Review

Not peer-reviewed version

Application of Machine Learning Techniques for Prediction of Soil Water Characteristics Curve: A State of the Art Review

Mohak Desai and Kaustav Chatterjee *

Posted Date: 10 February 2026

doi: 10.20944/preprints202602.0792.v1

Keywords: soil suction; machine learning; SWCC; SWRC; unsaturated soil mechanics

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 February 2026 doi:10.20944/preprints202602.0792.v1

Review

Application of Machine Learning Techniques for Prediction of Soil Water Characteristics Curve: A State of the Art Review

Mohak Desai [1] and Kaustav Chatterjee [2,*]

[1] Senior Data Analyst, NIQ, Vadodara
[2] PhD, School of Civil and Environmental Engineering, Oklahoma State University, Stillwater, Oklahoma: 74075, United States of America
* Correspondence: kaustav.chatterjee10@okstate.edu

Abstract

Soil suction is a crucial factor affecting the hydraulic and mechanical property of unsaturated soils, playing an important role in geotechnical, geoenvironmental, and hydrological engineering applications such as slope stability, foundation design and irrigation planning. Conventionally, measuring and modeling soil suction and its associated curves like Soil Water Characteristic Curve (SWCC) and Soil Water Retention Curve (SWRC) require extensive, time-consuming tests in the laboratory. Recent progress in Machine Learning (ML) offers powerful as well as data-driven and reliable alternatives ways that can enhance the efficiency and accuracy of suction-related predictions across a wide range of soil conditions. This study aims to cover the current state of the art research on the integration of ML techniques into the prediction and analysis of soil suction behavior. Studies utilized various algorithms including Random Forest (RF), Extreme Gradient Boosting (XGBoost), Artificial Neural Networks (ANNs), Support Vector Machine (SVM), Multi-Expression Programming (MEP), K-Nearest Neighbors (KNN), and AdaBoost (AB) to predict soil suction. These models demonstrated high predictive performance ($R^2 > 0.90$ in majority cases) based on soil parameters which can be easily evaluated like soil texture, bulk density, climate parameters, and remotely sensed data. Overall, this study covers the understanding of the current research gap related to SWCC and SWRC using different data driven and ML techniques.

Keywords: soil suction; machine learning; SWCC; SWRC; unsaturated soil mechanics

---

Introduction

Soil suction is a very important factor in geotechnical engineering and unsaturated soil mechanics, affecting slope stability, behavior of expansive soils and soil load bearing capacity. Soil suction is the capacity of soil to hold the water due to capillary and adsorptive forces. Soil suction can be subdivided into two types, matric suction and osmotic suction. Matric suction is created by capillary forces due to soil's pore structures, depending on the distribution and size of the pores. Osmotic suction arises due to the availability of dissolved salts in soil water, affecting the movement of water in soil.

Established techniques to evaluate soil suction include tensiometers, pressure plate apparatus, filter paper method, and psychrometers. Tensiometer is one of the commonly used tools which can measure matric suction directly, but it can be used for measuring low suction (below 100 kPa) range. Pressure plate apparatus can be employed for measuring higher suction values. In this method, soil samples are subjected to controlled air pressure to calculate water retention characteristics. Filter paper technique is another conventional technique where equilibrium moisture content of filter paper in contact with soil is used to calculate suction. Dew point hygrometers and psychrometers can calculate total suction by evaluating the relative humidity of the soil atmosphere.

These traditional methods to evaluate soil suction are widely used but also have some limitations such as time-consuming experiments, manual data acquisition, and limited accuracy due to sensitivity of the instruments, environment variability and human errors in some cases. ML can alleviate these problems by analyzing the large datasets obtained from different sources like remote sensing, sensor network and weather data. Different ML models were employed by researchers, practitioners and geotechnical engineers to measure soil suction and to gain insight into the complicated relationship between soil properties and suction behavior.

This study aims to introduce a comprehensive summary of the application of state-of-the art data driven techniques for measuring soil suction and related properties. The research commenced with selection of keywords (data-driven techniques, soil suction, unsaturated soil mechanics, ML), followed by literature acquisition from different databases such as Science Direct, Google Scholar, and some reputed journals. After literature acquisition, the contents of different literatures were summarized in the body of the paper and in a table. Moreover, this chapter provides a brief insight into some of the popular ML models. Figure 1 presents an overview of the methodology used in this research.

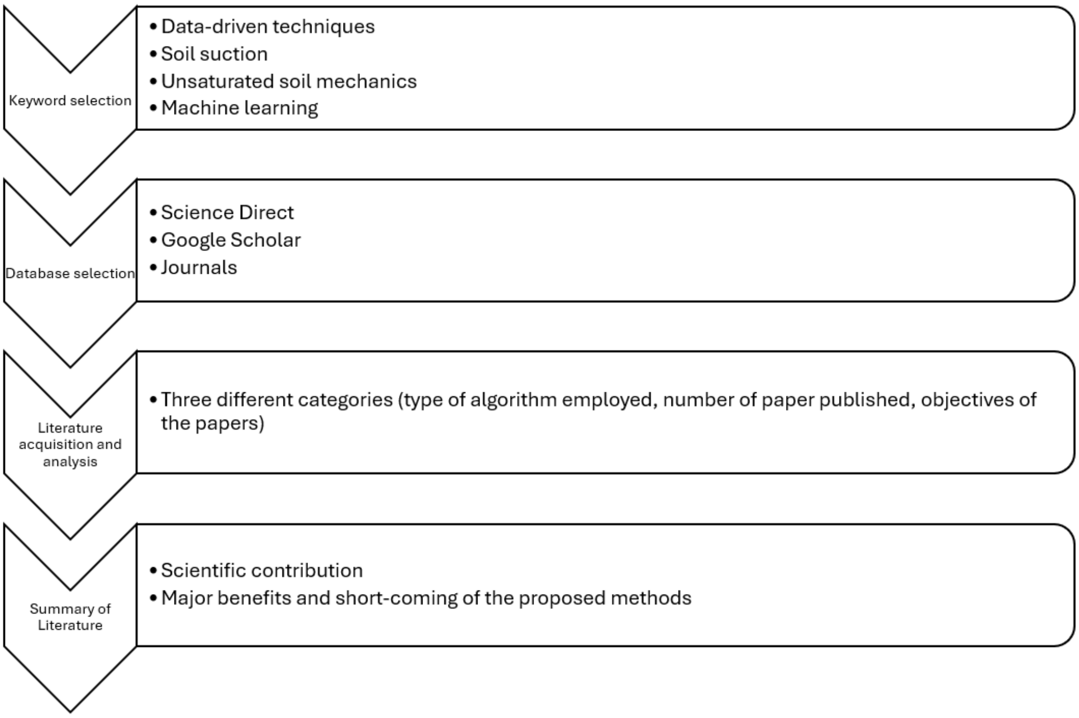
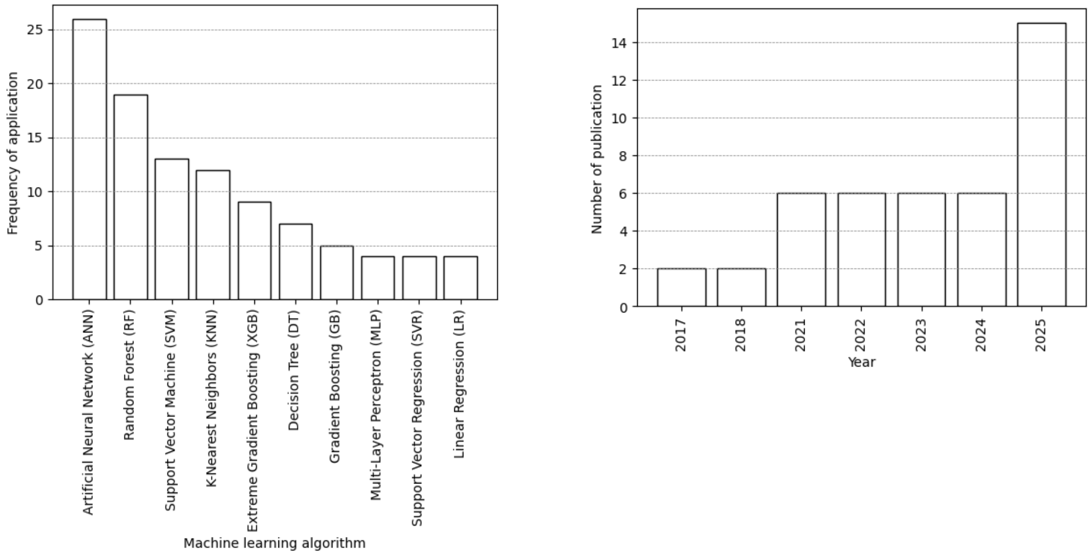


Figure 1. Methodology of research.

The literature collected in this study was analyzed in the three different categories namely: (a) ML algorithm employed for predicting parameters related to unsaturated properties of soil, (b) number of papers published in different years, and (c) objectives of the papers. Analyzing the ML algorithms employed by different researchers provided the information about the most efficient machine learning algorithms for predicting unsaturated properties of soil. Figure 2a shows the bar plot of commonly used algorithms and their frequency. ANN was the most popular algorithm followed by random forest algorithm. The reason for higher application of ANN by researcher due to was its ability to capture the non-linear relationship in the data. Random Forest was frequently used by researchers because it is an ensemble-based algorithm and can perform non-linear modelling without over-fitting. Figure 2b shows the number of papers published in different years. Analyzing the number of papers published during different years revealed that ML has become a popular tool among engineers and researchers in the last five years, with 15, 6, 6, 6, and 6 papers published in 2025, 2024, 2023, 2022 and 2021, respectively. Moreover, analyzing the paper with respect to research

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 February 2026 doi:10.20944/preprints202602.0792.v1

objective reveals that majority of the research covers SWCC which is considered as the principal factor affecting relationship between soil suction and water content followed by SWRC and UWC.



(a) (b)

Figure 2. (a) Frequency of application of ML algorithms, (b) numbers of papers published each year.

State-of-the-Art Machine Learning Algorithms

Decision Tree (DT)

DT is the boolean based data classification and regression technique, performs automatic decisions by starting at the root (top level) and moving down through a series of boolean based questions until it finally reaches a final answer at the leaf. This tree-based model automatically takes the decision of which and how many questions to be asked and where to split the data to reach the final leaf based on “YES/NO” condition. DT functions like a flowchart to make any decisionstep by step where internal nodes serve as attribute tests, branches serve as attribute values and leaf nodes serve as decisions. Figure 3 illustrates the schematic representation of DT algorithm.

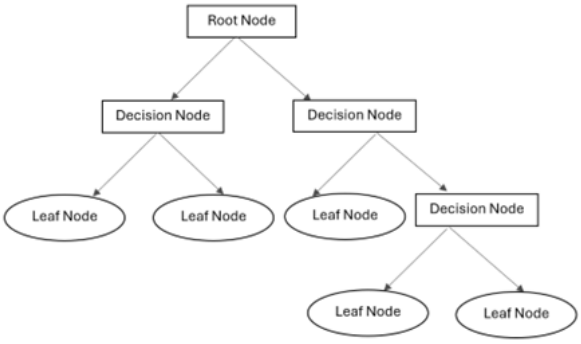


Figure 3. Schematic representation of DT.

Support Vector Machine (SVM)

SVM is a supervised ML algorithm utilized for both data classification and regression It works by defining the optimal hyperplane that maximally splits data points of different classes, creating the maximum possible gap between two classes of data to minimize the classification errors. A higher margin correlates to the higher performance of a model on unseen or new data. This algorithm is very useful when a user needs to have a binary classification such as car vs. bike, human vs. animal, girl vs boy. Figure 4 shows a schematic representation of the SVM model. Sometimes, it is not possible to split data linearly, in those cases kernel transformation techniques can be used to separate the non-linear data into higher dimensions.

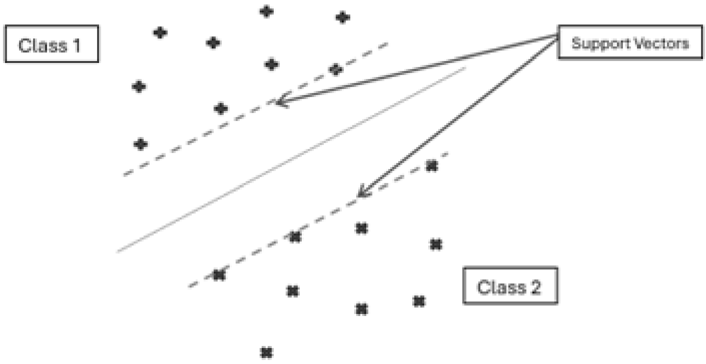


Figure 4. Schematic representation of SVM.

Random Forest (RF)

RF (Chatterjee et al. 2024, Parajulee et al. 2025, Rana et al. 2025) is an ensemble-based machine learning model composed of multiple DTs to make better predictions.. In the ensemble model, each tree works on a different part of the data, and their outcomes are combined by voting for classification or averaging for regression. Basically, it provides the output by merging the results from numbers of DTs and by considering the output which is common from majority of DTs. The advantages of RF are decreases overfitting, predict non-linear relationship, and robust to outliers. Moreover, RF model showcases the capability to handle the missing data, and this model can let the users know that which features are most useful for making prediction. RF can be utilized for both classification (i.e. prediction of types) and regression (i.e. predicting numbers or amount).

K-Nearest Neighbors (KNN)

KNN is the type of supervised ML model which separates objects based on its nearest neighbor's class in the dataset. The KNN model assumes that objects near each other are alike. In this method, first the value of K is chosen, which is the number of neighbors. Then the K number of nearest points to the new point is chosen. The label or value of a new point is determined based on the most common value among the neighbors. In the KNN algorithm, K is a number that informs algorithms how many nearby neighbors look at while making decisions. For instance, if the value of K is 5 while deciding type of vegetable, then the algorithm will look at 5 closest neighbor vegetables to the new one. If 4 out of 5 vegetables are tomato and 1 is potato, then the algorithm will give output as tomato because most of its neighbors are tomatoes.

Gradient Boosting (GB)

GB (Chatterjee et al. 2024, Parajulee et al. 2025, Rana et al. 2025), just like AdaBoost method, is an ensemble technique that inflates estimation accuracy by combining multiple weak learners. In this model, each new model tries to correct the mistakes created by the previous model. GB creates models by reducing a loss function using gradient descent, offering greater flexibility and accuracy for complex, nonlinear problems. The overall prediction is obtained by adding all appropriately weighted individual models. GB is usually opted in geotechnical engineering because of its capability to capture relationships between soil properties, especially when predicting soil suction or SWCC parameters.

Estimation of SWCC

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 February 2026 doi:10.20944/preprints202602.0792.v1

The relationship between matric suction (Ψ) of the soil and water content can be defined as SWCC. It is a principal factor impacting hydraulic and mechanical performance of an unsaturated soil. The SWCC is commonly stated using degree of saturation (S_r), volumetric water content (θ) and gravimetric water content (w). A typical SWCC curve contains three zones (i.e. boundary effect zone, transition zone and residual zone), and Air Entry Value (AEV), residual water content and saturated water content. From practical viewpoint, the SWCC is used to determine Hydraulic conductivity function, shear strength function, permeability using models such as van Genuchten (1980), Brooks and Corey. (1964), and Fredlund and Xing. (1994).

Since the direct estimation of unsaturated hydraulic properties is a challenging task, the prediction based on SWCC is easy and practical approach in geotechnical analysis. Filter paper method, axis translation method, potentiometer, chilled-mirror dewpoint and pressure plate apparatus are some of the laboratory methods to find out SWCC. The suction measurement range of each technique ranges from kPa to MPa – depending upon the technique. The SWCC can be employed in shear strength estimation, landfill design, infiltration analysis, slope stability analysis and seepage modeling. Soil suction can be linked with effective stress with the use of SWCC. It can be ultimately useful to figure out shear strength using models (Vanapalli. 1996).

In mid 2000s, the researchers-initiated adoption of ML to estimate SWCC. Johari et al. (2006) were the early explorers in this area, developed a Genetic Programming (GP) based model to estimate SWCC using basic soil properties such as initial water content, void ratio, clay and silt content, and normalized soil suction. This study proved that ML could provide accurate results (R^2 up to 0.93) comparable to traditional methods, with less laboratory work. In the mid-2010s Nikhil et al. (2016) and Zainal et al. (2018) introduced neural network-based approaches where researchers illustrated that ANNs could predict log nonlinear suction–moisture relationships successfully with high accuracy ($R^2 = 0.83–0.99$). These studies proved that neural networks do not need conventional curve fitting approach for suction estimation. Due to these merits, the methods became more accurate and efficient for soil characterization.

Between 2020 and 2025, various ML algorithms (i.e. KNN, SVM, and regression-based methods) were adopted for more efficient estimation of SWCC. Ramos-Rivera et al. (2021) showcased that for dual porosity soil structures, KNN achieved moderate accuracy with a R^2 ranging between 0.82 and 0.88. Yang et al. (2021) carried out a study of data-driven soil suction prediction, found out that RF model as the most robust model with highest accuracy for predicting SWCC. In 2023, Amir. (2023) proposed an equilibrium suction prediction model that incorporates climatic parameters like temperature, relative humidity and precipitation. This study showed how ML can combine climatic conditions with soil behavior.

In mid 2020s, ensemble-based algorithms including GB, XGB and RF were mostly employed for SWCC prediction due to two major benefits: (a) strong generalization and (b) minimized overfitting. Khanh et al. (2023) and Nazem et al. (2023) conducted large-scale research in which they developed ML models on thousands of samples of varying soil textures. In both studies, high accuracy was achieved ($R^2 > 0.95$), and reliability and scalability of ML-pedotransfer functions were demonstrated. Savio et al. (2024) published research on addressing unique soil conditions like tropical bimodal soils. GB (CatBoost) model was tailored for tropical bimodal soil for SWCC prediction which achieved accuracy of $R^2 = 0.90$ for testing database. In the same year, Guangchang et al. (2024) developed a Multi-Factor SWCC prediction model, that incorporates temperature, salinity, and deformation simultaneously. This study used Bayesian Regularization Neural Network, and it provided a quicker and more flexible way to estimating SWCC.

The research in year 2025 was based on ML frameworks which are more interpretable and have low uncertainty. For instance, Xuzhen et al. (2025) utilized ANN for predicting SWCC parameters. This prediction model minimizes prediction uncertainty and improves reliability. Moreover, Junjie et al. (2025) presented a new parameter called “effective degree of aggregation” which is determined by using ML. It improved suction estimation by recording effects of compaction and initial water content. In the same year Manoj et al. (2025) showcased that suction and moisture estimation can be

doi:10.20944/preprints202602.0792.v1

scaled to regional and landscape level with the use of remote sensing and ML. To sum up, these studies show advancement of ML from simple neural networks to hybrid frameworks, proving ML as a scalable and cost-effective way of estimating SWCC for numerous soil and environmental conditions.

Projection of Soil Water Retention Curve

Throughout last two decades, the application of ML techniques for SWRC modeling has grown from basic neural network-based models to ensemble-based and hybrid prediction frameworks. Initially, Sharad et al. (2004) showcased the ability of a three-layer feed-forward ANN to record the nonlinear relationship between soil suction and water content. This model presented low Root Mean Square Error (RMSE) value (0.002–0.006), showcasing ANNs can form drying, wetting, and scanning paths of the SWRC. This work verifies that ML could cut down the dependency on conventional suction measurement approaches despite having limitations like need for well-distributed datasets.

Krzysztof et al. (2017) further studied ANNs and worked on the prediction of main wetting branch of SWRC from its drying branch, without the requiring labor-intensive laboratory wetting tests. Additionally, this study shows that ANN based model can capture complex nonlinear relationships between soil properties and hydraulic behavior. This model achieved RMSE of 0.021 m^3/m^3. In other words, the accuracy of ANN model was recorded above 95–96%. This work proves that ML not only recreates laboratory tests but also predicts parts of the curve that are difficult to measure.

In the upcoming years, research was done by combining multiple ML algorithms and larger soil datasets. Enzo et al. (2022) created ML-based framework to predict SWRC from basic soil characterization parameters. Also, provided a methodology to generate a SWRC by fitting analytical models to ML predictions. In this study, Ensemble models like Extremely Randomized Trees achieved high accuracies (training R^2 up to 0.99 and testing R^2 around 0.84–0.90). During the same time, Savio et al. (2023) employed large dataset, 794 soil samples combined with basic physical parameters such as texture, porosity, gravel content, and plasticity index. These studies showcased that ML-generated SWRC points could be fitted into analytical models like van Genuchten. 1980 or Costa and Cavalcante. (2021), which shows their direct use in geotechnical engineering.

In recent times, Adel et al. (2024) analyzed two ANN-based strategies —pointwise prediction and continuous prediction of van Genuchten (van Genuchten. 1980) parameters. The outcome showed that pointwise models achieved better accuracy with RMSE of 0.027, whereas continuous models gave consistency of the full curve. Simultaneously, Milan et al. (2025) worked on regression-based models such as Multiple Linear Regression and SVM. Additionally, the authors suggested pedotransfer functions for estimating the drying branch of water retention curves. This study noted that SVMs offer enhanced predictive performance with correlation coefficients 0.878–0.925.

Estimation of Hydraulic Conductivity

Bashar et al. (2025) carried out work by using SWCC parameters like saturated volumetric water content, residual volumetric water content and air-entry suction in ANN model to estimate the unsaturated hydraulic conductivity which is closely connected to soil suction. ANN model achieved high precision (R2=0.9947 for training dataset and R2=0.9349 for testing dataset). It showcases model's capability to capture the correlation between soil suction and hydraulic conductivity. Moreover, this study highlights that prediction accuracy changes with soil texture. Loamy sand and silt loam soils offers steady predictions whereas sandy soils give more errors because of its unique pore structure as well as variability in suction–conductivity relationships. The key benefits of this model are high accuracy and its computational efficiency. On the contrary, key limitations include its variable accuracy across different soil textures.

Stress Prediction

The use of ML to understand and estimate stress has grown in past two decades. This practice initiated when Johari et al. (2013) adopted Gene Expression Programming (GEP) to estimate effective stress parameters of unsaturated soil. This work showed that computational intelligence could illustrate non-linear relationships between soil suction and stress-related parameters, and it does not require any prior assumptions. This model attained accuracy of $R^2 = 0.83$ and this research established a strong foundation for use of ML for stress prediction.

With increasing acceptance of ML, researchers explored more algorithms that are capable to handle large and complex datasets. Singh et al. (2023) merged conventional sensor-based, remote sensing, and ML approaches in their research to estimate soil moisture and suction. This research illustrates how ML techniques like RF, GB, SVM, and ANN can model the relation between soil suction and soil water retention by employing soil properties, climate variables, and remotely sensed data. Algorithms like RF and GB attained $R^2 \approx 0.95$ and verified that ML could exceed prior approaches.

In the coming years, researchers started exploring stress-dependent SWCC modeling. Seyed et al. (2024) carried out one of the first research to generate ML-based framework that explains effect of net stress on soil suction. The authors combined and used a database from 100+ tests and allowed models to understand the relationship between soil properties, net stress, and suction behavior. This study showed that stress influenced suction as much as soil type. From the range of ML models, RF again proved to be the most reliable model, with accuracy R^2 between 0.93 and 0.95. In this study, ML model was not just used to estimate suction but also used to capture mechanical–hydraulic coupling in unsaturated soils. This shift cut down on laboratory work and gave more realistic estimations.

Currently, studies are mostly engaged in specialized applications like defining swelling behavior or generating direct equations for effective stress estimation. Aolin et al. (2025) employed models like MLP, SVM, ELM to determine swelling index and swelling pressure of expansive soils, where ELM performed the best. In the same year, Jagan et al. (2025) worked on the determination of effective stress with improved ML techniques like GP and Multivariate Adaptive Regression Splines (MARS). Their models achieved high precision ($R^2 = 0.97$). These studies proved how ML for SWCC has improved from simple tools to accurate tools for prediction of suction and stress with better reliance.

Slope Stability Analysis

Yangyang et al. (2025) created a RF model to calculate stability of shallow slopes incorporating spatiotemporal variations in unsaturated soil moisture. Soil suction plays a key role in the stability of shallow slopes. RF models were set up by using volumetric water content data for maximum daily rainfall scenarios and predicted slope stability for various rainfall circumstances. This study shows the reliability of RF models in predicting nonlinear relationship between soil moisture and slope stability. The proposed RF model attained high precision of $MAE < 0.15$ and $RMSE < 0.20$. The key benefit of this model is its computational efficiency and accuracy compared to conventional physical models. Whereas limitations include dependency on stimulated training data and its zone-specific model.

Summary

This study underlines the revolutionary impact of ML in measuring soil suction and soil water characteristics curve (SWCC) parameters which are highly important to analyze the behaviour of an unsaturated soil in geotechnical engineering. Traditional methods to measure soil suction are labor intensive, costly as well as less accurate because of environmental factors and instrument sensitivity. ML algorithms provide data driven and scalable options to measure nonlinear relationship between soil properties and suction.

Table 1 summarizes the contribution of different researchers on the application of machine learning and data-driven techniques for determining soil suction. Several ML algorithms were reviewed including DTs, SVMs, RFs, GB, ANNs, KNN, GP, and hybrid approaches. Out of all these algorithms, ensemble methods like GB and RF outperformed every time with highest predictive accuracy as well as robustness. On the other hand, neural networks offered strong performance in modeling hysteresis and soil-water interactions.

Table 1. Summary of application of machine learning for predicting unsaturated properties of soil.

Author	Scientific Contribution, Advantages & Limitation	ML Algorithm Used
Alibrahim et al. (2025)	Contribution: Prediction of unsaturated hydraulic conductivity Advantage: Removal of unrealistic output Limitation: Variable accuracy for different soils	ANN
He et al. (2025)	Contribution: Reduction of uncertainty in SWCC Advantage: Improved accuracy Limitation: Risk of overfitting	ANN
Almuaythir et al. (2025)	Contribution: AI driven soil suction estimation Advantage: High accuracy for MDD & OMC prediction Limitation: less interpretability	XGB, RF, SVR, LSTMN, KNN
Li et al. (2025)	Contribution: Hybrid framework to integrate slope stability & ML Advantage: High computational efficiency Limitation: Zone specific model	RF
Cisty and Povazanova (2025)	Contribution: PTFs development to estimate drying branch of water retention curve Advantage: High accuracy Limitation: Less interpretability of ML models	SVM, MLR
Lamichhane et al. (2025)	Contribution: Identification of the most influential features Advantage: Provides performance metrics for ML models Limitation: Less discussion on soil suction physics	RF, SVR, ANN, XGB, CNN, LSTM
Wang and Vanapalli (2025)	Contribution: Capture nonlinear relationships between soil structure & compaction characteristics Advantage: High accuracy Limitation: Validation requirement for general applicability	PSO-SVR, MGGP
Jagan et al. (2025)	Contribution: GP & MARS models comparison and validation Advantage: High accuracy and low errors Limitation: Computationally intensive	GP, MARS
Zhang and Vanapalli (2025)	Contribution: Prediction of soil suction leveraging ML Advantage: Applicable to a wide range of soil types Limitation: Difficult & time-consuming models	Multilayer Perceptron, SVM, ELM
Liu et al. (2025)	Contribution: Proposal of a fusion feature matrix Advantage: Centimeter-level resolution Limitation: Retraining of model for different soil types	SVM, DNN, RT, Gaussian Regression
Li et al. (2025)	Contribution: Proposed ML framework for UWC prediction Advantage: Improved accuracy & simplified framework Limitation: Generalization Risk	ANN, SVM, RF, XGB
Pereira et al. (2025)	Contribution: ANN Modeling Strategies for SWCC Prediction Advantages: High accuracy and reduced experimental effort Limitation: Black box type nature	MLB, RBFN, ELM
Showkat et al. (2025)	Contribution: ML-Based SWCC Prediction Advantage: Time and cost effective Limitation: Data Dependency on training dataset	RF, XGB, MEP
Pereira et al. (2024)	Contribution: Development of SWCC for tropical soil Advantage: High accuracy & indirect prediction of SWCC Limitation: Need wide range of validation	GB, ANN

Yang et al. (2024)	Contribution: Multi factorial prediction of water content Advantage: High accuracy with automated assessment Limitation: Low interpretability of models	BRNN
Mojtahedi et al. (2024)	Contribution: Soil suction modeling using ML Advantage: Reduced laboratory testing efforts Limitation: Computationally intensive	MLP-NN, GMDH-NN
Abdallah (2024)	Contribution: Comparative evaluation of ML Approaches Advantage: SWRC shape consistency in continuous Prediction Limitation: No hybrid models were tested	Pointwise ANN Continuous ANN
Sharma et al. (2024)	Contribution: Developed ML framework to estimate SWCC Advantage: Robustness across various models Limitation: No real-world dataset validation	MLR, SVR, DTR, RFR, ANN
Yang et al. (2024)	Contribution: physics-informed method to estimate SWCC Advantage: Good performance with limited data Limitation: Computational complexity	PINNs
Li et al. (2024)	Contribution: Development of framework for predicting UWC Advantage: No Need for Predefined Equations Limitation: Data Dependency on training data	RF, XGB, KNN, SVR, BPNN
He et al. (2024)	Contribution: Probabilistic Modeling of SWCC Parameters Advantage: Rigorous Validation to avoid overfitting Limitation: Data Dependency on training data	Bayesian Models, ANN
Huang and Wang (2024)	Contribution: Developed BPNN based model Advantage: Improved Accuracy Limitation: Potential overfitting and complex setup	BPNN
Bakhshi et al. (2023)	Contribution: Integration of image analysis with ML Advantage: Easy to measure soil suction Limitation: Dependency on quality and size of training dataset	GB, DT, RF, ANN, SVM, KNN, LR
Javid (2023)	Contribution: Soil suction and diffusivity estimation Advantage: Cost and time reduction for diffusivity estimation Limitation: Tested model is site specific	NLSR, Ridge regression
Pereira et al. (2023)	Contribution: Applying ML models to estimate SWRC Advantage: High accuracy, time and cost effective Limitation: Lack of interpretability	RF, DT, ERT, SVM, KNN
Pham et al. (2023)	Contribution: Development of ML based PTF to predict SWCC Advantage: Strong generalization with low overfitting Limitation: Single models like SVM had low accuracy	KNN, SVM, DT, NN, RF, GB, XGB
Singh et al. (2023)	Contribution: Application of ML to predict soil suction Advantage: High accuracy, Model is adaptive to different soil Limitation: Models require high computational resources	RF, SVM, ANN, GB, KNN, DT
Nazem et al. (2023)	Contribution: Demonstration of use of ML in modeling SWCC Advantage: High predictive accuracy Limitation: Limited size of the dataset	PSO-XGB PSO-RF PSO-SVR
Qin et al. (2023)	Contribution: Development of an Improved SVM Model Advantage: High Prediction Accuracy Limitation: Complexity in Model Setup	SVM, SVM-PSO
Albuquerque et al. (2022)	Contribution: Development of ML framework to predict SWRC Advantage: High accuracy of Decision tree model Limitation: Limited data and risk of overfitting	MLP, SVM, KNN, DT, RF, ERT
Li et al. (2022)	Contribution: Integration of ML for Predicting Soil Properties Advantage: High accuracy with limited data Limitation: Long computational time	RF, ANN
Gupta at al. (2022)	Contribution: Global-scale mapping of SWCC using ML.	RF

	Advantage: Improved representativeness and robustness Limitation: Reliance on Predicted Soil Properties	
Li at al. (2022)	Contribution: Use of RF & ANN to predict soil properties Advantage: Use of Log Transformation for Better Accuracy Limitation: Models are complex to train and interpret	RF, ANN
Onyelowe et al. (2022)	Contribution: Integration of ML with SWRC Prediction Advantage: Applicable to a wide range of soils Limitation: Complexity in Measurement	SVM, ANN, KNN, RF, XGB
Yang et al. (2021)	Contribution: Soil suction data driven prediction Advantage: Time and cost effective Limitation: Low predictive accuracy	DT, SVM, KNN, GB, RF
Ramos-Rivera et al. (2021)	Contribution: Application of KNN for SWCC prediction Advantage: Laboratory tests reduction Limitation: Prediction time increases with large datasets	KNN
Nobahar and Khan (2021)	Contribution: Development of model for soil matric suction Advantage: High Prediction Accuracy Limitation: ANN models lack transparency	ANN
Sesha et al. (2021)	Contribution: ML-Like approach for pattern recognition Advantage: Reduction in Overestimation Limitation: Complexity in Implementation	Nelder-Mead Simplex Algorithm
Cheng et al. (2021)	Contribution: Used MGGP for suction response to rainfall Advantage: High reliability and applicability Limitation: Limited Spatial and Temporal Scope	MGGP
Pham et al. (2019)	Contribution: Neural network-based PTFs to predict SWCC Advantage: Robustness Across Soil Types Limitation: Complexity in Network Design	Feedforward Neural Networks
Zainal and Fadhil (2018)	Contribution: Developed ANN based model to estimate SWCC Advantage: Decent predictive performance for multiple soils Limitation: Risk of overfitting	ANN
Saha et al. (2018)	Contribution: Developed ANN based SWCC estimation model Advantage: Separate models for plastic & non-plastic Soils Limitation: Small Training and Validation Sets	ANN
Lamorski et al. (2017)	Contribution: Estimation of main wetting branch of the SWRC Advantage: Practicable in large-scale applications Limitation: Requires training when applied to new regions	ANN
Saha et al. (2017)	Contribution: Developed ANN-based models to predict SWCC Advantage: High accuracy with less experimental efforts Limitation: The model can't adapt new data unless retrained	ANN
Nikhil et al. (2016)	Contribution: Developed ANN based model to estimate SWCC Advantage: Integration of ML with analytical models Limitation: Accuracy may decrease without retraining	ANN
Johari et al. (2013)	Contribution: Use of GEP to model unsaturated soil behavior Advantage: Provides interpretable expressions for practical use. Limitation: Comparison with deep learning models is not done.	Gene Expression Programming
Johari et al. (2011)	Contribution: Genetic Algorithm-Based Neural Network to model the mechanical behavior of unsaturated soils. Advantage: Improved Prediction Accuracy Limitation: Data Dependency on training data	ANN, GABNN
Johari et al. (2011)	Contribution: Development and Comparison of GBNN & GP Advantage: Time and cost effective Limitation: Model Complexity and less interpretability	GBNN, GP
Yusuf (2007)	Contribution: Developed a predictive model for total soil Advantage: Reduces the need for time-consuming tests	ANN

	Limitation: Black Box Nature: Lack of interpretability	
Johari et al. (2006)	Contribution: Development of GP model to estimate SWCC Advantage: High Accuracy Limitation: Complex model, dependency on training dataset	Genetic Programming
Jain et al. (2004)	Contribution: ANN based prediction of SWRC Advantage: High accuracy with low RMSE Limitation: Model is less transparent and less interpretable	ANN

Additionally, research on remote sensing, image analysis and climate data enhanced the estimation capacity of ML frameworks for soil suction. All these advancements improve slope stability analysis, foundation design as well as soil moisture management and reduces dependency on laboratory testing.

In contrast, ML models also had some constraints. The efficiency of ML models depends on quality, size, and representativeness of training datasets. Some ML models are very difficult to interpret due to its complex nature. However, by using ML algorithms, the way of studying soil suction and SWCC can be revolutionized. Engineers and researchers have a better way to study soil behavior quickly and efficiently.

Future Directions

The findings of the literature review suggest that there is a wide range of application of machine learning algorithms for prediction of SWCC, SRCC and different soil-water parameters. Moreover, some of the studies used deep learning algorithms such as CNN and LSTM to develop different relationships related to the unsaturated behavior of soils. One of the common trends reported from different studies are the accuracy of models, interpretability of models, lack of large volumes of data for model development and validation, and models cannot predict behaviors for different types of soils.

Model accuracy and application of models for different types of soil can be increased by using large language models for predicting the unsaturated behaviors of soil. Large language models are based on transformer architecture (Ansari et al. 2025, Chatterjee et al. 2026) and are developed on huge volumes of data. The huge volume of data increases the generalizability as well as the accuracy of the models.

The lack of data for training machine learning and deep learning models can be solved by augmenting experimental data leveraging data augmentation techniques such as random noise. Moreover, numerical modelling using different commercial and educational software can be used for generating data for model development.

The problem of the interpretability of the models can be solved using different statistical models such as multi-linear regression, logistic regression, lasso and ridge regression or can be solved by using tree-based machine learning models such as decision-tree, random forest and gradient boosting algorithm.

References

1. Abdallah, A. (2022, May). Artificial neural network prediction of the water retention curve from physical soil parameters: comparing continuous and pointwise approaches. In *20th International Conference on Soil Mechanics and Geotechnical Engineering*.
2. Albuquerque, E. A. C., Borges, L. P. D. F., Cavalcante, A. L. B., & Machado, S. L. (2022). Prediction of soil water retention curve based on physical characterization parameters using machine learning. *Soils and Rocks*, *45*(3), e2022000222.
3. Alibrahim, B., Habib, M., & Habib, A. (2025). Utilizing soil–water characteristic curve parameters in custom artificial neural network models to predict the unsaturated hydraulic conductivity. *Discover Artificial Intelligence*, *5*(1), 1-15.

doi:10.20944/preprints202602.0792.v1

4. Almuaythir, S., Zaini, M. S. I., & Lodhi, R. H. (2025). Predicting soil compaction parameters in expansive soils using advanced machine learning models: a comparative study. *Scientific Reports*, *15*(1), 24018.
5. Ansari, F., Chatterjee, K., Li, J. Q., Wang, K., & Golalipour, A. Multi-Object Pavement Surface Feature Detection with CNN and Transformer Deep Learning Architecture. In *Airfield and Highway Pavements 2025* (pp. 350-359).
6. Bakhshi, A., Alamdari, P., Heidari, A., & Mohammadi, M. H. (2023). Estimating soil–water characteristic curve (SWCC) using machine learning and soil micro-porosity analysis. *Earth Science Informatics*, *16*(4), 3839-3860.
7. Chatterjee, K., Vivanco, D., Yang, X., & Li, J. Q. (2024). Enhancing Pavement Performance through Balanced Mix Design: A Comprehensive Field Study in Oklahoma. In *International Conference on Transportation and Development 2024* (pp. 511-522).
8. Chatterjee, K., Li, J. Q., Ansari, F., Munna, M. R., Parajulee, K., & Schwennesen, J. (2026). Hybrid LSTM-Transformer Models for Profiling Highway–Railway Grade Crossings. *Journal of Transportation Engineering, Part A: Systems*, *152*(2), 04025138.
9. Cheng, Z. L., Yang, S., Zhao, L. S., Tian, C., & Zhou, W. H. (2021). Multivariate modeling of soil suction response to various rainfall by multi-gene genetic programing. *Acta Geotechnica*, *16*(11), 3601-3616.
10. Cisty, M., & Povazanova, B. (2021, November). Evaluation of water retention curves by regression and machine learning methods. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1203, No. 3, p. 032088). IOP Publishing.
11. Costa, M. B. A. D., & Cavalcante, A. L. B. (2021). Bimodal soil–water retention curve and k-function model using linear superposition. *International Journal of Geomechanics*, *21*(7), 04021116.
12. dos Santos Pereira, S. A., de FN Gitirana Jr, G., Mendes, T. A., & de Aquino Gomes, R. (2025). Artificial neural networks for the prediction of the soil-water characteristic curve: An overview. *Soil and Tillage Research*, *248*, 106466.
13. dos Santos Pereira, S. A. Predicting the Soil-Water Characteristic Curve of Tropical Bimodal Soils Using Gradient Boosting.
14. dos Santos Pereira, S. A., Silva Junior, A. C., Mendes, T. A., Gitirana Junior, G. D. F. N., & Alves, R. D. (2024). Prediction of soil–water characteristic curves in bimodal tropical soils using artificial neural networks. *Geotechnical and Geological Engineering*, *42*(5), 3043-3062.
15. Erzin, Y. (2007). Artificial neural networks approach for swell pressure versus soil suction behaviour. *Canadian Geotechnical Journal*, *44*(10), 1215-1223.
16. Fazel Mojtahedi, S. F., Akbarpour, A., Darzi, A. G., Sadeghi, H., & van Genuchten, M. T. (2024). Prediction of stress-dependent soil water retention using machine learning. *Geotechnical and Geological Engineering*, *42*(5), 3939-3966.
17. Gupta, S., Papritz, A., Lehmann, P., Hengl, T., Bonetti, S., & Or, D. (2022). Global mapping of soil water characteristics parameters—fusing curated data with machine learning and environmental covariates. *Remote Sensing*, *14*(8), 1947
18. He, X., Cai, G., & Sheng, D. (2025). Indirect models for SWCC parameters: Reducing prediction uncertainty with machine learning. *Computers and Geotechnics*, *177*, 106823.
19. Huang, Y., & Wang, Z. (2024). Exploring the Hydraulic Properties of Unsaturated Soil Using Deep Learning and Digital Imaging Measurement. *Water*, *16*(24), 3550.
20. Jagan, J., Vinod, B. R., Gobinath, S., Samui, P., & Das, G. J. (2025). Comparative analysis of machine learning models for predicting effective stress parameters in unsaturated soils. *Modeling Earth Systems and Environment*, *11*(5), 345.
21. Jain, S. K., Singh, V. P., & Van Genuchten, M. T. (2004). Analysis of soil water retention data using artificial neural networks. *Journal of Hydrologic Engineering*, *9*(5), 415-420.
22. Javid, A. H. (2023). *Variation of soil suction and application of remote sensing in evaluating unsaturated soil behavior within vadose zone* (Doctoral dissertation, Oklahoma State University).
23. Johari, A., Habibagahi, G., & Ghahramani, A. (2006). Prediction of soil–water characteristic curve using genetic programming. *Journal of geotechnical and geoenvironmental engineering*, *132*(5), 661-665.

24. Johari, A., Habibagahi, G., & Ghahramani, A. (2011). Prediction of SWCC using artificial intelligent systems: A comparative study. *Scientia Iranica*, *18*(5), 1002-1008.
25. Johari, A., Habibagahi, G., & Nakhaee, M. (2013). Prediction of unsaturated soils effective stress parameter using gene expression programming. *Scientia Iranica*, *20*(5), 1433-1444.
26. Johari, A., Javadi, A. A., & Habibagahi, G. (2011). Modelling the mechanical behaviour of unsaturated soils using a genetic algorithm-based neural network. *Computers and Geotechnics*, *38*(1), 2-13.
27. Lamichhane, M., Mehan, S., & Mankin, K. R. (2025). Soil moisture prediction using remote sensing and machine learning algorithms: A review on progress, challenges, and opportunities. *Remote Sensing*, *17*(14), 2397.
28. Lamorski, K., Šimůnek, J., Sławiński, C., & Lamorska, J. (2017). An estimation of the main wetting branch of the soil water retention curve based on its main drying branch using the machine learning method. *Water Resources Research*, *53*(2), 1539-1552.
29. Li, J., Zhou, P., Pu, Y., Ren, J., Zhang, F., & Wang, C. (2024). Comparative analysis of machine learning techniques for accurate prediction of unfrozen water content in frozen soils. *Cold Regions Science and Technology*, *227*, 104304.
30. Li, M., Ma, S., Li, J., Ren, J., & Wang, C. (2025). Application of machine learning for predicting unfrozen water content in frozen soils: A review. *Cold Regions Science and Technology*, 104711.
31. Li, Y., Rahardjo, H., Satyanaga, A., Rangarajan, S., & Lee, D. T. T. (2022). Soil Database Development In Singapore with the Application of Machine Learning Methods in Soil Properties Prediction. *Available at SSRN 4047079*.
32. Li, Y., Rangarajan, S., Cheng, Y., Rahardjo, H., & Satyanaga, A. (2025). Random forest-based prediction of shallow slope stability considering spatiotemporal variations in unsaturated soil moisture. *Scientific Reports*, *15*(1), 8751.
33. Li, Y., Rahardjo, H., Satyanaga, A., Rangarajan, S., & Lee, D. T. T. (2022). Soil database development with the application of machine learning methods in soil properties prediction. *Engineering Geology*, *306*, 106769.
34. Liu, G., Tian, S., Wang, Q., Wang, H., & Kong, L. (2025). High-resolution measurement of moisture filed at soil surface with interfered image processing method and machine learning techniques. *Journal of Hydrology*, *652*, 132623.
35. Nazem, M., Kardani, N., Moridpour, S., & Zhou, A. (2023). Prediction of Soil-Water Characteristic Curve using optimised machine learning approaches. In L. Zdravkovic, S. Kontoe, A. Tsiampousi, & D. Taborda (Eds.), Proceedings of the 10th European Conference on Numerical Methods in Geotechnical Engineering (NUMGE 2023). International Society for Soil Mechanics and Geotechnical Engineering. https://doi.org/10.53243/NUMGE2023-312
36. Nikhil, N. V., Seok, Y., Lee, S. R., & Lee, D. H. (2016). ANN based estimation of SWCC fitting parameters for Korean weathered soil considering in-situ characteristics. In *The 2016 world congress on advances on Civil, Environmental, and Materials Research.(ACEM16)*.
37. Nobahar, M., & Khan, M. S. (2021). Prediction of matric suction of highway slopes using autoregression artificial neural network (ANN) model. In *Geo-extreme 2021* (pp. 40-50).
38. Onyelowe, K. C., Mojtahedi, F. F., Azizi, S., Mahdi, H. A., Sujatha, E. R., Ebid, A. M., ... & Aneke, F. I. (2022). Innovative overview of SWRC application in modeling geotechnical engineering problems. *Designs*, *6*(5), 69.
39. Parajulee, K., Chatterjee, K., & Li, J. (2025). Leveraging Original Equipment Manufacturer Vehicle Sensor Data for Enhanced Roadway Safety. *International Journal of Pavement Research and Technology*, 1-18.
40. Pham, K., Kim, D., Yoon, Y., & Choi, H. (2019). Analysis of neural network based pedotransfer function for predicting soil water characteristic curve. *Geoderma*, *351*, 92-102.
41. Pham, K., Kim, D., Le, C. V., & Won, J. (2023). Machine learning-based pedotransfer functions to predict soil water characteristics curves. *Transportation Geotechnics*, *42*, 101052.
42. Qin, W., & Fan, G. (2023). Estimation and predicting of soil water characteristic curve using the support vector machine method. *Earth Science Informatics*, *16*(1), 1061-1072.

43. Raghuram, A. S. S., Basha, B. M., & Raviteja, K. V. N. S. (2021). Variability characterization of SWCC for clay and silt and its application to infinite slope reliability. *Journal of Materials in Civil Engineering*, *33*(8), 04021180.
44. Ramos-Rivera, J., Parra-Holguín, D., Valencia-González, Y., & Echeverri-Ramírez, O. (2021). Estimating soil-water characteristic curve based on soil type and best-fitting regressions derived from a simplified method using Aburra Valley dataset. In *MATEC web of conferences* (Vol. 337, p. 02002). EDP Sciences.
45. Rana Munna, M., Chatterjee, K., Parajulee, K., & Li, J. Q. Effect of Pavement Surface Characteristics on Adverse Road Conditions. In *Airfield and Highway Pavements 2025* (pp. 360-369).
46. Saha, S., Gu, F., Luo, X., & Lytton, R. L. (2017). Prediction of soil-water characteristic curve using artificial neural network approach. In *PanAm Unsaturated Soils 2017* (pp. 124-134).
47. Saha, S., Gu, F., Luo, X., & Lytton, R. L. (2018). Prediction of soil-water characteristic curve for unbound material using Fredlund–Xing equation-based ANN approach. *Journal of Materials in Civil Engineering*, *30*(5), 06018002.
48. Sharma, S., Rathor, A. P. S., & Sharma, J. K. (2025). Prediction of soil water characteristic curve of unsaturated soil using machine learning. *Multiscale and Multidisciplinary Modeling, Experiments and Design*, *8*(1), 72.
49. Showkat, R., Jalal, F. E., & Babu, G. S. (2025). Estimation of Soil Water Characteristic Curve Using Machine-Learning Algorithms and Its Application in Embankment Response. *Journal of Computing in Civil Engineering*, *39*(3), 04025012.
50. Singh, A., Gaurav, K., Sonkar, G. K., & Lee, C. C. (2023). Strategies to measure soil moisture using traditional methods, automated sensors, remote sensing, and machine learning techniques: review, bibliometric analysis, applications, research findings, and future directions. *Ieee Access*, *11*, 13605-13635.
51. Van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil science society of America journal*, *44*(5), 892-898.
52. Vanapalli, S. K., Fredlund, D. G., Pufahl, D. E., & Clifton, A. W. (1996). Model for the prediction of shear strength with respect to soil suction. Canadian geotechnical journal, 33(3), 379-392.
53. Wang, J., & Vanapalli, S. (2025). A Framework for Estimating Matric Suction in Compacted Fine-Grained Soils Based on a Machine Learning-Assisted Conceptual Model. *International Journal for Numerical and Analytical Methods in Geomechanics*.
54. Yang, H. Q., Shi, C., & Zhang, L. (2025). Ensemble learning of soil–water characteristic curve for unsaturated seepage using physics-informed neural networks. *Soils and Foundations*, *65*(1), 101556.
55. Yang, S., Zheng, P. Q., Yu, Y. T., & Zhang, J. (2021, October). Probabilistic analysis of soil-water characteristic curve based on machine learning algorithms. In *IOP conference series: earth and environmental science* (Vol. 861, No. 6, p. 062030). IOP Publishing.
56. Yang, G., Liu, J., Liu, Y., Wu, N., & Liu, T. (2024). A Prediction Model for Soil–Water Characteristic Curve Based on Machine Learning Considering Multiple Factors. Buildings, 14(7), 2087.
57. Zainal, A. K. E., & Fadhil, S. H. (2018). Prediction of soil water characteristic curve using artificial neural network: a new approach. In *MATEC Web of Conferences* (Vol. 162, p. 01014). EDP Sciences.
58. Zhang, A., & Vanapalli, S. K. (2025). Estimation of the Swelling Index and Swelling Pressure of Expansive Soils Using Multiple Artificial Intelligence Techniques. *International Journal of Geomechanics*, *25*(11), 04025244.