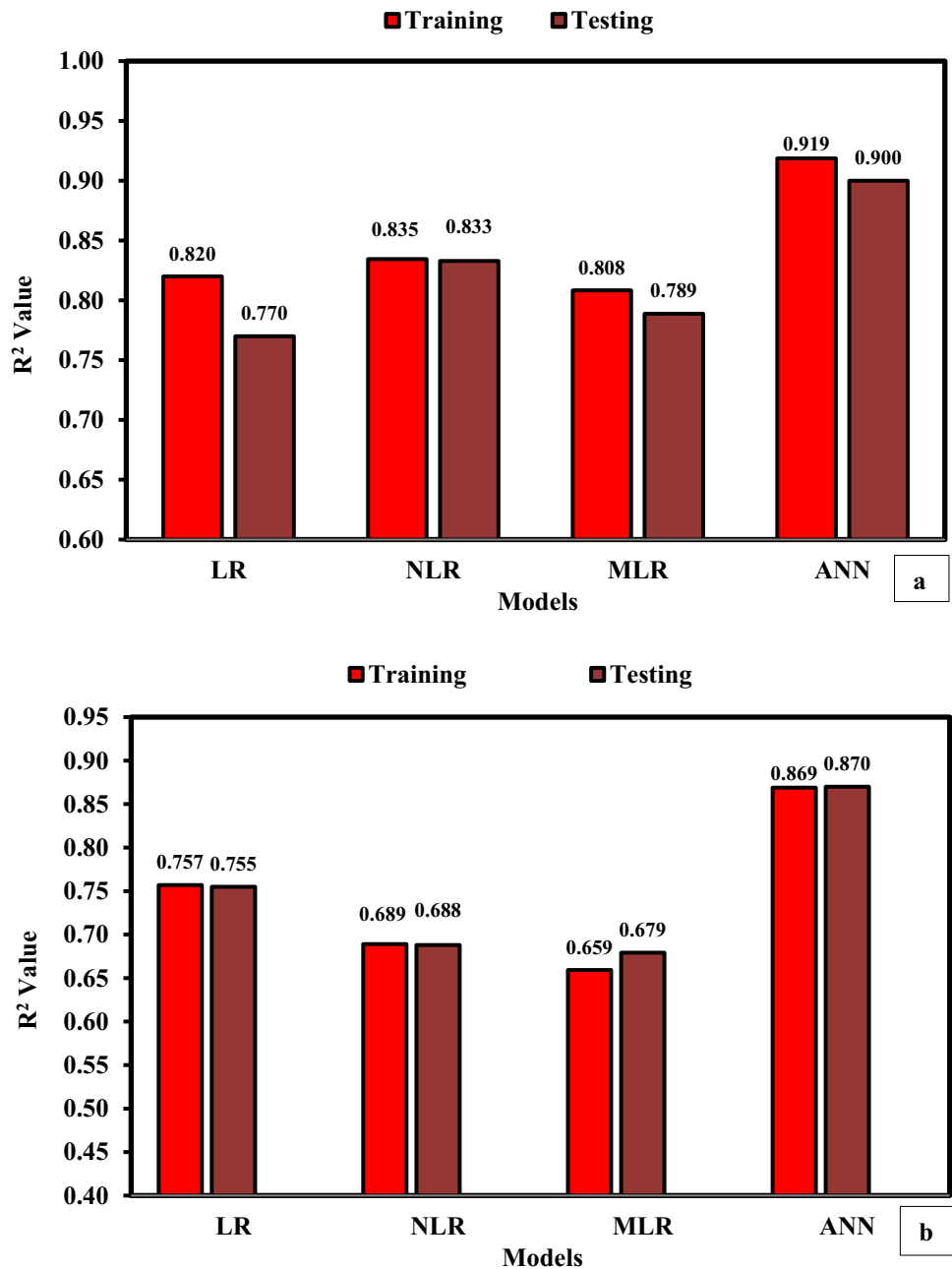
determining OMC and MDD. By optimizing the sum of error squares and the least square approach, which was performed in Excel using Solver to obtain the ideal value for the Equation, the present model’s weight of each parameter on the OMC and MDD was found. The following are the Equations for the LR model with various weight parameters [Eqs. (13) and (14)]:

$$\begin{aligned}
 \text{OMC} = & 0.96 - 0.05G - 0.04S + 0.03F + 0.72LL \\
 & + 0.07PL - 0.74PI \tag{14}
 \end{aligned}$$

$$\begin{aligned}
 \text{MDD} = & 27.13 - 0.025G - 0.045S - 0.073F - 0.215LL \\
 & + 0.04PL + 0.216PI \tag{15}
 \end{aligned}$$

As seen in the above Equations, among other input parameters, PI can have more influence on lowering the value of OMC; on the contrary, MDD value declines with the increase of LL. This seems to be consistent with the experimental findings reported in the literature that soil with higher LL has a lower MDD value. The obtained R^2 , RMSE, and MAE assessment parameters for the OMC are 0.82, 3.99, and 2.98%, respectively. For the MDD, R^2 , RMSE, and MAE are 0.76, 1.05, and 0.84 kN/m^3 , respectively. Furthermore, as shown in Figs. 11 and 12, the current model’s OBJ and SI values for OMC for the training dataset are 3.69 and 0.23, respectively, while these values are 1.07 and 0.062, correspondingly, for MDD.

Fig. 13 The R² values of **a** OMC and **b** MDD for all four models



7.2 NLR model

The NLR model is another model that might be interesting to examine for its performance among other models. Figure 6a, b depict the compaction characteristics values for predicted and real data obtained from the literature for training and testing datasets. According to the developed model, F plays a considerable role in changing the value of both OMC and MDD. The following are the suggested formulas for the NLR model with various variable parameters [Eqs. (15) and (16)]:

$$\begin{aligned}
 \text{OMC} = & 777.3 + 2.13(G)^{-0.213} + 0.000028(S)^{-6.31} - 784.6(F)^{-0.003} \\
 & - 0.137(LL)^{0.93} + 0.144(PL)^{1.38} + 2(PI)^{0.47}
 \end{aligned}
 \tag{16}$$

$$\begin{aligned}
 \text{MDD} = & 35.7 - 4(G)^{-0.046} - 94.4(S)^{-0.0013} + 60.8(F)^{-0.012} \\
 & - 0.63(LL)^{0.83} + 23.68(PL)^{0.077} + 0.35(PI)^{0.92}
 \end{aligned}
 \tag{17}$$

It should be noted that this equation cannot be used for pure fine-grained soil since it provides zero OMC and MDD. In order to overcome this issue, the value of G and S can be approximated to zero (not zero).

Table 3 The tested ANN architecture for OMC

	No. of hidden layers			R ²	MAE (%)	RMSE (%)	
	No. of hidden layers	No. of neurons in hidden layers					
		The left side	The middle	The right side			
1	5	0	0	0.9141	2.611	3.6104	
1	6	0	0	0.9155	2.6312	3.5765	
1	7	0	0	0.9115	2.6715	3.6542	
1	8	0	0	0.9187	2.5746	3.5139	
1	9	0	0	0.917	2.5828	3.5415	
1	10	0	0	0.9151	2.6054	3.5811	
2	6	0	6	0.9081	2.7394	3.7289	
2	7	0	7	0.9052	2.8017	3.8289	
2	8	0	8	0.9039	2.8266	3.8407	
2	10	0	10	0.9028	2.8214	3.8718	
3	7	7	7	0.9047	2.6454	3.7887	
3	8	8	8	0.9002	2.6975	3.8701	
3	10	10	10	0.8979	2.7101	3.911	

For OMC, this model’s R², RMSE, and MAE assessment parameters are 0.79, 4.22%, and 2.72%, respectively, whereas these values for MDD are 0.69, 1.062, and 0.84 kN/m³ correspondingly. Furthermore, the models’ OBJ for OMC and MDD is 3.38% and 1.09 kN/m³, respectively. Moreover, the SI values for OMC and MDD are 0.214 and 0.055, respectively.

7.3 MLR model

The comparison between the real compaction characteristics values from literature and predicted values are illustrated in Fig. 7a, b for both phases (training and testing datasets). The following equation shows that PL is the most influential parameter affecting the OMC. Changing the PL value can substantially change the value of OMC. However, both G and S contents have more influence on MDD. Equations (18) and (19) can be used to forecast the OMC and MDD, respectively, for the MLR model with various variable parameters:

$$\begin{aligned}
 \text{OMC} &= 0.632*(G)^{-0.0314}(S)^{-0.0142}(F)^{0.08}(LL)^{0.04}(PL)^{0.955}(PI)^{-0.0153} \\
 &\hspace{15em} (18)
 \end{aligned}$$

$$\begin{aligned}
 \text{MDD} &= 39.74*(G)^{0.01}(S)^{0.01}(F)^{-0.03}(LL)^{-0.12}(PL)^{-0.12}(PI)^{0.02} \\
 &\hspace{15em} (19)
 \end{aligned}$$

The R², RMSE, and MAE assessment parameters for this model are 0.81, 4.098%, and 3.02%, respectively, for OMC, while, for MDD, these are 0.66, 1 kN/m³, and 0.8, respectively. Furthermore, the current model’s OBJ for OMC and MDD are 3.78 and 1.15, respectively. In addition, the SI values of the training dataset for both OMC and MDD are 0.239 and 0.054, respectively.

7.4 ANN model

The authors investigated several hidden layers, neurons, momentum, learning rate, and iterations to achieve high ANN efficiency. Tables 3 and 4 indicate several ANN architectures examined to obtain the most optimum ANN model for OMC and MDD. Consequently, it was detected that when the ANN contains one hidden layer, eight neurons on the left side (as shown in Fig. 8a), 0.1 momentum, 0.2 learning rate, and 2000 iterations, the OMC is best predicted. However, MDD was best predicted with three hidden layers and five neurons on each side (Fig. 8b). The comparison of both predicted and real values of the compaction characteristics are demonstrated in Fig. 9a, b for both phases (training and testing datasets). The ANN model performs better in predicting OMC value than other models. Nevertheless, compared to the other models, the ANN model only performed better in providing R² value in predicting MDD. Regarding the OMC, the ANN model’s R², RMSE, and MAE assessment parameters are 0.92, 3.51, and 2.574%, respectively. However, for predicting MDD, R², RMSE, and MAE values are 0.86, 1.164 kN/m³, and 2.31 kN/m³, respectively. Moreover, the present model’s OBJ for OMC and MDD are 3.29 and 2.03, respectively. Further, SI values for the training dataset for both OMC and MDD are 0.202 and 0.068, respectively.

8 Model comparisons

In order to categorize the effectiveness of the constructed models, six different quantitative tools, MAE, R², RMSE, OBJ, a20-index, and SI, were used. Figures 13, 14, and 15 illustrate R², RMSE, and MAE, respectively.

Table 4 The tested ANN architecture for MDD

No. of hidden layers	No. of neurons in hidden layers			R ²	MAE (kN/m ³)	RMSE (kN/m ³)
	The left side	The middle	The right side			
1	5	0	0	0.8543	0.9175	1.2119
1	6	0	0	0.8499	0.9109	1.2194
1	7	0	0	0.8563	0.8991	1.198
1	8	0	0	0.8585	0.8806	1.183
1	9	0	0	0.8583	0.8827	1.185
3	5	5	5	0.8618	0.8689	1.1647
2	6	0	6	0.8574	0.8741	1.1769
3	6	6	6	0.858	0.8822	1.1809
3	8	8	8	0.86	0.87	1.1626
3	10	10	10	0.8572	0.892	1.1929
2	8	0	8	0.8604	0.88	1.1744

As far as the OMC is concerned, according to the figures, the ANN outperforms LR, NLR, and MLR models as it has a higher R² and lowers RMSE and MAE values. To assess the reliability of the models, a new engineering index, the a20—index, was used. The advantage of the proposed a20 – index is that it states the number of samples that match expected values with a variance of less than 20% from experimental values [11]. Concerning this index, Fig. 10a compares the predicted and measured OMC of the models. The figure indicated that NLR provided a higher a20- index value of 0.73, and it was 0.7 for the ANN model, which means that 73% of samples with the predicted values have deviated 20% from the actual values of OMC. Thus, the NLR model predicts OMC with the least percentage of error. The OBJ for all the models is shown in Fig. 11. The OBJ values of LR, NLR, MLR, and ANN models are 3.69, 3.38, 3.78, and 3.29, respectively. Even though the OBJ values for all the models are very close, the ANN model has the least OBJ value compared to the others. Concerning all the assessment tools, the ANN approach provides more reliable outcomes and accurately forecasts OMC.

The SI values are exhibited in Fig. 12 for the training and testing phases. Figure 12a shows that the testing phases for LR, NLR, and MLR are between 0.2 and 0.3, indicating fair performance. Nevertheless, the SI value for the ANN model is on a merge of good performance. This contradicts the other phases, as LR, NLR, and MLR perform well. In light of these analyses, although all models can, to some extent, be used to forecast the OMC from physical soil properties, the ANN models are proven to have a distinct performance.

Concerning the prediction of MDD, in Fig. 10b, all the models provide a distinction prediction of MDD, and more

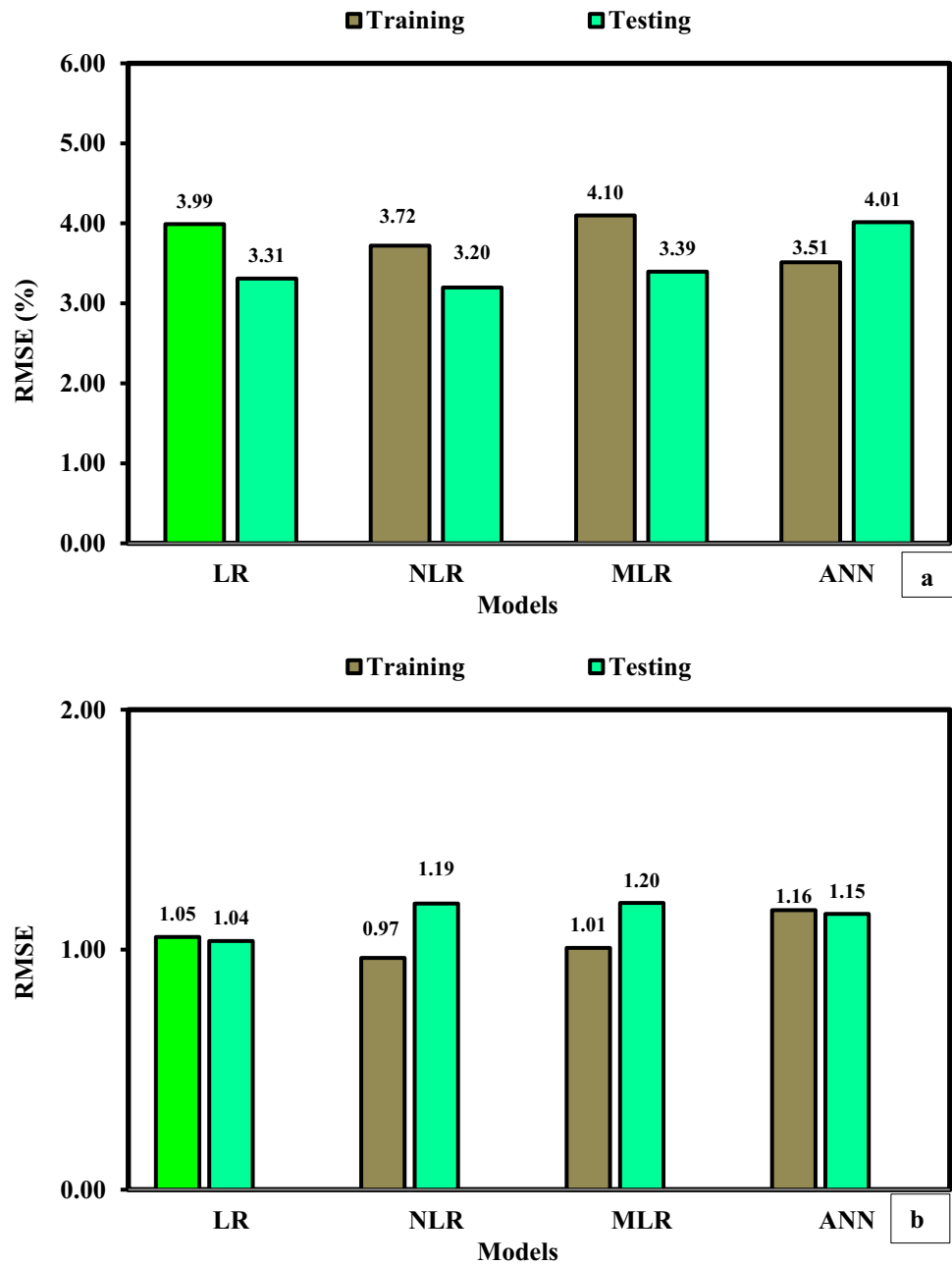
than 99% of the predicted samples were within $\pm 20\%$ deviation from the actual data. ANN model provided a slightly higher a20-index. Regarding Fig. 11b, the OBJ values of LR, NLR, MLR, and ANN are 1.07, 1.09, 1.15, and 2.03, respectively. LR model has the least OBJ compared to the other models. In this regard, the LR, MLR, and ANN somewhat perform well in predicting MDD. According to Fig. 12b, the SI values for LR, NLR, MLR, and ANN models are 0.062, 0.055, 0.057, and 0.068, respectively. These values show that all the models are excellent at predicting MDD.

Regarding all the statistical assessment criteria, all the statistical assessment criteria, all the models perform well in predicting MDD, and the percentages of errors are less than 20%. However, the ANN model provides a higher R² value, and LR can still be a reliable model to predict MDD with an R² value of 0.76.

9 Sensitivity analysis

In order to obtain the evident influence of each parameter on the OMC and MDD, sensitivity analyses were performed for the models. The MLR model was selected since the value of all input parameters is greater than zero; this can provide the real contribution of each parameter in the model and the influence of each parameter on OMC and MDD. All training data were combined throughout these analyses, and a single input variable was excluded each time. The RMSE, R², and MAE were individually determined so that the influence of each parameter could be observed. Tables 5 and 6 indicate the variation of evaluation criteria for predicting OMC and MDD, respectively.

Fig. 14 The RMSE values of **a** OMC and **b** MDD for all four models



The most efficient input parameter is the one that can remarkably change the values of statistical tools (increasing RMSE, MAE, and decreasing R^2). Table 5 indicates the outcomes of the sensitivity study on predicting OMC. The findings show that PL is the most dominant parameter influencing the value of OMC. The R^2 value of the model changed from 0.808 to 0.615 when excluding PL. On the contrary, the value of RMSE has increased from 4.098 to 5.8.06. Table 6 clarifies the influence of a single input parameter on MDD. When removing G, the R^2 value declined from 0.66 to 0.62; however, the values of RMSE and MAE increased to 1.052 and 0.83, respectively.

Therefore, PL can be considered a dominant input parameter that significantly influences the value of OMC, while, for quantifying MDD, G plays a remarkable role.

10 Recommendations for future work

- i. Refinement of model integration: explore methods to integrate different models (ANN, NLR, LR, MLR) to harness the strengths of each model in a combined predictive approach. Ensemble techniques or hybrid models could potentially yield

Fig. 15 The MAE value of **a** OMC and **b** MDD for all the four models

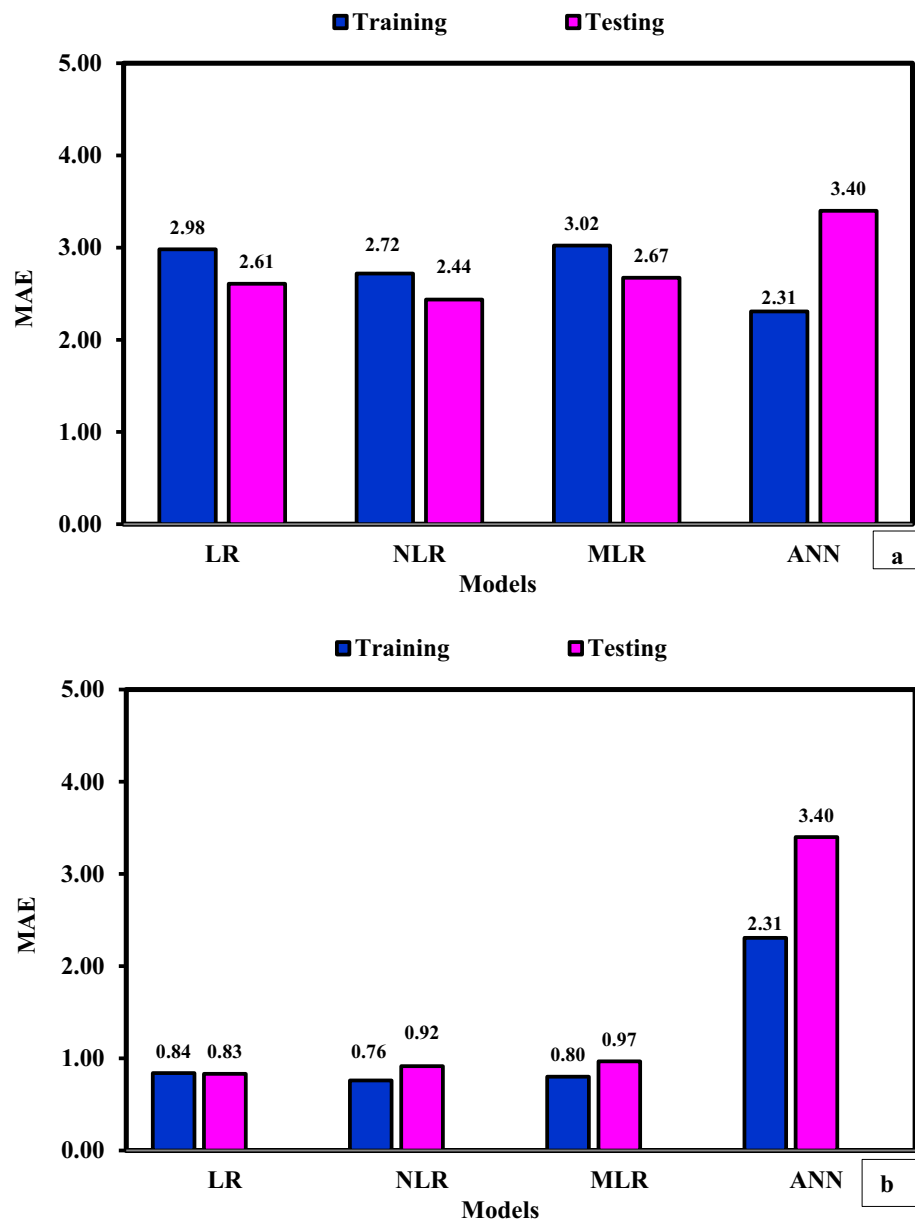


Table 5 Sensitivity analysis for MLR model on OMC

Nos.	Input parameters	Removed parameter	R ²	MAE	RMSE	Ranking
1	G, S, F, LL, PL, PI		0.808	3.02	4.098	-
2	S, F, LL, PL, PI	G	0.798	3.12	4.202	2
3	G, F, LL, PL, PI	S	0.807	3.02	4.106	4
4	G, S, LL, PL, PI	F	0.803	3.085	4.147	3
5	G, S, F, PL, PI	LL	0.808	3.02	4.099	5
6	G, S, F, LL, PI	PL	0.615	3.41	5.806	1
7	G, S, F, LL, PL	PI	0.808	3.02	4.098	6

more accurate and robust predictions by leveraging the unique capabilities of individual models.

ii. Additional input parameters: consider expanding the input parameters beyond the six currently utilized factors. Incorporating other relevant soil

Table 6 Sensitivity analysis for MLR model on MDD

Nos.	Input parameter	Removed parameter	R ²	MAE	RMSE	Ranking
1	G, S, F, LL, PL, PI	–	0.66	0.8	1	–
2	S, F, LL, PL, PI	G	0.62	0.8314	1.052	1
3	G, F, LL, PL, PI	S	0.6538	0.802	1.0155	4
4	G, S, LL, PL, PI	F	0.6357	0.8277	1.041	2
5	G, S, F, PL, PI	LL	0.6561	0.805	1.012	5
6	G, S, F, LL, PI	PL	0.6487	0.803	1.023	3
7	G, S, F, LL, PL	PI	0.6593	0.8005	1.007	6

properties or environmental factors that might influence compaction characteristics could enhance the models' predictive capabilities.

- iii. Data augmentation and validation: continuously collect new data and augment the existing dataset with additional samples to improve the models' training and validation. This can help ensure the models remain accurate and effective when applied to a wider range of soil variations.
- iv. Regional variability analysis: investigate how the predictive models perform across regions with varying soil compositions and climates. Understanding how these models generalize to geographical areas can enhance their applicability and reliability in diverse construction contexts.
- v. Dynamic modelling for real-time applications: develop dynamic models that can continuously adapt and predict compaction characteristics based on real-time data from sensors embedded in construction equipment. This could facilitate on-site decision-making and optimize the compaction process as it progresses.
- vi. Incorporation of advanced ai techniques: explore the integration of advanced artificial intelligence techniques, such as deep learning or reinforcement learning, to further improve the predictive accuracy of the models. These techniques might capture more complex nonlinear relationships within the dataset.
- vii. Field validation and case studies: conduct field validation studies where predictions from the models are compared against actual on-site compaction results. Case studies in real construction scenarios can provide practical insights into the models' effectiveness and help refine their application in practical settings.
- viii. Optimization algorithms: develop optimization algorithms that utilize the predictive models to suggest the most efficient and effective combinations of soil properties to achieve the desired compaction characteristics, considering both resource utilization and construction timelines.

11 Conclusions

Developing a predictive model to anticipate Optimum Moisture Content (OMC) and Maximum Dry Density (MDD) is crucial for leveraging fundamental soil characteristics such as soil gradation and plasticity. These parameters are pivotal in gauging soil suitability, simplifying the selection process from various soil sources, and circumventing the requirement for standard Proctor tests. An analysis of 2162 datasets, encompassing Gravel (G), Sand (S), Fine content (F), Plastic Limit (PL), Liquid Limit (LL), and Plasticity Index (PI) parameters, unveiled significant revelations:

Several models—Linear Regression (LR), Nonlinear Regression (NLR), Multilinear Regression (MLR), and Artificial Neural Networks (ANN)—were devised to predict OMC and MDD. Evaluation using diverse metrics highlighted the superior performance of the ANN model. When it came to predicting OMC, it had higher R-squared (R²) values and lower Objective (OBJ), Root Mean Squared Error (RMSE), Scatter Index (SI), and Mean Absolute Error (MAE) values.

All models demonstrated robust predictive accuracy for OMC and MDD, with SI values below 0.2.

The ANN model particularly excelled in predicting OMC, displaying an OBJ value approximately 12% lower than LR, 2.7% lower than NLR, and 15% lower than MLR. Conversely, for MDD prediction, the ANN model presented a higher OBJ value compared to LR, NLR, and MLR.

Although the ANN model showcased the highest R² value, LR, NLR, and MLR models also exhibited commendable R² values, hovering around 0.8, signifying reliability in predicting OMC based on parameters G, S, F, PL, LL, and PI. The LR model's R² value (0.78) for MDD implies relative reliability in this context.

The model equations underscore the importance of Plastic Limit (PL) and Liquid Limit (LL) in projecting OMC, while the proportions of Gravel (G) and Sand (S) emerge as pivotal factors in determining MDD.

Sensitivity analyses identified Plastic Limit (PL) as the primary influencer in predicting OMC, while Gravel (G) holds substantial sway in quantifying MDD.

While the ANN model outshines others in OMC and MDD prediction, LR models might offer more practicality owing to their transparent equations grounded in soil physical properties. This contrasts with the ANN's inherently complex “black box” approach.

Funding None.

Availability of data and materials The data supporting the conclusions of this article are included.

Declarations

Conflict of interest None.

Ethical approval and consent to participate Not applicable.

References

- Abasi N (2013) A new empirical equation for compression behavior of unconsolidated clayey soils. *J Civ Eng Ferdowski Univ Mashhad (Persian Lang)* 24:41–56
- Abdalla A, Mohammed AS (2022) Hybrid MARS-, MEP-, and ANN-based prediction for modeling the compressive strength of cement mortar with various sand size and clay mineral meta-kaolin content. *Arch Civil Mech Eng* 22(4):1–16
- Adeoye AS, Alo BA, Abdu-Raheem YA (2018) Assessment of geotechnical properties of migmatite-derived residual lateritic soil from Ado-Ekiti South Western Nigeria as Barrier in Sanitary Landfill. *IJIRSET* 7:7454–7462
- Ahmed SS, Hossain N, Khan AJ, Islam MS (2016) Prediction of soaked CBR using index properties, dry density and unsoaked cbr of lean clay. *Malays J Civ Eng* 28:270–283
- Al-Khafaji AN (1993) Estimation of soil compaction parameters by means of Atterberg limits. *Q J Eng Geol Hydrogeol* 26:359–368. <https://doi.org/10.1144/GSL.QJEGH.1993.026.004.10>
- Albrecht BA, Benson CH (2001) Effect of desiccation on compacted natural clays. *J Geotech Geoenviron Eng* 127:67–75. [https://doi.org/10.1061/\(ASCE\)1090-0241\(2001\)127:1\(67\)](https://doi.org/10.1061/(ASCE)1090-0241(2001)127:1(67))
- Ali HF, Hama RAJ, Hama Kareem MI, Muhedin DA (2019) A correlation between compaction characteristics and soil index properties for fine-grained soils. *Polytech J* 9:93–99. <https://doi.org/10.25156/ptj.v9n2y2019.pp93-99>
- Ali TS, Fakhraldin MK (2016) Soil parameters analysis of Al-Najaf City in Iraq: case study. *J Geotech Eng* 3:56–62
- Alzabeebee S, Mohamad SA, Al-Hamd RKS (2021) Surrogate models to predict maximum dry unit weight, optimum moisture content and California bearing ratio from grain size distribution curve. *Road Mater Pavem Des.* <https://doi.org/10.1080/14680629.2021.1995471>
- Arshid MU, Kamal MA (2020) Appraisal of bearing capacity and modulus of subgrade reaction of refilled soils. *Civ Eng J* 6:2120–2130. <https://doi.org/10.28991/cej-2020-03091606>
- Asteris PG, Skentou AD, Bardhan A, Samui P, Pilakoutas K (2021) Predicting concrete compressive strength using hybrid ensembling of surrogate machine learning models. *Cem Concr Res* 145:106449
- Bekele A (2017) Correlation of CBR with index properties of soils in Sululta Town. MSc thesis, Addis Ababa University
- Bellezza I, Fratolocchi E (2006) Effectiveness of cement on hydraulic conductivity of compacted soil–cement mixtures. *Proc Inst Civ Eng—Gr Improv* 10:77–90. <https://doi.org/10.1680/grim.2006.10.2.77>
- Benson CH, Trast JM (1995) Hydraulic conductivity of thirteen compacted clays. *Clays Clay Miner* 43:669–681. <https://doi.org/10.1346/CCMN.1995.0430603>
- Bera A, Ghosh A (2011) Regression model for prediction of optimum moisture content and maximum dry unit weight of fine grained soil. *Int J Geotech Eng* 5:297–305. <https://doi.org/10.3328/IJGE.2011.05.03.297-305>
- Vipulanandan C, Mohammed A (2020) Effect of drilling mud bentonite contents on the fluid loss and filter cake formation on a field clay soil formation compared to the API fluid loss method and characterized using Vipulanandan models. *J Pet Sci Eng* 189:107029
- Burroughs VS (2001) Quantitative criteria for the selection and stabilisation of soils for rammed earth wall construction. University of New South Wales Sydney, Australia
- Das BM (2021) Principles of geotechnical engineering. Cengage Learning
- Daita R, Drnevich V, Kim D (2005) Family of compaction curves for chemically modified soils. *Jt Transp Res Prog* 98
- David Suits L, Sheahan T, Inci G et al (2003) Experimental investigation of dynamic response of compacted clayey soils. *Geotech Test J* 26:10750. <https://doi.org/10.1520/GTJ11328J>
- Demiralay I, Guresinli YZ (2010) A study on the consistency limits and compactibility of the soils of Erzurum Plain. *J Ataturk Univ Fac Agric (Turk Lang)* 10:77–93
- Di Sante M (2020) On the compaction characteristics of soil-lime mixtures. *Geotech Geol Eng* 38:2335–2344. <https://doi.org/10.1007/s10706-019-01110-w>
- Djokovic K, Rakic D, Ljubojev M (2013) Estimation of soil compaction parameters based on the Atterberg limits. *Min Metall Eng Bor.* <https://doi.org/10.5937/mmweb1304001D>
- Duque J, Fuentes W, Rey S, Molina E (2020) Effect of grain size distribution on California Bearing Ratio (CBR) and modified proctor parameters for granular materials. *Arab J Sci Eng* 45:8231–8239. <https://doi.org/10.1007/s13369-020-04673-6>
- Dway SMM, Thant DAA (2014) Soil compression index prediction model for clayey soils. *Int J Sci Eng Tech Res* 3:2458–2462
- Farooq K, Khalid U, Mujtaba H (2016) Prediction of compaction characteristics of fine-grained soils using consistency limits. *Arab J Sci Eng* 41:1319–1328. <https://doi.org/10.1007/s13369-015-1918-0>
- Felt EJ (1965) Compactibility. In: *Methods of soil analysis: part 1 physical and mineralogical properties, including statistics of measurement and sampling*. Wiley Online Library, pp 400–412
- Firomsa W, Quezon ET (2019) Parametric modelling on the relationships between Atterberg limits and compaction characteristics of fine-grained soils. *Int J Adv Res Eng Appl Sci* 8:1–20
- Fondjo AA, Theron E, Ray RP (2021) Estimation of optimum moisture content and maximum dry unit weight of fine-grained soils using numerical methods. *Walailak J Sci Technol* 18:22722–22792
- Foreman DE, Daniel DE (1986) Permeation of compacted clay with organic chemicals. *J Geotech Eng* 112:669–681. [https://doi.org/10.1061/\(ASCE\)0733-9410\(1986\)112:7\(669\)](https://doi.org/10.1061/(ASCE)0733-9410(1986)112:7(669))
- Gunaydin O (2009) Estimation of soil compaction parameters by using statistical analyses and artificial neural networks. *Environ Geol* 57:203–215. <https://doi.org/10.1007/s00254-008-1300-6>

32. Gurtug Y, Sridharan A (2015) Prediction of compaction behaviour of soils at different energy levels. *Int J Eng Res Dev* 7:15–18. <https://doi.org/10.29137/umagd.379757>
33. Gurtug Y, Sridharan A (2002) Prediction of compaction characteristics of fine-grained soils. *Géotechnique* 52:761–763. <https://doi.org/10.1680/geot.2002.52.10.761>
34. Ali HFH (2023) Soft computing models to predict the compaction characteristics from physical soil properties. *Eng Technol J* 41(05):698–715
35. Hama Ali HF (2023) Utilizing multivariable mathematical models to predict maximum dry density and optimum moisture content from physical soil properties. *Multiscale and Multi Model Exp Des* 6(4):603–627
36. Han Z, Vanapalli SK (2016) Relationship between resilient modulus and suction for compacted subgrade soils. *Eng Geol* 211:85–97. <https://doi.org/10.1016/j.enggeo.2016.06.020>
37. Harris MT (1969) A study of the correlation potential of the optimum moisture content, maximum dry density, and consolidated drained shear strength of plastic fine-grained soils with index properties
38. Hassan J, Alshameri B, Iqbal F (2021) Prediction of California Bearing Ratio (CBR) using index soil properties and compaction parameters of low plastic fine-grained soil. *Transp Infrastruct Geotechnol*. <https://doi.org/10.1007/s40515-021-00197-0>
39. Vieira e Silva, A., Leme, R. F., da Silva Filho, F. C., Moura, T. E., & Ayala, G. R. L. (2021). Empirical models to predict compaction parameters for soils in the State of Ceará, North-eastern Brazil.
40. Hong L (2008) Optimization and management of materials in earthwork construction. PhD thesis, Iowa State University
41. Hopkins TC (1970) Relationship between soil support value and Kentucky CBR. Res Report, Commonw Kentucky, Dep Highw Frankfort, Kentucky
42. Horpibulsuk S, Katkan W, Apichatvullop A (2008) An approach for assessment of compaction curves of fine grained soils at various energies using a one point test. *Soils Found* 48:115–125. <https://doi.org/10.3208/sandf.48.115>
43. Horz RC (1983) Evaluation of revised manual compaction rammers and laboratory compaction procedures. Army Eng Waterw Exp Stn Vicksbg MS Geotech Lab
44. Hussain A, Atalar C (2020) Estimation of compaction characteristics of soils using Atterberg limits. *IOP Conf Ser Mater Sci Eng* 800:012024. <https://doi.org/10.1088/1757-899X/800/1/012024>
45. Hussain AHA (2016) Prediction of compaction characteristics of over-consolidated soils. MSc thesis, Near East University
46. Ibrahim SF (2013) Baghdad subgrade resilient modulus and liquefaction evaluation for pavement design using load cyclic triaxial strength. *J Environ Earth Sci* 3:125
47. Isik F, Ozden G (2013) Estimating compaction parameters of fine- and coarse-grained soils by means of artificial neural networks. *Environ Earth Sci* 69:2287–2297. <https://doi.org/10.1007/s12665-012-2057-5>
48. Jalal FE, Xu Y, Iqbal M et al (2021) Predicting the compaction characteristics of expansive soils using two genetic programming-based algorithms. *Transp Geotech* 30:100608. <https://doi.org/10.1016/j.trgeo.2021.100608>
49. Jesmani M, Manesh AN, Hoseini SMR (2008) Optimum water content and maximum dry unit weight of clayey gravels at different compactive efforts. *EJGE* 13:1–14
50. Ahmed C, Mohammed A, Tahir A (2020) Geostatistics of strength, modeling and GIS mapping of soil properties for residential purpose for Sulaimani City soils, Kurdistan Region, Iraq. *Model Earth Syst Environ* 6(2):879–893
51. Alim MA, Akhi AH, Alam MA, Roknuzzaman M (2021) Prediction of compaction characteristics of soil using plastic limit. In *Proceedings of International Conference on Planning, Architecture & Civil Engineering held at Rajshahi University of Engineering & Technology, Rajshahi, Bangladesh* (pp. 215–220)
52. Karakan E, Demir S (2018) Effect of fines content and plasticity on undrained shear strength of quartz-clay mixtures. *Arab J Geosci* 11:1–12. <https://doi.org/10.1007/s12517-018-4114-1>
53. Karimpour-Fard M, Machado SL, Falamaki A et al (2019) Prediction of compaction characteristics of soils from index test's results. *Iran J Sci Technol Trans Civ Eng* 43:231–248
54. Kaushal V, Guleria SP (2015) Geotechnical investigation of black cotton soils. *Int J Adv Eng Sci* 5:15–22
55. Kolay E, Baser T (2014) Estimating of the dry unit weight of compacted soils using general linear model and multi-layer perceptron neural networks. *Appl Soft Comput* 18:223–231. <https://doi.org/10.1016/j.asoc.2014.01.033>
56. Korde M, Yadav RK (2015) A study of correlation between CBR value and physical properties of some soils. *Int J Emerg Technol Adv Eng J* 5:237–239
57. Kulhawy FH (1991) Drilled shaft foundations. *Foundation engineering handbook*. Springer, Boston, pp 537–552
58. Liang RY, Rabab'ah S, Khasawneh M (2008) Predicting moisture-dependent resilient modulus of cohesive soils using soil suction concept. *J Transp Eng* 134:34–40. [https://doi.org/10.1061/\(ASCE\)0733-947X\(2008\)134:1\(34\)](https://doi.org/10.1061/(ASCE)0733-947X(2008)134:1(34))
59. Luczak-Wilamowska B (2004) Basic soil properties of a number of artificial clay-sand mixtures determined as a function of sand content. In: *Engineering geology for infrastructure planning in Europe*. Springer, pp 308–315
60. Machado SL, Carvalho MF, Carvalho ZS et al (2020) Optimal ranges of soil index properties for diesel containment using compacted barriers. *Environ Geotech* 7:540–553. <https://doi.org/10.1680/jenge.18.00076>
61. Vipulanandan C, Mohammed AS (2021) 3-dimension stresses and new failure model to predict behavior of clay soils in various liquid limit ranges. *Arab J Geosci* 14(3):160
62. Mazari M, Nazarian S (2017) Mechanistic approach for construction quality management of compacted geomaterials. *Transp Geotech* 13:92–102. <https://doi.org/10.1016/j.trgeo.2017.08.001>
63. Mejias-Santiago M, Berney I V, Ernest S, Bradley CT (2013) Evaluation of a non-nuclear soil density gauge on fine-grained soils
64. Miller CJ, Yesiller N, Yaldo K, Merayyan S (2002) Impact of soil type and compaction conditions on soil water characteristic. *J Geotech Geoenviron Eng* 128:733–742. [https://doi.org/10.1061/\(ASCE\)1090-0241\(2002\)128:9\(733\)](https://doi.org/10.1061/(ASCE)1090-0241(2002)128:9(733))
65. Mohammad LN, Huang B, Puppala AJ, Allen A (1999) Regression model for resilient modulus of subgrade soils. *Transp Res Rec J Transp Res Board* 1687:47–54. <https://doi.org/10.3141/1687-06>
66. Mujtaba H (2015) Development of correlations between various geotechnical parameters for granular soils in Punjab. PhD Thesis, Univ Eng Technol Lahore, Pakistan
67. Mujtaba H, Farooq K, Sivakugan N, Das BM (2013) Correlation between gradational parameters and compaction characteristics of sandy soils. *Int J Geotech Eng* 7:395–401. <https://doi.org/10.1179/1938636213Z.00000000045>
68. Nagaraj HB, Reesha B, Sravan MV, Suresh MR (2015) Correlation of compaction characteristics of natural soils with modified plastic limit. *Transp Geotech* 2:65–77. <https://doi.org/10.1016/j.trgeo.2014.09.002>
69. Nagaraju TV, Gobinath R, Awoyera P, Abdy Sayyed MAH (2021) Prediction of California bearing ratio of subgrade soils using artificial neural network principles. In: *Communication and intelligent systems*. Springer, pp 133–146

70. Mawlood Y, Mohammed A, Hummadi R, Hasan A, Ibrahim H (2022) Modeling and statistical evaluations of unconfined compressive strength and compression index of the clay soils at various ranges of liquid limit. *J Test Eval* 50(1):551–569
71. Nesamatha R, Arumairaj PD (2015) Numerical modeling for prediction of compression index from soil index properties. *Electron J Geotech Eng* 20:4369–4378
72. Ng K, Chew YM, Osman MH, SK MG (2015) Estimating maximum dry density and optimum moisture content of compacted soils. *Int Confer Adv Civ Environ Eng B* 1–8
73. Nwaiwu CMO, Mezie EO (2021) Prediction of maximum dry unit weight and optimum moisture content for coarse-grained lateritic soils. *Soils Rocks* 44:1–10. <https://doi.org/10.28927/SR.2021.054120>
74. Ogbuchukwu PO, Okeke OC, Ahiarakwem CA, Ozotta OO (2019) Geotechnical properties of expansive soils in Awka and environs, Southeastern Nigeria, in relation to engineering problems. *Int J Appl Sci Res* 2:79–94
75. Olmez A (2007) Determination of compaction parameters by means of regression approaches. Master's thesis, Nigde University (Turkish Lang)
76. Omar M, Shanableh A, Basma A, Barakat S (2003) Compaction characteristics of granular soils in United Arab Emirates. *Geotech Geol Eng* 21:283–295. <https://doi.org/10.1023/A:1024927719730>
77. Omar M, Shanableh A, Mughieda O et al (2018) Advanced mathematical models and their comparison to predict compaction properties of fine-grained soils from various physical properties. *Soils Found* 58:1383–1399. <https://doi.org/10.1016/j.sandf.2018.08.004>
78. Ören AH (2014) Estimating compaction parameters of clayey soils from sediment volume test. *Appl Clay Sci* 101:68–72. <https://doi.org/10.1016/j.clay.2014.07.019>
79. Othman K, Abdelwahab H (2021) Prediction of the soil compaction parameters using deep neural networks. *Transp Infrastruct Geotechnol*. <https://doi.org/10.1007/s40515-021-00213-3>
80. Parkoh EA (2016) Prediction of compaction characteristics of lateritic soils in Ghana. Unpubl Master's thesis, Near East University
81. Pillai GS, Vinod P (2019) A framework for prediction of compaction parameters in standard proctor tests. In: *Recent advances in materials, mechanics and management: proceedings of the 3rd international conference on materials, mechanics and management (IMMM 2017)*, July 13–15, 2017, Trivandrum, Kerala, India, p 9
82. Piro NS, Mohammed AS, Hamad SM (2022) The impact of GGBS and ferrous on the flow of electrical current and compressive strength of concrete. *Constr Build Mater* 349:128639
83. Prashanth JP, Sivapullaiah PV, Sridharan A (2001) Pozzolanic fly ash as a hydraulic barrier in land fills. *Eng Geol* 60(1–4):245–252
84. Raju NV, Srimurali M, Prasad KN (2014) Functional correlations between compaction characteristics, un-drained shear strength and Atterberg limits. *IOSR J Mech Civ Eng* 11:109–115
85. Ramasubbarao G, Sankar SG (2013) Predicting soaked CBR value of fine grained soils using index and compaction characteristics. *Jordan J Civ Eng* 7:354–360
86. Ramiah BK, Viswanath V, Krishnamurthy H V (1970) Interrelationship of compaction and index properties. In: *Proc. 2nd South East Asian Conf on Soil Eng*, p 577
87. Ratnam UV, Prasad KN (2019) Prediction of compaction and compressibility characteristics of compacted soils. *Int J Appl Eng Res* 14:621–632
88. Ring G, Sallberg JR, Collins WH (1962) Correlation of compaction and classification test data. *Hwy Res Bull* 325:55–75
89. Saikia A, Baruah D, Das K et al (2017) Predicting compaction characteristics of fine-grained soils in terms of Atterberg limits. *Int J Geosynth Gr Eng* 3:18. <https://doi.org/10.1007/s40891-017-0096-4>
90. Verdugo R (2008) Singularities of geotechnical properties of complex soils in seismic regions. *J Geotech Geoenviron eng* 134(7):982–991.
91. Sathawara JK, Patel AK (2013) Comparison between soaked and unsoaked CBR. *Int J Adv Eng Res Stud* 2:132–135
92. Sawangsuriya A, Edil TB, Bosscher PJ (2009) Modulus-suction-moisture relationship for compacted soils in postcompaction state. *J Geotech Geoenviron Eng* 135:1390–1403. [https://doi.org/10.1061/\(ASCE\)GT.1943-5606.0000108](https://doi.org/10.1061/(ASCE)GT.1943-5606.0000108)
93. Sen B, Pal SK (2017) Compaction and consolidation characteristics of soils and correlations of parameters. *Int J Eng Technol Sci Res* 4:874–886
94. Setiawan B (2016) The preliminary study on the effect of coarse particles content on OMC and maximum dry unit weight: a case of Aceh's fill materials. *Aceh Int J Sci Technol* 5:75–81. <https://doi.org/10.13170/aijst.5.2.4877>
95. Shirur NB, Hiremath SG (2014) Establishing relationship between CBR value and physical properties of soil. *IOSR J Mech Civ Eng* 11:26–30
96. Shook JF, Fang HY (1961) Cooperative materials testing programs at the AASHO road test. *Highw Res Board Spec Rep* (66)
97. Sinha SK, Wang MC (2008) Artificial neural network prediction models for soil compaction and permeability. *Geotech Geol Eng* 26:47–64. <https://doi.org/10.1007/s10706-007-9146-3>
98. Sivrikaya O (2008) Models of compacted fine-grained soils used as mineral liner for solid waste. *Environ Geol* 53:1585–1595. <https://doi.org/10.1007/s00254-007-1142-7>
99. Sivrikaya O, Kayadelen C, Cecen E (2013) Prediction of the compaction parameters for coarse-grained soils with fines content by MLA and GEP. *Acta Geotech Slov* 10:29–41
100. Sivrikaya O, Soycan TY (2011) Estimation of compaction parameters of fine-grained soils in terms of compaction energy using artificial neural networks. *Int J Numer Anal Methods Geomech* 35:1830–1841. <https://doi.org/10.1002/nag.981>
101. Sivrikaya O, Soycan YT (2009) Estimation of compaction parameters of fine-grained soils using Artificial neural networks. In: *Proc. 2nd international conference on new developments in soil mechanics and geotechnical engineering*, pp 406–412
102. Soltani A, Deng A, Taheri A, Sridharan A (2019) Consistency limits and compaction characteristics of clay soils containing rubber waste. *Proc Inst Civ Eng—Geotech Eng* 172:174–188. <https://doi.org/10.1680/jgeen.18.00042>
103. Sotelo, M. J. (2012). Evaluation of non-nuclear devices in measuring moisture content and density of soils. The University of Texas at El Paso.
104. Sridharan A, Nagaraj HB (2005) Plastic limit and compaction characteristics of finegrained soils. *Proc Inst Civ Eng—Gr Improv* 9:17–22. <https://doi.org/10.1680/grim.2005.9.1.17>
105. Sridharan A, Rao SM, Joshi S (1990) Classification of expansive soils by sediment volume method. *Geotech Test J* 13:375–380. <https://doi.org/10.1520/GTJ10181J>
106. Taha OME, Majeed ZH, Ahmed SM (2018) Artificial neural network prediction models for maximum dry density and optimum moisture content of stabilized soils. *Transp Infrastruct Geotechnol* 5:146–168. <https://doi.org/10.1007/s40515-018-0053-2>
107. Talukdar DK (2014) A study of correlation between California Bearing Ratio (CBR) value with other properties of soil. *Int J Emerg Technol Adv Eng* 4:559–562
108. Tenpe A, Kaur S (2015) Artificial neural network modeling for predicting compaction parameters based on index properties of soil. *Int J Sci Res* 4:1198–1202

109. Tsegaye T, Fikre H, Abebe T (2017) Correlation between compaction characteristics and Atterberg limits of fine grained soils found in Addis Ababa. *Int J Sci Eng Res* 8:357–364
110. Verma G, Kumar B (2022) Artificial neural network equations for predicting the modified proctor compaction parameters of fine-grained soil. *Transp Infrastruct Geotechnol*. <https://doi.org/10.1007/s40515-022-00228-4>
111. Verma G, Kumar B (2022) Multi-layer perceptron (MLP) neural network for predicting the modified compaction parameters of coarse-grained and fine-grained soils. *Innov Infrastruct Solut* 7:78. <https://doi.org/10.1007/s41062-021-00679-7>
112. Wang MC, Huang CC (1984) Soil compaction and permeability prediction models. *J Environ Eng* 110:1063–1083. [https://doi.org/10.1061/\(ASCE\)0733-9372\(1984\)110:6\(1063\)](https://doi.org/10.1061/(ASCE)0733-9372(1984)110:6(1063))
113. Wells JER (2014) Calibration of non-nuclear devices for construction quality control of compacted soils. MSc thesis, University Kentucky
114. Wesley LD (1973) Some basic engineering properties of halloysite and allophane clays in Java, Indonesia. *Géotechnique* 23:471–494. <https://doi.org/10.1680/geot.1973.23.4.471>
115. Winta A (2016) Prediction of soaked california bearing ratio (Cbr) of fine grained soils from index properties. Case of Silty Clay Soils from Addis Abeba
116. Yoon H, Hyun Y, Ha K, Lee KK, Kim GB (2016) A method to improve the stability and accuracy of ANN-and SVM-based time series models for long-term groundwater level predictions. *Comput Geosci* 90:144–155
117. Zhang K, Frederick CN (2017) Experimental investigation on compaction and Atterberg limits characteristics of soils: Aspects of clay content using artificial mixtures. *KSCE J Civ Eng* 21:546–553. <https://doi.org/10.1007/s12205-017-1580-z>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.