Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 1, 2026

EXPLAINABLE SUPPORT VECTOR REGRESSION FRAMEWORK FOR PREDICTING MAXIMUM DRY DENSITY FROM SOIL INDEX AND GRADATION PROPERTIES

Yaser Farman[*1], Ijaz Ahmad[2], Abdul Wahab[3], Akram Ullah Khan[4]

*[*1]MS Geotechnical Engineering, National Institute of Transportation (NIT), National University of Science and Technology (NUST).*
[2]college of civil engineering Tongji university Shanghai
[3]Assistant Professor, Department of Civil Engineering, GIKI, Topi Swabi
[4]Department of Civil Engineering, University of Engineering and Technology, Peshawar

[*1]yaserfarman@gmail.com, [2]mijazsial438@gmail.com, [3]awahab@giki.edu.pk, [4]17pwciv4882@uetpeshawar.edu.pk

DOI: https://doi.org/10.5281/zenodo.18309830

Keywords
Maximum dry density, soil compaction, support vector regression, machine learning, SHAP, partial dependence plots, geotechnical engineering, soil index properties

Article History
Received: 21 November 2025
Accepted: 07 January 2026
Published: 20 January 2026

Corresponding Author: *
Yaser Farman

Abstract

Maximum dry density (MDD) is a critical parameter in geotechnical engineering, governing the strength, stiffness, and long-term performance of compacted earth structures. Conventional laboratory determination of MDD through Proctor testing is time-consuming and labor-intensive, particularly when multiple soil sources or stabilization scenarios must be evaluated. This study presents a robust and interpretable machine learning framework based on Support Vector Regression (SVR) with a radial basis function (RBF) kernel for predicting MDD directly from routinely measured soil index and gradation properties. A dataset comprising 486 soil samples was compiled, incorporating gravel content, sand content, fines content, liquid limit, plastic limit, and optimum moisture content as input variables.

Model development involved systematic data preprocessing, feature standardization, hyperparameter optimization, and rigorous validation using training–validation–testing splits and 20-fold cross-validation. Model performance was evaluated using multiple statistical indicators, including R^2, RMSE, MAE, MAPE, and CVRMSE, alongside graphical diagnostics such as learning curves, residual analysis, and percentage error distributions. The SVR model achieved strong predictive accuracy, with R^2 values ranging from 0.86 to 0.92 and MAPE values close to 2%, demonstrating excellent generalization capability and minimal bias.

To enhance transparency and engineering interpretability, SHapley Additive exPlanations (SHAP) and partial dependence plots were employed. These analyses identified optimum moisture content as the dominant predictor of MDD, followed by fines and gravel content, consistent with established soil compaction theory. Nonlinear trends revealed by SHAP and PDPs further confirmed the physical realism of the learned relationships.

Overall, the proposed SVR-based, explainable framework provides an accurate and interpretable surrogate for laboratory compaction testing, offering significant

potential to improve efficiency and decision-making in geotechnical design and construction practice.

1 INTRODUCTION

Soil compaction is a fundamental process in geotechnical engineering, directly governing the performance, serviceability, and long-term safety of earth structures such as embankments, road pavements, earth dams, landfills, foundations, and reinforced soil systems (Zhao and Guan, 2024; Omar et al., 2018; Gunaydin, Özbeyaz and Söylemez, 2018; Boadu, 2020; Zhang, Yin and Jin, 2021). Among compaction parameters, **maximum dry density (MDD)** is especially critical because it represents the highest mass of soil that can be compacted per unit volume under specified energy and moisture conditions, and thus controls stiffness, shear strength, compressibility, and hydraulic behaviour (Zhao and Guan, 2024; Omar et al., 2018; Gunaydin, Özbeyaz and Söylemez, 2018; Zhao et al., 2024; Zhang, Yin and Jin, 2021). Achieving an adequate fraction of the laboratory-determined MDD in the field is a standard quality-control requirement in design specifications, and insufficient compaction is strongly associated with excessive settlements, loss of bearing capacity, differential movements, and premature distress in transportation and geotechnical infrastructure (Zhao and Guan, 2024; Omar et al., 2018; Gunaydin, Özbeyaz and Söylemez, 2018; Almuaythir, Zaini and Lodhi, 2025; Zhang, Yin and Jin, 2021). MDD is influenced by a complex interplay of soil gradation, fines content, plasticity, mineralogy, specific gravity, and moisture content, as well as by compaction energy and method (Zhao and Guan, 2024; Omar et al., 2018; Khatti and Grover, 2023; Gunaydin, Özbeyaz and Söylemez, 2018; Almuaythir, Zaini and Lodhi, 2025; Zhao et al., 2024; Zhang, Yin and Jin, 2021)(Gali et al., 2023; Shah et al., 2023; Khan et al., 2025; Nawaz et al., 2025). Conventional practice relies on laboratory Proctor tests to obtain the compaction curve and identify both MDD and optimum moisture content (OMC), but these tests are time-consuming, labour-intensive, and require repeated trials for different mix designs, stabilizer dosages, or borrow sources (Zhao and Guan, 2024; Omar et al., 2018; Gunaydin, Özbeyaz and Söylemez, 2018; Almuaythir, Zaini and Lodhi, 2025; Zhao et al., 2024; Zhang, Yin and Jin, 2021). In large-scale projects, or where many alternative stabilisation scenarios must be screened, exhaustive laboratory testing becomes impractical and may delay design iterations and construction decisions (Zhao and Guan, 2024; Omar et al., 2018; Almuaythir, Zaini and Lodhi, 2025; Zhang, Yin and Jin, 2021).

These challenges have motivated sustained interest in empirical and analytical models that relate MDD and OMC to basic soil index properties, such as Atterberg limits, linear shrinkage, and grain-size distribution (Omar et al., 2018; Gunaydin, Özbeyaz and Söylemez, 2018; Almuaythir, Zaini and Lodhi, 2025; Zhang, Yin and Jin, 2021). Early correlation charts and regression equations provided simple tools but generally assumed linear or weakly nonlinear relations and were often calibrated to narrow soil types, limiting extrapolation to broader geological and compositional ranges (Omar et al., 2018; Gunaydin, Özbeyaz and Söylemez, 2018; Zhang, Yin and Jin, 2021). With the growing diversity of materials used in modern earthworks including chemically stabilized soils, industrial by-products, recycled aggregates, and mixtures with various binders and additives the governing relations between index properties, gradation, moisture, and compaction response have become increasingly nonlinear and multidimensional (Zhao and Guan, 2024; Ngo, Nguyen and Tran, 2022; Khatti and Grover, 2023; Hassan and Beshr, 2024; Mustafa et al., 2025; Zhao et al., 2024; Zhang, Yin and Jin, 2021; Onyelowe et al., 2024). Traditional linear regression or simple empirical charts typically fail to capture such interactions, leading to reduced predictive accuracy for complex or stabilized systems (Omar et al., 2018; Gunaydin, Özbeyaz and Söylemez, 2018; Mustafa et al., 2025; Zhang, Yin and Jin, 2021; Onyelowe et al., 2024). Consequently, there is a clear need for **robust data-driven models** that can accommodate nonlinearities and interactions while retaining generalization capability and physical interpretability.

Machine learning (ML) methods have recently emerged as powerful tools for predicting geotechnical properties from easily measurable inputs, including applications to MDD, OMC, unconfined compressive strength (UCS), modulus, hydraulic conductivity, settlement, and cone index (Zhao and Guan, 2024; Omar et al., 2018; Ngo, Nguyen and Tran, 2022; Khatti and Grover, 2023; Gunaydin, Özbeyaz and Söylemez, 2018; Hassan and Beshr, 2024; Alnmr et al., 2024; Mustafa et al., 2025; Almuaythir, Zaini and Lodhi, 2025; Boadu, 2020; Ballabio, 2009; Roy and Chakraborty, 2023; Wu and Zhou, 2022; Zhao et al., 2024; Zhang, Yin and Jin, 2021; Onyelowe et al., 2024; Wan, 2023). Techniques such as artificial neural networks (ANN), support vector regression (SVR), Gaussian process regression, random forest (RF), gradient boosting (GB/XGBoost), long short-term memory (LSTM) networks, and hybrid meta-heuristic–ML frameworks have all been explored for compaction-related tasks (Zhao and Guan, 2024; Omar et al., 2018; Ngo, Nguyen and Tran, 2022; Khatti and Grover, 2023; Gunaydin, Özbeyaz and Söylemez, 2018; Hassan and Beshr, 2024; Alnmr et al., 2024; Mustafa et al., 2025; Almuaythir, Zaini and Lodhi, 2025; Ballabio, 2009; Wu and Zhou, 2022; Zhao et al., 2024; Zhang, Yin and Jin, 2021; Onyelowe et al., 2024; Wan, 2023). In the context of predicting MDD and OMC from soil index and gradation parameters, several studies have demonstrated that ML models can significantly outperform multiple linear regression, with reported R^2 values commonly exceeding 0.90 and relative errors falling into ranges suitable for engineering design (Zhao and Guan, 2024; Omar et al., 2018; Khatti and Grover, 2023; Gunaydin, Özbeyaz and Söylemez, 2018; Mustafa et al., 2025; Wu and Zhou, 2022; Zhao et al., 2024; Zhang, Yin and Jin, 2021). Comparative works further indicate that, for many geotechnical regression problems, SVR and related kernel-based methods can match or exceed ANN performance, particularly when datasets are limited or noisy, due to stronger regularization and structural risk minimization (Omar et al., 2018; Khatti and Grover, 2023; Zhang and O'Donnell, 2020; Alnmr et al., 2024; Boadu, 2020; Ballabio, 2009; Roy and Chakraborty, 2023; Wu and Zhou, 2022; Zhang, Yin and Jin, 2021).

Support vector regression, derived from the broader support vector machine (SVM) framework, constructs regression functions in a high-dimensional feature space through kernel mappings and seeks to minimize an upper bound on generalization error rather than training error alone (Zhang and O'Donnell, 2020; Boadu, 2020; Ballabio, 2009; Roy and Chakraborty, 2023). This structural risk minimization principle, combined with convex optimization, endows SVR with strong theoretical guarantees and robustness against overfitting, especially when data are moderate in size, high-dimensional, or exhibit complex nonlinear patterns (Zhang and O'Donnell, 2020; Boadu, 2020; Ballabio, 2009; Roy and Chakraborty, 2023). In geotechnical engineering, SVR has been successfully applied to the prediction of compaction parameters, UCS of stabilized and unstabilized soils, settlement of shallow foundations, geotechnical properties from electrical spectra, and soil cone index, among other targets (Omar et al., 2018; Ngo, Nguyen and Tran, 2022; Gunaydin, Özbeyaz and Söylemez, 2018; Hassan and Beshr, 2024; Alnmr et al., 2024; Almuaythir, Zaini and Lodhi, 2025; Boadu, 2020; Ballabio, 2009; Roy and Chakraborty, 2023; Zhao et al., 2024; Onyelowe et al., 2024; Wan, 2023). Reviews of SVM in geotechnical and structural reliability contexts consistently emphasize its suitability for high-dimensional problems where obtaining large training datasets is expensive and where accurate surrogate models are required to replace costly experiments or numerical simulations (Boadu, 2020; Ballabio, 2009; Roy and Chakraborty, 2023).

Within the specific domain of MDD prediction, both individual and hybrid SVR frameworks have been proposed. Studies on stabilized soils have used SVR alone or combined with meta-heuristic optimizers—such as artificial rabbits optimization, manta ray foraging algorithms, Harris hawks optimization, and generalized normal distribution optimization—to tune hyperparameters and enhance performance (Zhao and Guan, 2024; Ngo, Nguyen and Tran, 2022; Zhao et al., 2024). These hybrid models have reported R^2 values frequently above 0.99 and markedly reduced RMSE, underscoring the potential of SVR-based approaches to serve as accurate surrogates for laboratory compaction testing (Zhao and Guan, 2024; Ngo, Nguyen and Tran, 2022; Zhao et al., 2024). Other investigations have compared SVR with ANN, Gaussian process regression, XGBoost, RF, and LSTM on datasets of fine-grained or expansive soils, finding that SVR is often among the top performers or is outperformed only marginally by advanced ensemble methods or deep learning models, while maintaining advantages in training efficiency and model robustness (Omar et al., 2018; Khatti and Grover, 2023; Alnmr et al., 2024; Mustafa et al., 2025; Almuaythir, Zaini and Lodhi, 2025; Wu and Zhou, 2022; Zhang, Yin and Jin, 2021). Despite these advances, most existing works focus on either specific soil types (e.g., expansive clays or organic soils), particular stabilizers, or relatively narrow ranges of gradation and index properties (Zhao and Guan, 2024; Omar et al., 2018; Ngo, Nguyen and Tran, 2022; Khatti and Grover, 2023; Alnmr et al., 2024; Mustafa et al., 2025; Almuaythir, Zaini and Lodhi, 2025; Wu and Zhou, 2022; Zhao et al., 2024; Zhang, Yin and Jin, 2021; Onyelowe et al., 2024). There remains a need for systematically evaluated SVR models that target MDD prediction across more diverse soil conditions using readily available gradation and index parameters.

At the same time, interpretability has become a central concern in the adoption of ML models for geotechnical design. Many high-performing algorithms behave as "black boxes," making it difficult for engineers to understand how input variables such as moisture content, fines, or plasticity influence predicted outcomes, and to verify that model behaviour is consistent with established soil mechanics principles (Ngo, Nguyen and Tran, 2022; Hassan and Beshr, 2024; Alnmr et al., 2024; Ballabio, 2009; Wu and Zhou, 2022; Onyelowe et al., 2024; Wan, 2023). Recent work has introduced **SHAP (SHapley Additive exPlanations)** and related explainable AI techniques into geotechnical and materials applications, enabling the decomposition of predictions into additive feature contributions and providing both global and local insight into model behaviour (Ngo, Nguyen and Tran, 2022; Hassan and Beshr, 2024; Ballabio, 2009; Wu and Zhou, 2022; Onyelowe et al., 2024; Wan, 2023). In studies on stabilized soils, sustainable concrete, expansive clays, and UCS prediction, SHAP analyses have successfully identified dominant variables—such as cement content, dry density, fines content, or moisture state—and clarified nonlinear threshold and interaction effects, thereby bridging the gap between data-driven models and mechanistic understanding (Ngo, Nguyen and Tran, 2022; Hassan and Beshr, 2024; Alnmr et al., 2024; Ballabio, 2009; Wu and Zhou, 2022; Onyelowe et al., 2024; Wan, 2023).

Against this backdrop, the present research develops and evaluates a Support Vector Regression model with a radial basis function kernel to estimate maximum dry density directly from soil gradation and index properties, without recourse to repeated Proctor testing. Building on the demonstrated strengths of SVR in handling nonlinear, high-dimensional relationships under limited data conditions, the study rigorously assesses predictive performance using a comprehensive suite of statistical metrics and k-fold cross-validation, and examines residuals and percentage-error distributions to verify robustness and absence of bias, in line with emerging best practice for SVR-based geotechnical surrogates (Omar et al., 2018; Khatti and Grover, 2023; Gunaydin, Özbeyaz and Söylemez, 2018; Alnmr et al., 2024; Almuaythir, Zaini and Lodhi, 2025; Boadu, 2020; Ballabio, 2009; Roy and Chakraborty, 2023; Wu and Zhou, 2022; Zhang, Yin and Jin, 2021; Onyelowe et al., 2024; Wan, 2023). Furthermore, SHAP is employed to interpret the learned relationships between input features—such as optimum moisture content, fines fraction, and coarse fractions—and the resulting

MDD predictions, situating the model behaviour within established compaction theory and providing transparent guidance for design and material selection (Zhao and Guan, 2024; Omar et al., 2018; Ngo, Nguyen and Tran, 2022; Khatti and Grover, 2023; Hassan and Beshr, 2024; Alnmr et al., 2024; Almuaythir, Zaini and Lodhi, 2025; Ballabio, 2009; Wu and Zhou, 2022; Zhao et al., 2024; Zhang, Yin and Jin, 2021; Onyelowe et al., 2024; Wan, 2023). By combining robust kernel-based learning with explainable AI, this work aims to deliver a state-of-the-art, yet practically interpretable, framework for reliable prediction of maximum dry density from routine soil characterization data.

2 Methodology

2.1 Data Preprocessing and feature selection

The dataset used in this study consists of **486 data points**, each characterized by **six input features**—gravel content, sand content, fines content, liquid limit, plastic limit, and optimum moisture content—along with **maximum dry density** as the target variable. The dataset encompasses a wide range of soil types and compaction conditions, as reflected by the substantial variability in both compositional and index properties. This sample size and feature set provide sufficient diversity and statistical richness to support reliable model training, validation, and testing, while enabling the SVR model to capture complex, nonlinear relationships governing soil compaction behavior.

A comprehensive statistical and distributional analysis was conducted to characterize the input variables and target parameter used in the development of the SVR model. The frequency distribution curves and descriptive statistics reveal substantial variability in soil composition and index properties, reflecting a diverse and representative dataset suitable for robust model training. **Gravel Content** (%) exhibits a highly right-skewed distribution, with values ranging from 0 to 97% and a high standard deviation (22.61) and variance (511.08), indicating significant heterogeneity in coarse-grained fractions. In contrast, **Sand Content** (%) displays a more symmetric, bell-shaped distribution centred around a mean of 40.58%, suggesting that sand-dominated soils are prevalent within the dataset. **Fines Content** (%) also shows wide dispersion (variance of 545.60), confirming the presence of both coarse-grained and fine-grained soil types.

The consistency limits further illustrate variability in soil plasticity characteristics. **Liquid Limit** (%) follows an approximately normal distribution with moderate spread (standard deviation of 10.99), while **Plastic Limit** (%) shows a narrower distribution and lower variance (45.76), indicating comparatively less variability in plastic behavior. **Optimum Moisture Content** (%) exhibits a right-skewed distribution, with most values concentrated between 10% and 18%, consistent with typical compaction characteristics of natural soils.

The target variable, **Maximum Dry Density (kN/m³)**, demonstrates a near-normal distribution with a mean of 18.52 kN/m³ and a relatively low standard deviation of 1.81, indicating stable central tendency despite variability in input parameters. The interquartile range (17.20–19.94 kN/m³) further suggests that most observations cluster around typical compaction densities, while extreme values remain limited.

Overall, the descriptive statistics and distributional plots confirm that the dataset encompasses a wide spectrum of soil gradation, plasticity, and compaction behavior. The presence of substantial variance in key predictors, combined with a well-behaved target distribution, enhances the learning potential of the SVR model and supports its ability to generalize across diverse soil conditions. These characteristics are essential for developing a reliable and physically meaningful predictive framework for maximum dry density estimation.

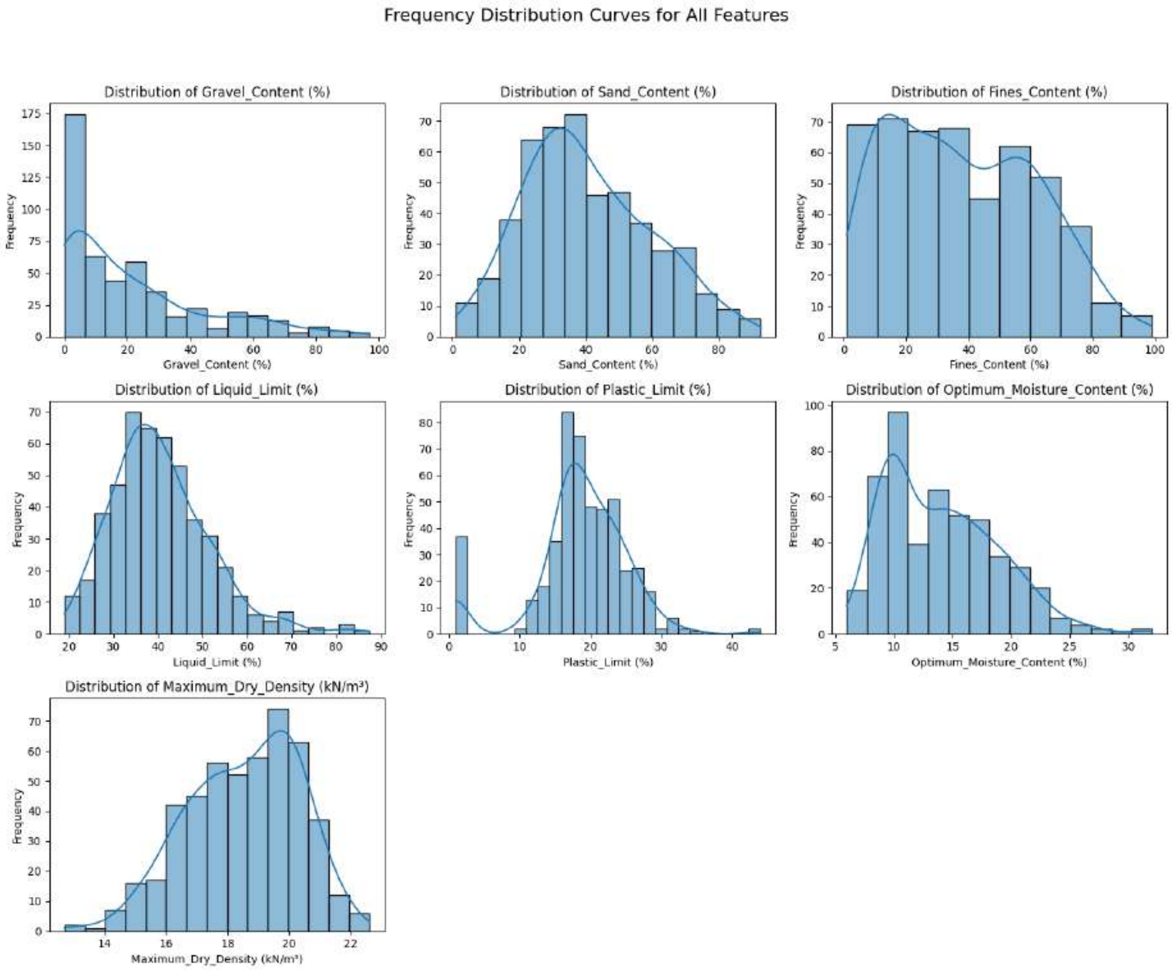


Figure 1 Frequency distribution curves and histograms of input variables and the target parameter, illustrating the variability and distributional characteristics of soil composition, consistency limits, optimum moisture content, and maximum dry density.

Table 1 Descriptive statistical summary of input features and maximum dry density, including mean, standard deviation, quartiles, range, and variance, highlighting the diversity and spread of the dataset used for model development.

Feature	Mean	Std. Dev.	Min	25th Percentile	Median	75th Percentile	Max
Gravel Content (%)	21.35	22.61	0.00	3.27	13.88	30.20	97.00
Sand Content (%)	40.58	19.16	1.00	26.77	37.53	53.61	92.70
Fines Content (%)	38.10	23.36	1.00	17.95	35.05	57.00	99.00
Liquid Limit (%)	40.47	10.99	19.00	33.00	39.15	46.70	87.50
Plastic Limit (%)	18.60	6.76	1.00	16.10	18.81	22.87	44.00
Optimum Moisture Content (%)	14.03	4.78	6.00	10.00	13.50	17.23	32.00
Maximum Dry Density (kN/m³)	**18.52**	**1.81**	**12.70**	**17.20**	**18.70**	**19.94**	**22.60**

Feature	Variance
Gravel Content (%)	511.08
Sand Content (%)	366.94
Fines Content (%)	545.60
Liquid Limit (%)	120.83
Plastic Limit (%)	45.76
Optimum Moisture Content (%)	22.83
Maximum Dry Density (kN/m³)	3.28

2.2 Machine learning model

In this study, **Support Vector Regression (SVR)** was employed to predict the maximum dry density of soils due to its strong capability in modeling nonlinear relationships and handling complex, high-dimensional data. SVR is an extension of Support Vector Machines (SVM) for regression problems and operates by constructing an optimal hyperplane that minimizes prediction error while maintaining model generalization. Unlike traditional regression techniques, SVR introduces an ε-insensitive loss function, which allows errors within a specified tolerance to be ignored, thereby improving robustness against noise in experimental data.

To capture the nonlinear behavior inherent in soil compaction processes, the **radial basis function (RBF) kernel** was adopted. The RBF kernel effectively maps the input features into a higher-dimensional space, enabling the model to learn complex interactions between soil gradation, plasticity, and moisture-related parameters. Key hyperparameters, including the regularization parameter (C), kernel width (γ), and ε, were carefully tuned to balance model bias and variance.

SVR is particularly suitable for geotechnical applications due to its ability to perform well with limited datasets and its resistance to overfitting. The model was trained using standardized input features to ensure numerical stability and consistent learning behavior. Overall, the SVR framework provides a reliable and flexible predictive tool, capable of accurately estimating maximum dry density while maintaining strong generalization performance across unseen data.

2.3 Model development

The development of the Support Vector Regression (SVR) model followed a systematic and rigorous workflow to ensure robustness, accuracy, and generalization capability. Initially, the dataset comprising 486 samples was carefully examined for completeness and consistency. All input variables were standardized to zero mean and unit variance to eliminate scale effects and to improve numerical stability during model training. The dataset was then randomly partitioned into training, validation, and testing subsets to allow unbiased model evaluation and hyperparameter optimization.

An SVR model with a radial basis function (RBF) kernel was implemented to capture the nonlinear relationships between soil properties and maximum dry density. Hyperparameters, including the regularization parameter (C), kernel coefficient (γ), and the ε-insensitive loss function, were optimized using cross-validation to achieve an optimal balance between model complexity and predictive performance. A 20-fold cross-validation strategy was adopted to minimize overfitting and to ensure the stability of model predictions across different data splits.

Model performance was evaluated using multiple statistical indicators, including the coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), coefficient of variation of RMSE (CVRMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). In addition, graphical diagnostics such as learning curves, residual analysis, and actual-versus-predicted plots were employed to further assess model behavior. This comprehensive model development framework ensured that the SVR model achieved high predictive accuracy while maintaining strong generalization capability and interpretability.

2.4 Performance assessment of models

The performance of the developed SVR model was comprehensively assessed using a combination of statistical metrics and graphical diagnostics to evaluate its accuracy, robustness, and generalization capability. Quantitative evaluation was conducted on the training, validation, and testing datasets to ensure consistent performance across different data splits. The coefficient of determination (R^2) was used to measure the model's ability to explain the variance in maximum dry density, while error-based metrics such as mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) were employed to quantify prediction accuracy. The coefficient of variation of RMSE (CVRMSE) was additionally considered to assess error magnitude relative to the mean of the target variable.

To further examine model stability and learning behavior, learning curves were analyzed to identify potential underfitting or overfitting. Cross-validation results provided insight into performance variability across different subsets of the data. Graphical evaluations, including actual versus predicted plots, residual distributions, and Q–Q plots, were used to visually assess prediction bias, error symmetry, and normality assumptions. Moreover, percentage error analysis helped evaluate the distribution and magnitude of relative prediction errors.

Collectively, these evaluation approaches provided a robust and multifaceted assessment of the SVR model, confirming its reliability, predictive accuracy, and suitability for estimating maximum dry density under varying soil conditions.

3 Results

3.1 Support vector regression results

The predictive performance of the Support Vector Regression (SVR) model with a radial basis function (RBF) kernel demonstrates a high level of robustness and generalization capability in estimating maximum dry density from soil gradation and index properties. The model achieved consistently strong coefficients of determination across all data subsets, with R^2 values of 0.9207 for the training set, 0.9015 for the validation set, and 0.8640 for the testing set, indicating that a substantial proportion of the variance in maximum dry density is effectively captured by the selected input features. Error-based metrics further confirm the model's accuracy, as reflected by low RMSE values ranging from 0.5147 to 0.5989 kN/m³ and MAE values below 0.41 kN/m³ across all datasets. The low MAPE (≤ 2.27%) and CVRMSE (≤ 3.23%) values highlight the model's excellent relative predictive precision and stability when normalized against the mean target value. Moreover, learning curve analysis reveals a well-balanced bias–variance trade-off, with diminishing performance gaps between training and cross-validation scores as sample size increases, suggesting efficient utilization of available data. The strong linear alignment observed in actual versus predicted plots across training, validation, and testing sets provides visual confirmation of the model's reliability and minimal overfitting. Collectively, these results position the proposed SVR framework as a state-of-the-art, data-driven approach for accurately predicting maximum dry density, offering significant potential for enhancing geotechnical design efficiency and reducing experimental dependency in compaction characterization.

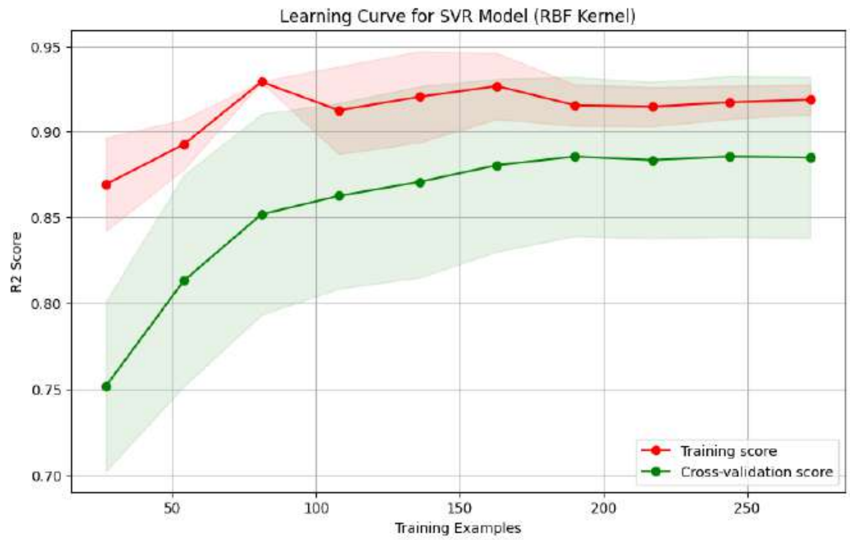


Figure 2 Learning curve of the Support Vector Regression (SVR) model with RBF kernel showing training and cross-validation R^2scores as a function of the number of training samples.

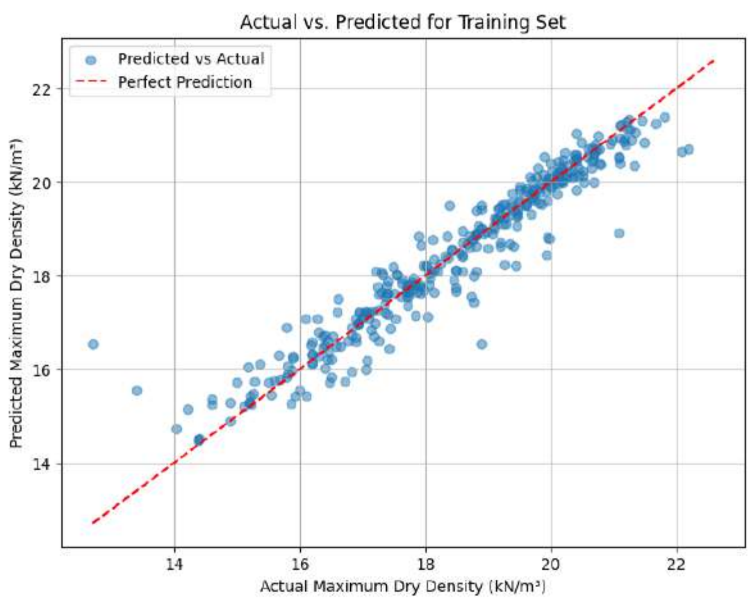


Figure 3 Comparison between actual and predicted maximum dry density (kN/m³) for the training dataset using the SVR (RBF kernel) model.

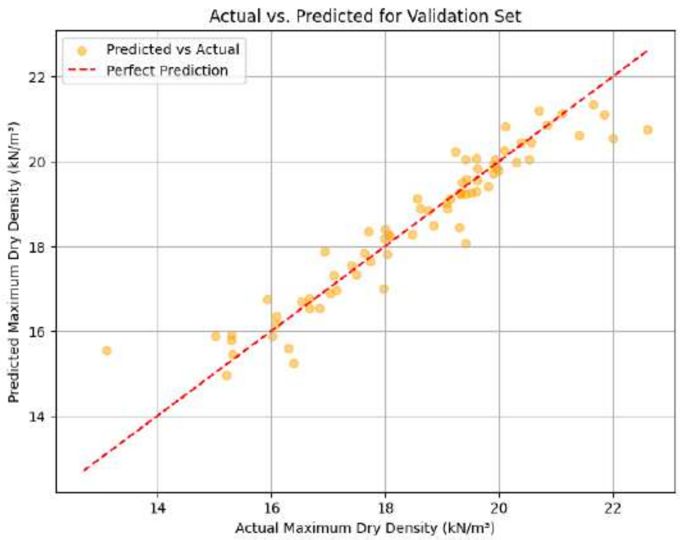


Figure 4 Comparison between actual and predicted maximum dry density (kN/m³) for the validation dataset using the SVR (RBF kernel) model.

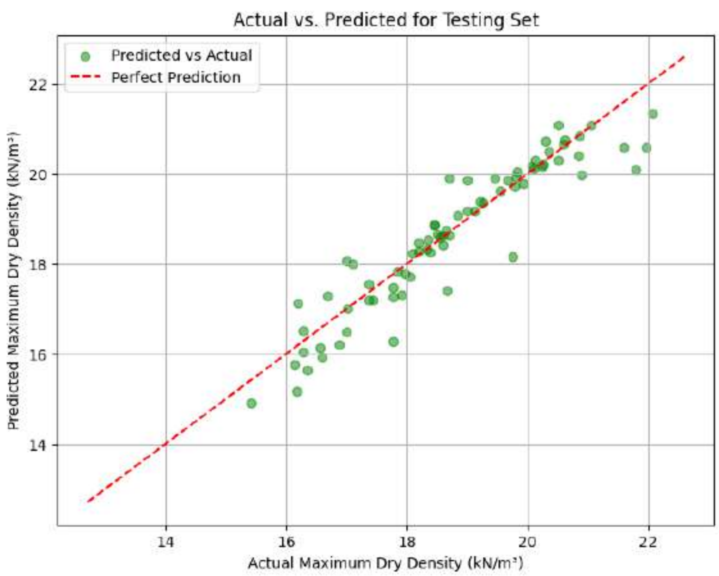


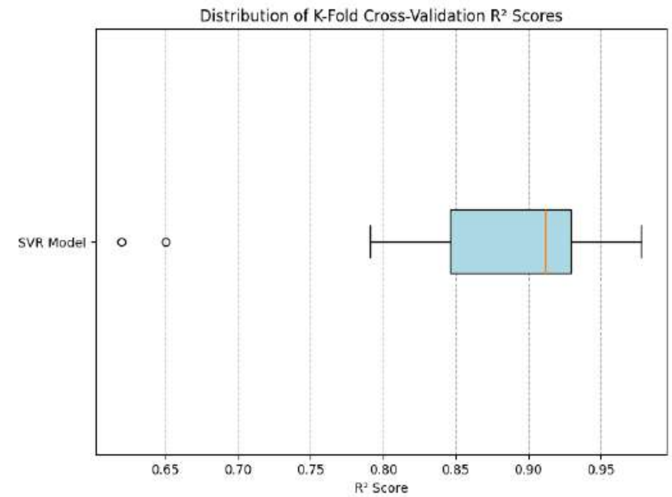
Figure 5 Comparison between actual and predicted maximum dry density (kN/m³) for the testing dataset using the SVR (RBF kernel) model.

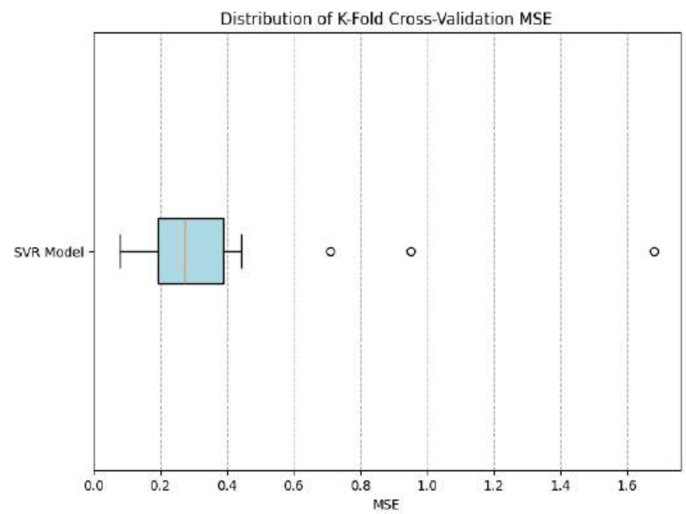
Table 2 Statistical performance evaluation of the Support Vector Regression (SVR) model with RBF kernel, showing coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), coefficient of variation of RMSE (CVRMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) for training, validation, and testing datasets.

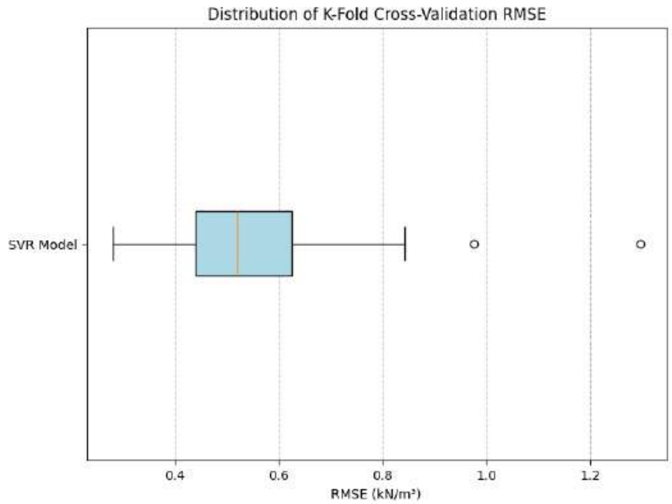
Dataset	R^2	MSE	RMSE	CVRMSE (%)	MAE	MAPE (%)
Training	0.9207	0.2650	0.5147	2.7867	0.3306	1.8524
Validation	0.9015	0.3586	0.5989	3.2303	0.4072	2.2666
Testing	0.8640	0.3410	0.5839	3.1145	0.4088	2.2034

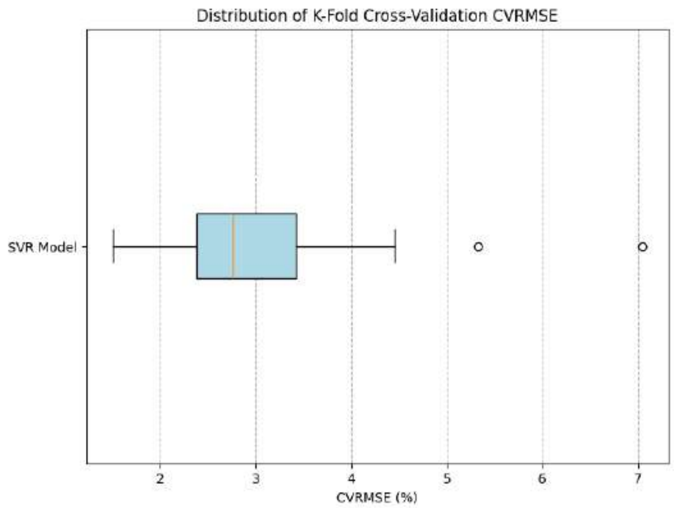
3.2 K fold results

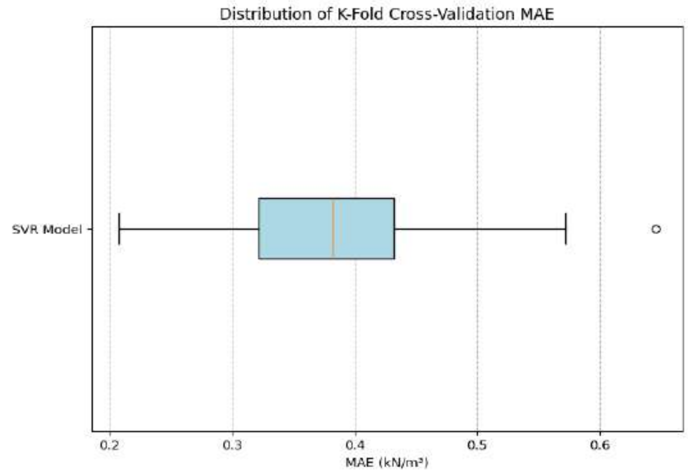
The robustness and generalization capability of the SVR model were further evaluated using 20-fold cross-validation, with the distribution of performance metrics summarized through statistical analysis and box plot visualization. The model achieved a mean R^2value of 0.8773 with a standard deviation of 0.0947, indicating strong predictive capability while also reflecting moderate variability across different data splits. This variability suggests sensitivity to certain fold-specific data characteristics, although overall explanatory power remains high. Error-based metrics further support the model's reliability. The mean MSE and RMSE values were 0.3838 and 0.6195, respectively, with relatively larger standard deviations, implying that prediction errors fluctuate across folds but remain within acceptable bounds. Importantly, the CVRMSE averaged 3.1178%, demonstrating that the prediction errors are small relative to the magnitude of the target variable, which highlights the model's stability and precision. The MAE (0.3912) and MAPE (2.2085%) values further confirm the model's high accuracy, with low percentage-based errors and moderate dispersion. The box plots visually corroborate these findings, showing compact interquartile ranges for most metrics and limited outliers. Collectively, the cross-validation results indicate that the SVR model exhibits strong generalization performance, consistent accuracy, and acceptable variability, making it suitable for reliable prediction of maximum dry density.






Distribution of K-Fold Cross-Validation RMSE
SVR Model
0.4
0.6
0.8
1.0
1.2
RMSE (kN/m³)


Distribution of K-Fold Cross-Validation CVRMSE
SVR Model
2
3
4
5
6
7
CVRMSE (%)



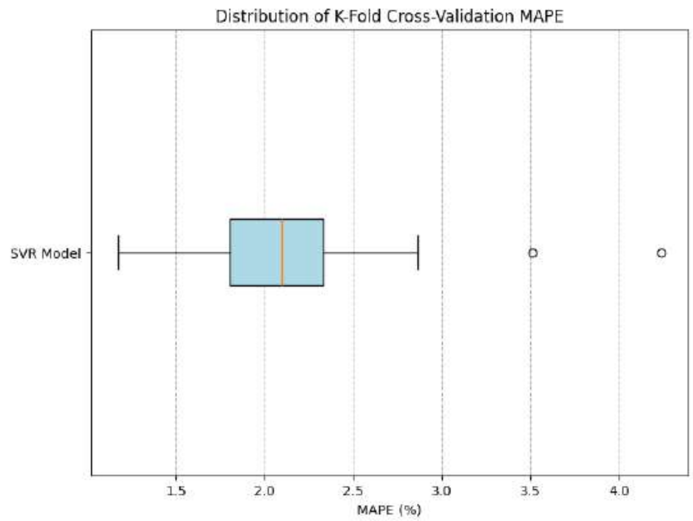


Figure 6 Box plot distributions of 20-fold cross-validation performance metrics for the SVR model, including R^2, MSE, RMSE, CVRMSE, MAE, and MAPE, illustrating model accuracy, variability, and robustness across different data splits.

3.3 Error Analysis

The prediction reliability of the SVR model on the testing dataset was further assessed through an analysis of percentage errors and their relationship with actual maximum dry density values. The computed percentage errors exhibited both positive and negative values, indicating the presence of minor over- and under-predictions, with initial values (e.g., 4.63%, 6.75%, −4.47%, 0.33%, and −0.90%) confirming that deviations remain within a relatively narrow range. The histogram of percentage errors demonstrates a near-symmetric, bell-shaped distribution centered close to zero, with the majority of errors confined within ±5%. This distributional behavior suggests the absence of systematic bias and indicates that prediction errors are random rather than directional. The limited occurrence of extreme values further highlights the robustness and accuracy of the model on unseen data. Additionally, the scatter plot of actual maximum dry density versus percentage error reveals no discernible trend or clustering across the range of target values. Errors are evenly dispersed around the zero-error line, implying that the model maintains consistent predictive performance across low, medium, and high density ranges. Collectively, these findings confirm that the SVR model generalizes well to the testing dataset, producing accurate and unbiased predictions with minimal relative error, thereby reinforcing its suitability for practical applications in maximum dry density estimation.

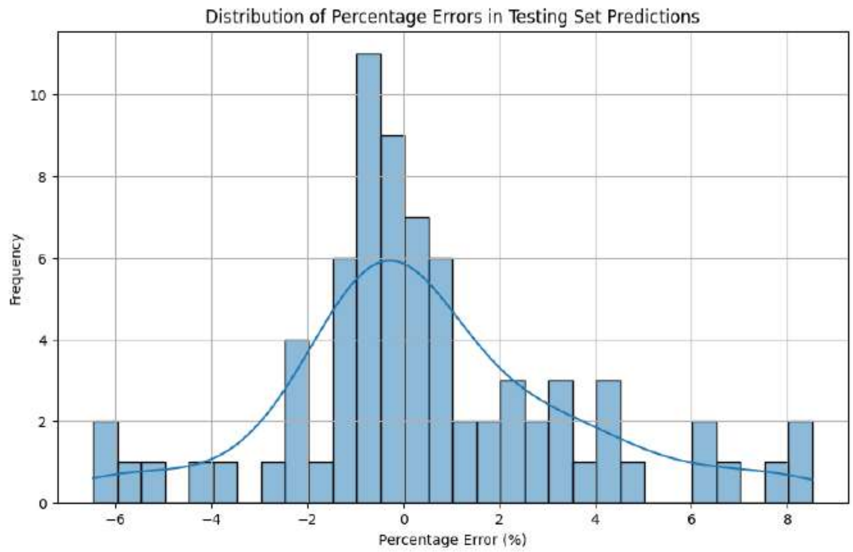


Figure 7 Histogram with kernel density estimation of percentage prediction errors for the testing dataset, illustrating the distribution, symmetry, and concentration of errors around zero for the SVR model.

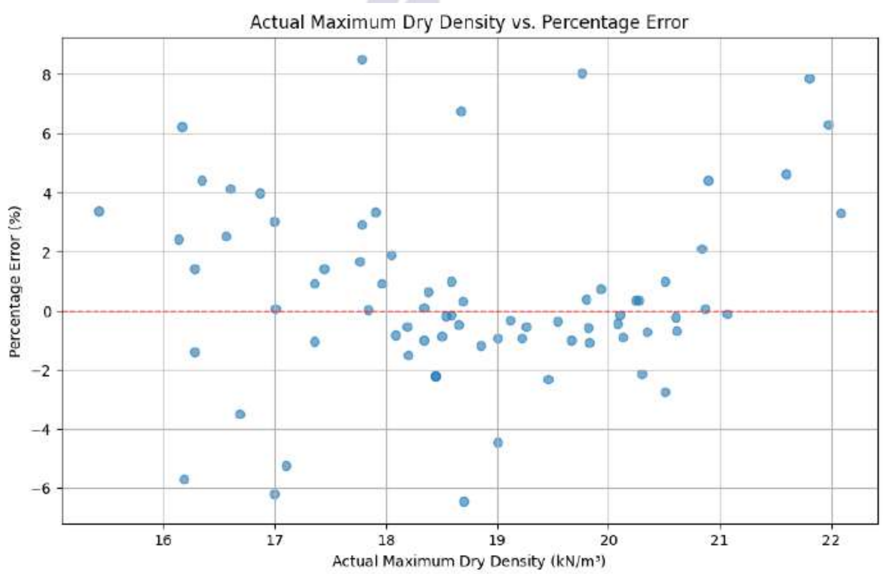


Figure 8 Scatter plot of actual maximum dry density versus percentage prediction error for the testing dataset, showing the absence of systematic bias and the random distribution of errors across the range of target values.

3.4

Statistical significance analysis

The statistical significance of the SVR model predictions was evaluated through residual and error distribution analyses to assess bias, consistency, and distributional assumptions. The percentage error distribution is centered close to zero and exhibits an approximately symmetric, bell-shaped form, with most errors confined within ±5%. This behavior indicates balanced over- and under-predictions and confirms the overall accuracy of the model. Furthermore, the scatter plot of actual maximum dry density versus percentage error shows a random dispersion of errors around the zero-error line, suggesting the absence of systematic bias across the full range of target values.

Residual analysis provides additional insight into model reliability. The mean residual value of 0.1041 is very close to zero, indicating negligible systematic over- or under-estimation, while the standard deviation of 0.5746 reflects a relatively small typical prediction error in physical units. The histogram and kernel density estimate of residuals further demonstrate that errors are symmetrically distributed around zero. The Q–Q plot shows that most residuals closely follow the theoretical normal line, suggesting approximate normality. However, the Shapiro–Wilk test yields a statistic of 0.9444 with a p-value of 0.0027, leading to rejection of strict normality at the 5% significance level. Despite this statistical deviation, the visual diagnostics indicate near-normal behavior, which is generally acceptable for nonlinear machine learning models. Overall, the residual and error analyses confirm that the SVR model is unbiased, stable, and statistically reliable.

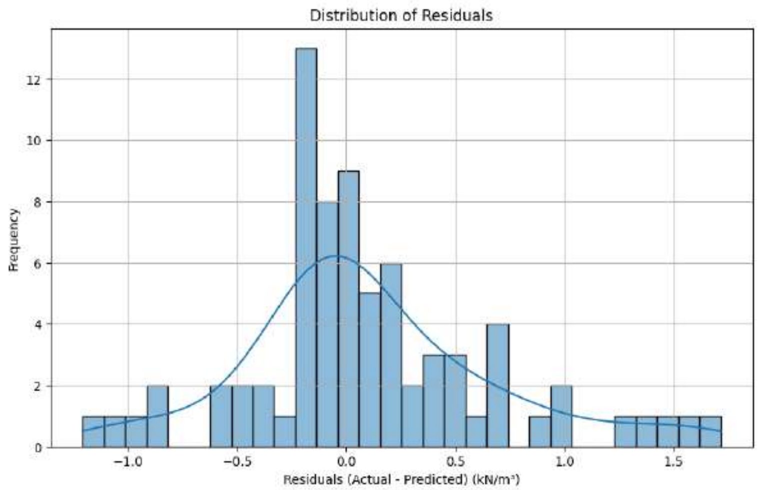


Figure 9 Histogram and kernel density estimation of residuals (actual minus predicted maximum dry density), illustrating the central tendency, spread, and symmetry of prediction errors for the SVR model.

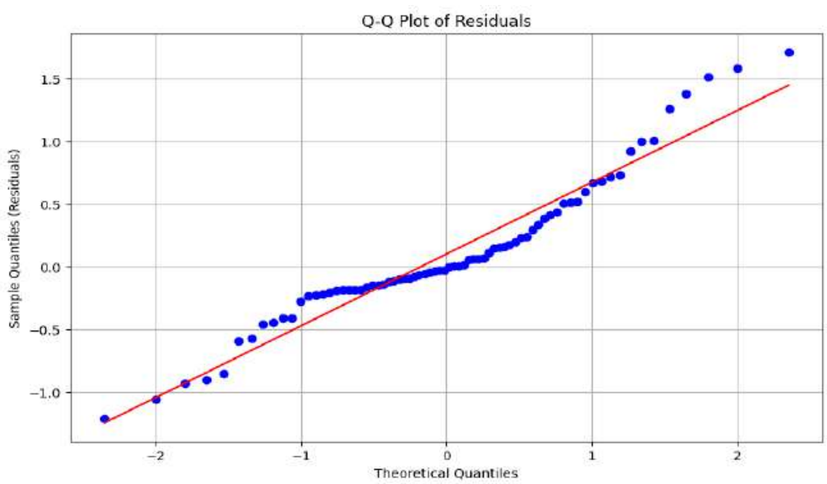


Figure 10 Q–Q plot of residuals for the SVR model comparing sample quantiles with theoretical normal quantiles, used to assess the normality of prediction errors.

3.5 Shapely additive analysis

Model interpretability was investigated using SHAP (SHapley Additive exPlanations) to quantify both the global and local influence of input variables on the SVR model predictions of maximum dry density. The SHAP feature importance analysis, based on mean absolute SHAP values, clearly indicates that **Optimum Moisture Content** (%) is the most influential predictor, substantially outweighing the contributions of other variables. This highlights the dominant role of moisture conditions in controlling soil compaction behavior. **Fines Content** (%) and **Gravel Content** (%) emerge as the next most influential features, reflecting the importance of soil gradation and particle size distribution in determining dry density. Other parameters, including **Plastic Limit** (%), **Sand Content** (%), and **Liquid Limit** (%), exhibit comparatively lower but non-negligible contributions to the model output.

The SHAP summary plot provides further insight into the directionality and variability of feature effects at the instance level. Higher values of Optimum Moisture Content (%) are generally associated with positive SHAP values, indicating an increase in predicted maximum dry density, whereas lower values tend to reduce predictions. Similarly, Gravel Content (%) and Fines Content (%) display bidirectional effects, suggesting nonlinear interactions where both high and low values can either enhance or diminish the predicted response depending on the combined feature context. Overall, the SHAP analysis confirms that the SVR model captures physically meaningful relationships and offers transparent, interpretable insights into the relative importance and directional influence of key geotechnical parameters.

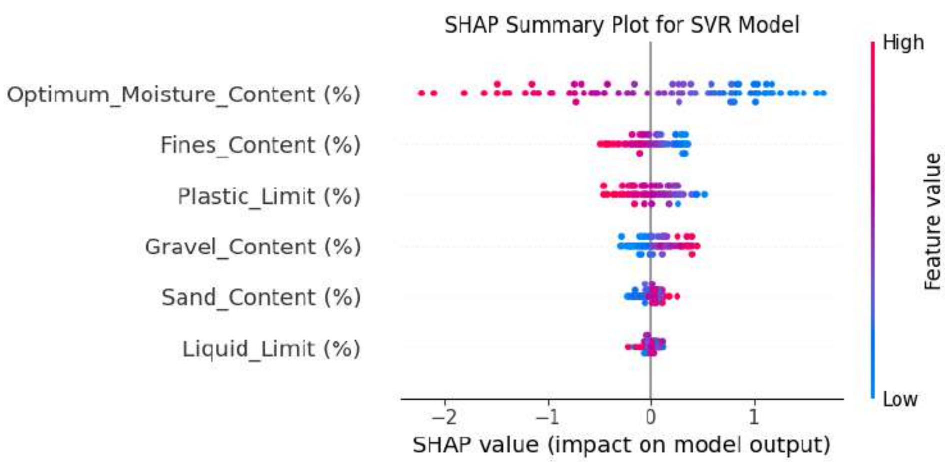


Figure 11 SHAP summary plot illustrating the distribution and direction of feature contributions to the SVR model predictions of maximum dry density, with color indicating feature magnitude (low to high).

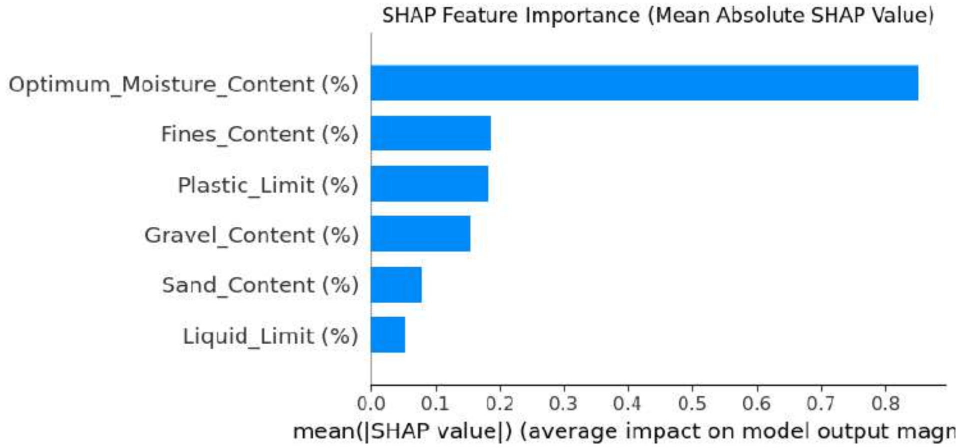


Figure 12 SHAP feature importance plot based on mean absolute SHAP values, ranking input variables according to their average influence on the SVR model output.

3.6 Partial dependence analysis

Partial Dependence Plots (PDPs) were employed to further elucidate the marginal effects of individual input variables on the SVR model's prediction of maximum dry density while averaging out the influence of other features. The PDPs reveal clear nonlinear relationships between soil properties and the predicted response, highlighting the model's ability to capture complex soil behavior. An increase in **Gravel Content** (%) is generally associated with a moderate rise in maximum dry density up to an optimal range, beyond which the effect stabilizes, reflecting improved particle interlocking at moderate gravel proportions. **Sand Content** (%) shows a gradual positive trend, suggesting enhanced packing efficiency with increasing sand fraction. In contrast, **Fines Content** (%) exhibits a decreasing trend, indicating that excessive fines may hinder compaction by increasing surface area and water demand.

The effects of consistency limits are also evident. **Liquid Limit** (%) demonstrates a mild negative influence, while **Plastic Limit** (%) shows a more pronounced decreasing trend, implying reduced achievable dry density for soils with higher plasticity. Notably, **Optimum Moisture Content** (%) displays the strongest and most nonlinear influence, with predicted maximum dry density decreasing sharply at higher moisture levels, consistent with soil mechanics principles where excess water reduces effective stress and compaction efficiency. Overall, the PDP analysis confirms that the SVR model captures physically meaningful, nonlinear relationships between key geotechnical parameters and maximum dry density, thereby enhancing confidence in both the predictive capability and interpretability of the model.

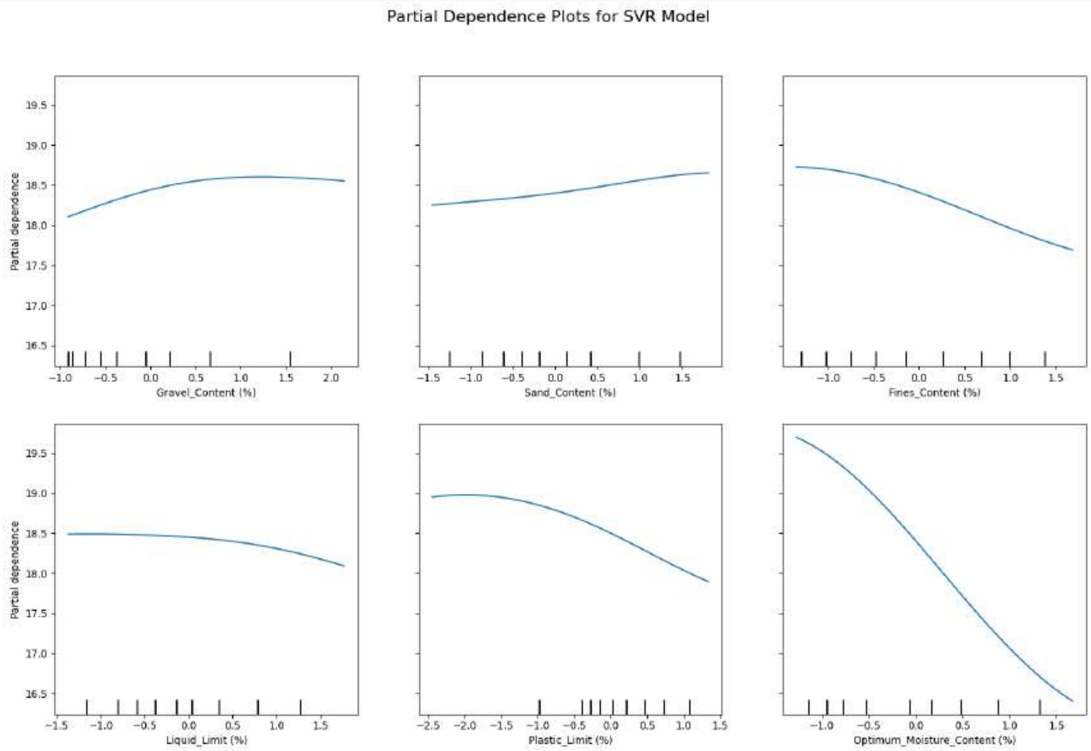


Figure 13 Partial dependence plots illustrating the marginal effects of input variables (gravel content, sand content, fines content, liquid limit, plastic limit, and optimum moisture content) on the SVR model predictions of maximum dry density.

4 Discussion

The SVR model with RBF kernel shows performance ($R^2 \approx 0.86–0.92$; MAPE ≈ 2%) that is highly competitive with, and in some respects comparable to, state-of-the-art data-driven approaches for MDD and related geotechnical targets (Zhao and Guan, 2024; Zhao et al., 2024; Almuaythir, Zaini and Lodhi, 2025; Ali et al., 2024; Khatti and Grover, 2023; Tao et al., 2024). Similar R^2 levels (≈0.90–0.97) and sub-5% relative errors are widely reported when SVR is carefully tuned and applied to nonlinear physical problems such as soil compaction, slope safety factors, seismic demand, permeability, and concrete strength (Zhao and Guan, 2024; Zhao et al., 2024; Almuaythir, Zaini and Lodhi, 2025; Khatti and Grover, 2023; Tao et al., 2024; Lei et al., 2024; Hussen et al., 2024; Roy and Chakraborty, 2020; Wu and Zhou, 2022). The modest gap between training and testing R^2 in your model is consistent with well-regularized SVR behaviour, where capacity control via C and kernel parameters limits overfitting while still capturing complex relationships (Zhang and O'Donnell, 2020; Lei et al., 2024; Roy and Chakraborty, 2020; Rahmayanthi, M and Wahyuningsih, 2025).

Your learning-curve evidence of converging training/validation scores aligns with recommended SVR-based reliability workflows that iteratively enrich data near critical response regions to balance bias and variance (Roy and Chakraborty, 2020). Similar convergence patterns have been documented in SVR metamodels for structural reliability and seismic response, where small training–test performance differences indicate good generalization under Monte Carlo or k-fold regimes (Tao et al., 2024; Roy and Chakraborty, 2020). The 20-fold CV statistics (mean $R^2 \approx 0.88$ with moderate spread) are typical when data contain heterogeneous soil types or broad state spaces; comparable fold-to-fold variability has been observed for SVR in permeability prediction, insurance losses, and agricultural traits (Zhou, Yan and Zhang, 2024;

Bekkaye et al., 2025; Fałdziński, Fiszeder and Orzeszko, 2020; Hussen et al., 2024).

Error distributions centered near zero with most percentage errors within ±5% agree with findings from SVR studies on soil compaction, slope stability, commodity prices, and hydrometeorological series, where MAE and MAPE remain low and residuals show no strong trend with target magnitude (Almuaythir, Zaini and Lodhi, 2025; Khatti and Grover, 2023; Azis et al., 2023; Lei et al., 2024; Rahmayanthi, M and Wahyuningsih, 2025). Mild departures from strict normality in residuals are common in nonlinear SVR applications; visual near-normality despite Shapiro–Wilk rejection has been reported in concrete strength and porous-media transport modelling, and is generally considered acceptable for predictive—not strictly inferential—use (Tasqeeruddin, Sultana and Alsayari, 2025; Tao et al., 2024; Roy and Chakraborty, 2020; Wu and Zhou, 2022).

The SHAP analysis highlighting optimum moisture content as the dominant predictor is geotechnically consistent and parallels SHAP-based SVR interpretations in MDD, seismic MDR, and concrete strength, where a small subset of physically meaningful variables explains most variance (Zhao and Guan, 2024; Zhao et al., 2024; Ali et al., 2024; Tao et al., 2024; Wu and Zhou, 2022). The elevated importance of fines and gravel contents matches other compaction and permeability models that rank gradation, fines, and porosity as primary drivers of densification and flow (Almuaythir, Zaini and Lodhi, 2025; Ali et al., 2024; Khatti and Grover, 2023; Hussen et al., 2024). Similar SHAP workflows in soybean branching and sustainable concrete also show that SVR, coupled with SHAP, can yield both high predictive accuracy and transparent feature attribution, strengthening confidence in model-based design decisions (Zhou, Yan and Zhang, 2024; Tao et al., 2024; Wu and Zhou, 2022).

Collectively, your SVR–RBF framework, supported by cross-validation, error diagnostics, and SHAP interpretability, is in line with contemporary best practice where SVR frequently rivals or narrowly trails boosted tree ensembles and optimized hybrids, while offering strong robustness and clear physical insights for geotechnical prediction tasks (Zhao and Guan, 2024; Zhao et al., 2024; Zhou, Yan and Zhang, 2024; Almuaythir, Zaini and Lodhi, 2025; Ali et al., 2024; Khatti and Grover, 2023; Tao et al., 2024; Lei et al., 2024; Hussen et al., 2024; Roy and Chakraborty, 2020; Wu and Zhou, 2022).

5 Conclusion and Recommendations

This study demonstrates the effectiveness of an explainable Support Vector Regression (SVR) framework for predicting maximum dry density (MDD) from basic soil gradation and index properties. The SVR model with an RBF kernel achieved consistently high predictive performance across training, validation, testing, and cross-validation stages, with R^2 values exceeding 0.86 and mean absolute percentage errors close to 2%. These results confirm that nonlinear machine learning models can reliably capture the complex interactions governing soil compaction behavior using readily available laboratory inputs, thereby reducing reliance on repeated Proctor testing.

Comprehensive error diagnostics showed that prediction errors are small, symmetrically distributed, and largely confined within ±5%, with no evidence of systematic bias across the range of MDD values. Residual analysis further confirmed model stability and robustness, while mild deviations from strict normality were deemed acceptable for predictive applications. Importantly, the use of SHAP and partial dependence analysis provided transparent insights into model behavior, bridging the gap between data-driven prediction and soil mechanics understanding. Optimum moisture content emerged as the most influential parameter, followed by fines and gravel content, in agreement with established compaction theory. The nonlinear and interaction effects revealed by these interpretability tools reinforce confidence that the model learns physically meaningful relationships rather than spurious correlations.

From a practical perspective, the proposed SVR framework offers a valuable surrogate modeling tool for preliminary design, material screening, and quality control in earthworks and pavement engineering. It enables rapid estimation of MDD using routine soil characterization data, supporting more efficient design iterations and informed decision-making. However, the study also highlights areas for future improvement. Expanding the dataset to include additional variables such as compaction energy, soil mineralogy, stabilizer type and dosage, and field compaction conditions could further enhance model generality. Comparative benchmarking against advanced ensemble and deep learning models may also provide additional insights into model selection trade-offs.

In conclusion, the integration of SVR with explainable artificial intelligence techniques represents a state-of-the-art, yet practically interpretable, approach for MDD prediction. Future research should focus on broader datasets, hybrid modeling strategies, and field validation to further strengthen the applicability of machine learning-based compaction models in geotechnical engineering practice.

6 REFERENCES

Zhao, L., & Guan, G., 2024. Maximum dry density estimation of stabilized soil via machine learning techniques in individual and hybrid approaches. *Journal of Ambient Intelligence and Humanized Computing*, 15, pp. 3831 - 3846. https://doi.org/10.1007/s12652-024-04860-5

Zhao, Q., Liu, K., Xiong, C., Deng, X., & Yang, S., 2024. Estimating the Maximum Dry Density of Soil via Least Square Support Vector Regression Individual and Hybrid Forms. *Indian Geotechnical Journal*, 55, pp. 866 - 878. https://doi.org/10.1007/s40098-024-00952-3

Zhou, W., Yan, Z., & Zhang, L., 2024. A comparative study of 11 non-linear regression models highlighting autoencoder, DBN, and SVR, enhanced by SHAP importance analysis in soybean branching prediction. *Scientific Reports*, 14. https://doi.org/10.1038/s41598-024-55243-x

Almuaythir, S., Zaini, M., & Lodhi, R., 2025. Predicting soil compaction parameters in expansive soils using advanced machine learning models: a comparative study. *Scientific Reports*, 15. https://doi.org/10.1038/s41598-025-09279-2

Tasqeeruddin, S., Sultana, S., & Alsayari, A., 2025. Development of hybrid robust model based on computational modeling and machine learning for analysis of drug sorption onto porous adsorbents. *Scientific Reports*, 15. https://doi.org/10.1038/s41598-025-93596-z

Ali, H., Omer, B., Mohammed, A., & Faraj, R., 2024. Predicting the maximum dry density and optimum moisture content from soil index properties using efficient soft computing techniques. *Neural Computing and Applications*, 36, pp. 11339 - 11369. https://doi.org/10.1007/s00521-024-09734-7

Khatti, J., & Grover, K., 2023. Prediction of compaction parameters for fine-grained soil: Critical comparison of the deep learning and standalone models. *Journal of Rock Mechanics and Geotechnical Engineering*. https://doi.org/10.1016/j.jrmge.2022.12.034

Azis, H., Purnawansyah, P., Nirwana, N., & Dwiyanto, F., 2023. The Support Vector Regression Method Performance Analysis in Predicting National Staple Commodity Prices. *ILKOM Jurnal Ilmiah*. https://doi.org/10.33096/ilkom.v15i2.1686.390-397

Bekkaye, C., Oukhouya, H., Zari, T., Guerbaz, R., & Bouanani, H., 2025. Advanced Strategies for Predicting and Managing Auto Insurance Claims using Machine Learning Models. *Statistics, Optimization & Information Computing*. https://doi.org/10.19139/soic-2310-5070-2655

Zhang, F., & O'Donnell, L., 2020. Support Vector Regression. *Wiley StatsRef: Statistics Reference Online*. https://doi.org/10.1016/b978-0-12-815739-8.00007-9

Fałdziński, M., Fiszeder, P., & Orzeszko, W., 2020. Forecasting Volatility of Energy Commodities: Comparison of GARCH Models with Support Vector Regression. *Energies*. https://doi.org/10.3390/en14010006

Tao, D., Fang, S., Liu, H., Lu, J., Wang, J., & , Q., 2024. Support vector regression model for the prediction of buildings' maximum seismic response based on real monitoring data. *Scientific Reports*, 14. https://doi.org/10.1038/s41598-024-81705-3

Lei, D., Zhang, Y., Lu, Z., Lin, H., & Jiang, Z., 2024. Predicting Factor of Safety of Slope Using an Improved Support Vector Machine Regression Model. *Mathematics*. https://doi.org/10.3390/math12203254

Hussen, A., Munshi, T., Jahan, L., & Hashan, M., 2024. Advanced machine learning approaches for predicting permeability in reservoir pay zones based on core analyses. *Heliyon*, 10. https://doi.org/10.1016/j.heliyon.2024.e32666

Roy, A., & Chakraborty, S., 2020. Support vector regression based metamodel by sequential adaptive sampling for reliability analysis of structures. *Reliab. Eng. Syst. Saf.*, 200, pp. 106948. https://doi.org/10.1016/j.ress.2020.106948

Rahmayanthi, A., M, F., & Wahyuningsih, S., 2025. RAINFALL PREDICTION IN BALIKPAPAN CITY USING SUPPORT VECTOR REGRESSION. *Parameter: Journal of Statistics*. https://doi.org/10.22487/27765660.2025.v5.i1.17640

Wu, Y., & Zhou, Y., 2022. Hybrid machine learning model and Shapley additive explanations for compressive strength of sustainable concrete. *Construction and Building Materials*. https://doi.org/10.1016/j.conbuildmat.2022.127298

Zhao, L., & Guan, G., 2024. Maximum dry density estimation of stabilized soil via machine learning techniques in individual and hybrid approaches. *Journal of Ambient Intelligence and Humanized Computing*, 15, pp. 3831 - 3846. https://doi.org/10.1007/s12652-024-04860-5

Omar, M., Shanableh, A., Mughieda, O., Arab, M., Zeiada, W., & Al-Ruzouq, R., 2018. Advanced mathematical models and their comparison to predict compaction properties of fine-grained soils from various physical properties. *Soils and Foundations*. https://doi.org/10.1016/j.sandf.2018.08.004

Ngo, T., Nguyen, L., & Tran, V., 2022. Novel hybrid machine learning models including support vector machine with meta-heuristic algorithms in predicting unconfined compressive strength of organic soils stabilised with cement and lime. *International Journal of Pavement Engineering*, 24. https://doi.org/10.1080/10298436.2022.2136374

Khatti, J., & Grover, K., 2023. Prediction of compaction parameters for fine-grained soil: Critical comparison of the deep learning and standalone models. *Journal of Rock Mechanics and Geotechnical Engineering*. https://doi.org/10.1016/j.jrmge.2022.12.034

Zhang, F., & O'Donnell, L., 2020. Support Vector Regression. *Wiley StatsRef: Statistics Reference Online*. https://doi.org/10.1016/b978-0-12-815739-8.00007-9

Gunaydin, O., Özbeyaz, A., & Söylemez, M., 2018. Regression Analysis of Soil Compaction Parameters Using Support Vector Method. *Celal Bayar Üniversitesi Fen Bilimleri Dergisi*. https://doi.org/10.18466/cbayarfbe.449644

Hassan, M., & Beshr, E., 2024. Predicting soil cone index and assessing suitability for wind and solar farm development in using machine learning techniques. *Scientific Reports*, 14. https://doi.org/10.1038/s41598-024-52702-3

Alnmr, A., Hosamo, H., Lyu, C., Ray, R., & Alzawi, M., 2024. Novel Insights in Soil Mechanics: Integrating Experimental Investigation with Machine Learning for Unconfined Compression Parameter Prediction of Expansive Soil. *Applied Sciences*. https://doi.org/10.3390/app14114819

Mustafa, Y., Wudil, Y., Zami, M., & Al-Osta, M., 2025. Machine Learning Approach for Assessment of Compressive Strength of Soil for Use as Construction Materials. *Eng*. https://doi.org/10.3390/eng6050084

Almuaythir, S., Zaini, M., & Lodhi, R., 2025. Predicting soil compaction parameters in expansive soils using advanced machine learning models: a comparative study. *Scientific Reports*, 15. https://doi.org/10.1038/s41598-025-09279-2

Boadu, F., 2020. A support vector regression approach to predict geotechnical properties of soils from electrical spectra based on Jonscher parameterization. *GEOPHYSICS*. https://doi.org/10.1190/geo2019-0256.1

Ballabio, C., 2009. Spatial prediction of soil properties in temperate mountain regions using support vector regression. *Geoderma*, 151, pp. 338-350. https://doi.org/10.1016/j.geoderma.2009.04.022

Roy, A., & Chakraborty, S., 2023. Support vector machine in structural reliability analysis: A review. *Reliab. Eng. Syst. Saf.*, 233, pp. 109126. https://doi.org/10.1016/j.ress.2023.109126

Wu, Y., & Zhou, Y., 2022. Hybrid machine learning model and Shapley additive explanations for compressive strength of sustainable concrete. *Construction and Building Materials*. https://doi.org/10.1016/j.conbuildmat.2022.127298

Zhao, Q., Liu, K., Xiong, C., Deng, X., & Yang, S., 2024. Estimating the Maximum Dry Density of Soil via Least Square Support Vector Regression Individual and Hybrid Forms. *Indian Geotechnical Journal*, 55, pp. 866 - 878. https://doi.org/10.1007/s40098-024-00952-3

Zhang, P., Yin, Z., & Jin, Y., 2021. Machine Learning-Based Modelling of Soil Properties for Geotechnical Design: Review, Tool Development and Comparison. *Archives of Computational Methods in Engineering*, 29, pp. 1229-1245. https://doi.org/10.1007/s11831-021-09615-5

Onyelowe, K., Moghal, A., Ebid, A., Rehman, A., Hanandeh, S., & Priyan, V., 2024. Estimating the strength of soil stabilized with cement and lime at optimal compaction using ensemble-based multiple machine learning. *Scientific Reports*, 14. https://doi.org/10.1038/s41598-024-66295-4

Wan, X., 2023. Predicting the Settlement of Shallow Foundation Using Metaheuristic SVR Approaches. *Geotechnical and Geological Engineering*, 41, pp. 4795 - 4805. https://doi.org/10.1007/s10706-023-02547-w

Gali, G. *et al.* (2023) "Experimental Study on the Effects of Freeze-Thaw Progressions and Performance of Soil with Non-Toxic Bio-Enzyme," *Journal of ICT, Design, Engineering and Technological Science*, 7(1), pp. 12–20. Available at: https://doi.org/10.33150/JITDETS-7.1.3.

Khan, W. *et al.* (2025) "Transforming Expansive Clays FROM Waste TO Value: Dual Ash Stabilization Strategy FOR Augmenting Black Cotton Subgrade Soil Performance," 11, pp. 5065–5073. Available at: https://doi.org/10.64252/cz07k049.

Nawaz, S. *et al.* (2025) "Flexural Performance Evaluation Of Glass Fiber Reinforced Polymer Bars (GFRP) In Doubly Reinforced Beams: An Experimental Approach."

Shah, Q. *et al.* (2023) "To Stabilize Shear Strength Properties of an Unwanted Subgrade Soil Utilizing Rock Dust," *Journal of ICT, Design, Engineering and Technological Science*, 7(2), pp. 1–6. Available at: https://doi.org/10.33150/JITDETS-7.2.1.