Earth Science Informatics (2023) 16:3839–3860
https://doi.org/10.1007/s12145-023-01131-3

RESEARCH

Estimating soil–water characteristic curve (SWCC) using machine learning and soil micro-porosity analysis

Aida Bakhshi[1] · Parisa Alamdari[1] · Ahmad Heidari[2] · Mohmmad Hossein Mohammadi[2]

Received: 1 September 2023 / Accepted: 18 October 2023 / Published online: 27 October 2023

Abstract

This study explores soil water characteristic curve (SWCC) prediction through informatics and machine learning. Utilizing these techniques, SWCC prediction was significantly simplified, enabled by the Orange.3 data mining software's integration of diverse soil properties. This integration eliminated the need for extensive programming, establishing a link between scientific insights and engineering applications. Limitations emerged in models relying solely on matric suction for SWCC prediction, evident through a Mean Absolute Error exceeding 0.08 and an R-squared value below 40% in the test dataset. To enhance accuracy, a comprehensive approach encompassing various soil properties, such as bulk density, organic carbon content, and micro-porosity characteristics, was employed. The Gradient Boosting algorithm excelled, yielding near-perfect SWCC estimations with RMSE and P_i values of 0.016 and 0.03, respectively. Likewise, AB, Random Forest, and Tree models displayed highly accurate predictions with RMSE and P_i values below 0.03 and 0.04, respectively. However, Neural Network, SVM, kNN, and Linear Regression models showed no improvements, even with added soil properties. Feature importance analysis highlighted matric suction's critical role in select models and soil micro-porosity characteristics' contribution to lowering RMSE by up to 0.04. These findings are pivotal in understanding errors in SWCC prediction, especially in cases of matric suctions surpassing the SWCC inflection point, with these errors, though present, minimally impacting model efficacy due to diminishing variations at high matric suctions.

Keywords Data mining · Machine Learning · Prediction Feature importance · Predictive models · Soil water characteristic curve

Introduction

By synergizing geospatial data, remote sensing techniques, and diverse Earth science datasets, informatics plays a pivotal role in fostering a holistic comprehension of natural systems and processes (Sermet and Demir 2019). In the realm of engineering applications, this encompassing knowledge significantly enhances the precision of decision-making and predictive capabilities. For instance, within geotechnical engineering, Earth Science Informatics facilitates the integration of soil behavior, geological insights, and hydrological information into prognostic models, thereby refining the design and evaluation of critical structures such as foundations, dams, and slopes (Pham et al. 2023). This interdisciplinary methodology not only optimizes engineering outcomes but also advances sustainable solutions by meticulously accounting for the intricate interplay between natural and constructed environments.

Communicated by H. Babaie.

✉ Parisa Alamdari
p_alamdari@znu.ac.ir

Aida Bakhshi
aida_bakhshi@ut.ac.ir

Ahmad Heidari
ahaidari@ut.ac.ir

Mohmmad Hossein Mohammadi
mhmohmad@ut.ac.ir

[1] Department of Soil Science, Faculty of Agriculture, University of Zanjan, Zanjan 38791-45371, Iran

[2] Department of Soil Science, Faculty of Agricultural Engineering & Technology, College of Agriculture & Natural Resources, University of Tehran, Daneshkadeh Ave., Karaj 3158777871, Iran

The Soil Water Characteristic Curve (SWCC) offers valuable insights, both directly and indirectly, into the behavior of water within unsaturated soils (Zhai and Rahardjo 2012). The accurate determination of the SWCC for a given soil necessitates a combination of precise measurement techniques and predictive methodologies. Nonetheless, the field, laboratory, and computer vision-based measurements of SWCC are resource-intensive, laborious, time-consuming, and occasionally unfeasible due to challenges concerning scaling, spatial variability, and site inaccessibility (Achieng 2019). As a result, the utilization of modeling procedures has become a widely adopted approach for predicting SWCC (Dobarco et al. 2019).

The application of machine learning (ML) algorithms in soil moisture research has witnessed a substantial upsurge. These algorithms are favored for their non-parametric essence and adeptness in capturing intricate and non-linear associations (Padarian et al. 2019). ML techniques employed to estimate SWCC predominantly fall within the realm of supervised learning, which entails the provision of a labeled training dataset containing known output values. The model is then trained through algorithms applied to the input dataset, enabling the prediction of the desired output. The training process continues until the model attains the intended accuracy on the training dataset. Supervised learning finds widespread application in tasks encompassing classification and regression (Rani et al. 2022).

In various studies, researchers have used different algorithms to understand the complex relationship between soil properties and water content. These methods range from traditional models to advanced ones like artificial neural networks (ANN), support vector machines (SVM), random forests (RF), and deep neural networks (DNN). Neural network models consist of input, hidden, and output layers, with the number of hidden layers determined by problem complexity. For general geotechnical engineering, it's often found that a single hidden layer suffices (Wang et al. 2022). Incorporating data preprocessing with Bayesian regularization neural networks, Pham et al. (2019) showcased the capability to enhance the precision of predicting SWCC and illustrated that three-hidden-layer BRNN-PTF showed a considerable outperformance to predict the soil water content. The effectiveness of these algorithms was assessed through metrics including the Root Mean Squared Error (RMSE) and R-squared (R^2) values, thereby furnishing insights into the prognostic capacities of the models. A comprehensive synthesis of these studies is presented in Table 1, outlining the spectrum of employed ML algorithms, their corresponding performance metrics, noteworthy observations, and the specific features assimilated within the models.

Several PTFs for predicting SWCC with acceptable accuracy were proposed by researchers (Pachepsky et al. 2006; Leij et al. 2004; Børgesen et al. 2008). However, rare attention has been given to assessing the significance of the features encompassed within the provided dataset. While Pham et al. (2023) undertook the endeavor of determining the importance of database features to construct their ML-PTFs and assessed its effectiveness in SWCC estimation, they did not specifically focus on the role of soil porosity in water retention within soil (Tuller et al. 2004).

The works by Fredlund and Rahardjo (1993) as well as Hopmans and Dane (1986) underscore the significance of matric suction in governing water dynamics and mechanical responses within soils of varying compositions, encompassing sandy and silty soils. While the assertion by Achieng (2019) maintains that a deep learning approach enables the exclusive prediction of SWCC using soil matric suction as the sole input feature, other investigations posit that precise SWCC predictions frequently derive advantages from a broader spectrum of input parameters, obtained through laboratory analyses and image processing methods. The incorporation of supplementary parameters like soil texture, porosity, particle size distribution, and mineral composition augments the predictive efficacy of ML models.

Nguyen et al. (2014) and Vereecken et al. (2010) reached the conclusion that incorporating soil structure information into Pedotransfer Functions (PTFs) holds the potential to enhance their performance. They further recommended in-depth exploration to determine the robustness of these improvements across various data mining techniques and diverse categories of PTFs. Employing ImageJ's built-in capability for soil porosity analysis, Bakhshi et al. (2023) demonstrated that the water retention capacity and SWCC pattern are contingent on soil pore geometrics. This feature yields valuable output parameters, encompassing porosity surface area (Total Area of Porous Regions, cm^2), volume (Total Number of Porous Voxels × Voxel Volume, cm^3), elongation (Major Axis Length/Minor Axis Length, dimensionless), flatness (Average Length of Major Plane/Average Length of Minor Plane), sphericity ($4\pi \times \text{area}/\text{perimeter}^2$, dimensionless), and compactness (volume of the porous region/surface area of the porous region, dimensionless).

Building upon these findings, our study leveraged intricate soil structural attributes derived from image analysis as inputs for the ML technique employed. To this end, in conjunction with other frequently employed algorithms, we assessed the application of gradient boosting (GB) and Ada Boost (AdaB) in estimating SWCC using ML within the Orange.3 data mining software. The predictive exercise was undertaken under two scenarios: 1) using matric suction as the sole predefined input, and 2) integrating an array of input parameters garnered from both laboratory measurements and image analysis techniques.

Table 1 Review on using machine learning algorithm to estimate SWCC

Reference	Used ML algorithm	performance (RMSE-R^2)	Special comment	Used features
Pham et al. 2023	kNN NN SVM DT RF GB XGB	0.066–0.847 0.070–0.828 0.097–0.662 0.052–0.902 0.038–0.921 0.033–0.964 0.028–0.972	XGB-PTF showed the best performance in predicting soil moisture content (θ)	H, Matric suction Sand percentage Silt percentage Clay percentage P_b, Soil bulk density P_p, Soil particle density n
Pham et al. 2019	NN	0.028–0.977	Three-hidden-layer BRNN-PTF showed a considerable outperformance to predict the soil water content	H, Matric suction Sand percentage Silt percentage Clay percentage P_b, Soil bulk density P_p, Soil particle density porosity
Achieng 2019	ANN DNN SVM	0.028–0.97	RBF-based support vector regression (SVR) outperformed SVR with linear and polynomial kernels, single-layer ANN, and deep neural network (DNN) models	Soil water content at specific Matric suction
Diao et al. 2021	ANN	0.03–0.95	The best estimation of θ was from the model based on the combined use of spectral feature parameters and an ANN algorithm	Spectral features
Senyurek et al. 2020	ANN RF SVM	0.061 – None 0.052 – None 0.065 – None	Particularly, soil texture features are identified as the most influential ancillary input for the SM prediction	CYGNSS reflectivity TES incidence angle NDVI Volumetric Water Content Terrain elevation clay proportions silt proportions
Amanabadi et al. 2019	ANN ANFIS	0.032–0.96 0.035–0.92	Comparative results, regardless of the soil texture, indicate that the ANN having $\log\lvert h\rvert$, θs, sand, clay, and silt fractions, porosity, and bulk density as its input variables results in the highest performance	$\log\lvert h\rvert$, matric suction saturated water content Sand percentage Silt percentage Clay percentage P_b, Soil bulk density P_p, Soil particle
Cai et al. 2019	deep learning regression network	None – 0.98	The DNNR has ability to keep prediction error near the zero point	Average temperature average pressure Relatively humidity average Wind speed Land temperature Daily precipitation Initial soil moisture

Table 1 (continued)

Reference	Used ML algorithm	performance (RMSE-R^2)	Special comment	Used features
Nguyen et al. 2017	MLR ANN SVM kNN	0.068–0.84 0.053–0.82 0.052–0.88 0.049–0.89	The evaluation of results confirm the superiority of the ANN and kNN approaches in modelling the relationship between soil and water as a complex system even when a limited dataset is available	θ_h, Soil water content at specific Matric suction Organic carbon percentage P_b, Soil bulk density Sand percentage Silt percentage Clay percentage
Gunarathna et al. 2019	ANN kNN RF	0.048–0.52 0.046–0.57 0.043–0.60	The RF algorithm showed the most robust machine learning algorithm for developing PTFs for tropical Sri Lankan soils	θ_h, Soil water content at 10, 33, and 1500 kPa matric suction Organic carbon percentage P_b, Soil bulk density Sand percentage Silt percentage Clay percentage
Rastgou et al. 2020	RF	~0.04–~0.98	It was found that the RF method had a better performance ($P < 0.05$) than the NLR method in the training step of the prediction of the SWRC	Broad range of soil physical and hydraulic properties

Material and methods

Soil sampling, treatment preparation, and experimental setup

This study delved into the intricate effects of diverse treatments on soil porosity and SWCC within soil samples obtained from distinct textural classes in Central Iran. The samples originated from Arenosols (coordinates 35° 54′ N and 50° 32′ E) and Vertisols (coordinates 36° 22′ N and 49° 35′ E) and comprised loamy sand and silty clay soils.

Soil sampling and analysis

Topsoil samples (0–10 cm) were collected, dried, and sieved to achieve a particle size of 2 mm, ensuring uniformity in the analysis. Established methods were employed to evaluate pivotal soil properties, vital for predicting SWCC. These properties included soil organic carbon (SOC) (Walkley and Black 1934), serving as an indicator of organic matter content; particle size distribution (PSD) (Gee and Or 2002), revealing soil texture composition; cation exchange capacity (CEC) (Rhoades 1983), a measure of ions retention capacity; electrical conductivity (EC) (Rhoades 1996), reflecting soil salinity; pH (Thomas 1996), indicating soil acidity; and parameters characterizing soil porosity (a, n, θ_s, and θ_r) (Dexter et al. 2008). To preserve sample integrity, bulk density was determined through the core method, utilizing Kopecky rings (5 cm in height and 5 cm in diameter) (Grossman and Reinsch 2002).

Treatment preparation and experimental setup

The soil samples underwent a comprehensive range of treatments, each meticulously designed to investigate specific soil responses. This included the application of various levels of $CaCO_3$, ranging from 0 to 5%, to assess the influence of calcium carbonate on soil properties (Huang et al. 2016). Similarly, $Fe_2O_3.7H_2O$, varying from 0 to 2%, was introduced to explore the effects of iron oxide (Li et al. 2021). The incorporation of vermicompost, at varying levels (0% to 2%), allowed insights into how organic carbon and nutrient content impacted soil characteristics (Demir 2020). Furthermore, combined treatments involving $CaCO_3$, $Fe_2O_3.7H_2O$, and vermicompost in specific ratios (1.5%, 0.5%, and 1%, respectively), as well as higher levels, were investigated. Additionally, treatments were prepared based on Sarkar et al. (2014), where organic matter, iron oxide, and $CaCO_3$ were removed at specific levels.

Cation treatment and structural degradation

To comprehend the influence of cations on soil structure, solutions containing $CaCl_2$ and NaCl at concentrations of 0, 5, 10, and 20 meq L^{-1} (Mi et al. 2018) were employed for irrigation during the incubation period. This facilitated an examination of how varied cation levels affected soil behavior. The study also encompassed a comprehensive analysis of degraded treatments, achieved through a tailored consolidation process designed to replicate conditions resulting from natural degradation.

Incubation and testing

Following treatment application, the soil samples were placed in pots and incubated at room temperature (24 ~ 26 °C). To emulate real-world scenarios, the samples underwent numerous cycles of shrink-swell and wetting–drying, a process repeated 20 times. Rigorous monitoring was conducted to capture any variations. Rewetting was done until moisture content equal to field capacity was achieved by carefully adding water to the sponge cover placed on top of the columns to avoid disturbing soil conditions.

Sample count and analysis

The study encompassed a total of 128 samples, involving diverse combinations of amendment treatments and degraded treatments. A subset of samples was selected for direct measurements of the SWCC, while others underwent preparation for subsequent image analysis through impregnation with a mixture of polyester resin, catalyst, hardener, and fluorescent dye. For a comprehensive breakdown of specific treatments, consult Table 2.

Determination of the SWCC

The SWCC was constructed by combining the results obtained for water content at both low matric suctions (0, 10, 20, 40, and 70 cm) using a sandbox apparatus (Cresswell et al. 2008) and higher matric suctions (100, 300, 500, 1000, 3000, 5000, 9000, and 15,000 cm) using pressure plate/pressure membrane apparatus (Dane et al. 2002). Despite the acknowledged methodological limitations, this approach stands as the most prevalent technique for SWCC measurement (Schindler et al. 2012). Undisturbed samples were used to determine the lower matric suctions ranging from 0 to 1000 cm, while disturbed samples were utilized for matric suctions ranging from 3000 to 15,000 cm.

Table 2 Initial soil samples properties

Soil Texture	Particle size distribution			Structural and hydraulic properties						CEC ($cmol^+$. kg^{-1})	EC ($\mu S.cm^{-1}$)	OC (%)	pH
	Clay (%)	Silt (%)	Sand (%)	Bulk Density ($kg.m^{-3}$)	A (cm^{-1})	n (-)	Θ_s ($g{\cdot}g^{-1}$)	Θ_r ($g.g^{-1}$)	h_i (cm)				
Loamy Sand	12.1	5.3	82.6	1340	0.023	2.07	0.247	0.058	70	4.8	782	0.12	8.12
Silty Clay	41.8	48.1	10.1	1290	0.003	1	0.382	0.175	300	24.1	426	0.42	7.49

Sample preparation, imaging, and image preprocessing

A total of 128 soil samples, having undergone pre-treatment, were subjected to a methodical impregnation procedure involving a mixture of polyester resin and styrene in a 5:1 ratio (Eben et al. 2020), accompanied by suitable amounts of hardener and catalyst. To enhance the visibility of soil pores during the forthcoming digital imaging phase, a brightener, 2 g.L^{-1} of fluorescent dye, was deliberately introduced into the mixture (Ringrose-Voase 1996). This strategic inclusion served to augment the luminance of pores under UV illumination, facilitating their subsequent visual analysis.

The impregnation process unfolded within plastic containers that were housed in a meticulously regulated environment within a vacuum desiccator. The desiccator underwent an evacuation process set at 8 psi for a duration of 2 h. This step assumed paramount importance, ensuring comprehensive resin infiltration throughout the samples and the effective displacement of air from the pores. Consequent to this evacuation, the samples were refilled with the impregnation mixture and subjected to an additional two-hour vacuum cycle, thereby optimizing resin penetration. Following this sequence, the samples were diligently sealed to counteract the rapid volatilization of styrene. Approximately seven days subsequent to sealing, the samples were unsealed, allowing for the gradual and natural volatilization of styrene over time, ultimately leading to the desired hardening of the polyester resin. This polymerization process culminated after an average duration of approximately 75 days (Wei et al. 2019).

Upon the completion of the resin hardening phase, the samples underwent meticulous cutting and polishing procedures. Each individual sample was meticulously subjected to two horizontal and two vertical cuts, which collectively resulted in the exposure of four proximate surfaces, thus providing an extensive range of viewing angles for subsequent imaging. Notably, the imaging process was executed within an environment carefully configured as a controlled dark room, a setting that was equipped with specialized UV lamps. The strategic utilization of these lamps aimed to maximize the fluorescence emission of the dye embedded within the pores, thereby significantly enhancing their visibility. The images were captured using a digital camera boasting a resolution of 12 MP and an aperture of f/1.8.

Following the successful acquisition of the color images, the next crucial step entailed their systematic preprocessing within the ImageJ software. This versatile software platform facilitated an array of operations essential for effective analysis. Specifically, the color images were subject to grayscale conversion, a step that transformed the images into a grayscale format, subsequently enhancing their suitability for further analysis. To accentuate the visual distinction between pores and solid regions, thresholding was systematically applied to the grayscale images, resulting in their conversion into binary images. This binary representation enabled a sharp demarcation between pores, represented as white pixels, and solid areas, depicted as black pixels.

The stacking of these binary images yielded an ensemble of four 3D volumes for each individual sample. These volumes served as the foundational data for the subsequent analysis, providing a multi-dimensional perspective of the spatial distribution of pores within the samples. The ensuing analysis drew extensively from the specialized 2D and 3D plugins embedded within the ImageJ software. These plugins facilitated the extraction of key parameters characterizing the identified pores. This encompassed pivotal parameters including the determination of 3D porosity, pore sphericity, aspect ratio, and the orientation of pore objects. The orientation was expressed through two angles: φ, representing the angular deviation between the horizontal plane and the long axis of the pore channel (ranging from 0° to 90°), and θ, representing the azimuthal orientation of the long axis on the horizontal plane (ranging from 0° to 180°). Furthermore, critical metrics such as pore space surface area and sphericity were directly ascertained through the utilization of ImageJ. Integral to the analysis was the calculation of porosity, which manifested as the fraction of image volume characterized by pore space.

ML procedure

Exploring ML models

Acknowledging the potential of ML models to unveil complex data patterns, these models were employed to unravel intricate relationships within the acquired soil dataset. However, it was recognized that the effectiveness of these models depended on the quality, quantity, and representativeness of the training dataset. Table 3 provides a comprehensive overview of the ML algorithms employed in this study for the prediction of SWCC. Each algorithm is described along with its key hyperparameters, strengths, and limitations.

Orange.3

Ahangar-Asr et al. (2012) emphasized that the simplicity of a procedure and its capability to apply multiple models simultaneously are key factors in determining the priority of a method for estimating SWCC. In line with this, we utilized Orange.3 software, which offers a user-friendly and efficient ML process. This approach facilitated a rapid comparison of diverse fitted models, encompassing Gradient Boosting, Ada Boost, Decision Tree, Random Forest, Neural Network, Support Vector Machine, k-Nearest Neighbors, and Linear Regression. Furthermore, the Feature Importance widget

Table 3 Overview of the utilized machine learning algorithms to predict the Soil Water Characteristic Curve

Machine learning algorithm	Description	Key hyperparameters	Strengths	Limitations	Citations
K-nearest neighbor (kNN)	A non-parametric algorithm known for its performance across data distributions	Number of nearest neighbors, Leaf size	Good performance, Data distribution agnostic	Computationally expensive	Guo et al. 2003
Support vector machine (SVM)	A versatile non-parametric algorithm	Regularization parameters (C and γ)	High-dimensional data compatibility	Resource-demanding	Cortes and Vapnik 1995
Decision tree (DT)	A white-box model providing interpretable insights	max_depth, max_leaf_nodes	Interpretable, No specific data preprocessing	Overfitting risk, Suboptimal results	Myles et al. 2004
Neural network (NN)	NN learns patterns through training by optimizing an objective function	Number of hidden layers, Nodes	model complex non-linear multivariate problems with the advantages of learning capacity and memory simulation	Challenging structure tuning	Zhang et al. 2020; Abiodun et al. 2018
RandomForest (RF)	A ensemble model aggregating DTs through bootstrap sampling	max_depth, max_leaf_nodes, Number of DTs (n_number)	Homogeneous ensemble, High predictability	Variable model complexity	Belgiu and Drăguţ 2016
Gradient boosting (GB)	GB is a sequential learning method chaining weak learners (DTs) for superior results	Regularization parameters, Distributed weighted quantile sketch algorithm	State-of-the-art results, Sequential learning	Overfitting, Computational cost	Natekin and Knoll 2013
AdaBoost (AB)	AB is an ensemble method enhancing model performance by iteratively reweighting misclassified samples	Number of weak learners, Learning rate	Improved generalization, Versatility	Sensitive to noisy data	Freund and Schapire 1996
Linear regression (LR)	LR models the linear relationship between variables	None	Interpretable, Easy to implement	Limited to linear relationships	N/A

was used to determine the relative importance of input features in predicting SWCC with a minimal dataset.

Statistical analysis

The statistical analysis included the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Relative Root Mean Squared Error (RRMSE), the Pearson correlation coefficient (R), Performance Index (P_i, Jalal et al. 2021), and the Willmott's index of agreement (d1, Zhang et al. 2020; Achieng 2019) collectively quantify the predictive performance and reliability of the models. These metrics provide a comprehensive view of how well the algorithms capture the complex relationships inherent in SWCC data.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(o_i - t_i)^2}{n}} \tag{1}$$

$$MAE = \frac{\sum_{i=1}^{n}|o_i - t_i|}{n} \tag{2}$$

$$RRMSE = \frac{1}{|\bar{o}|}\sqrt{\frac{\sum_{i=1}^{n}(o_i - t_i)^2}{n}} \tag{3}$$

$$R = \frac{\sum_{i=1}^{n}(o_i - \bar{o})(t_i - \bar{t})}{\sqrt{\sum_{i=1}^{n}(o_i - \bar{o})^2\sum_{i=1}^{n}(t_i - \bar{t})^2}} \tag{4}$$

$$P_i = \frac{RRMSE}{1 + R} \tag{5}$$

$$d1 = 1 - \frac{\sum_{i=1}^{n}|t_i - o_i|}{|o_i - \bar{o}| + |t_i - \bar{t}|} \tag{6}$$

Where o_i and t_i represent the i-th actual and predicted output values, respectively. $\bar{o}$ and $\bar{t}$ denote the average values of the actual and predicted output values, respectively. The parameter n signifies the number of samples under consideration, and it is worth noting that in our analysis, no parameters were excluded during the regression process.

Feature importance analysis

As suggested by Pham et al. (2023), the analysis of feature importance employed the Shapley additive explanations (SHAP) technique, which allowed for a quantitative assessment of the significance of the selected features. This methodology revolves around the assessment of model variance using a variance-based measurement approach (Molnar 2020). Notably, the geotechnical field has widely adopted the Shapley value approach (Wadoux and Molnar 2022; Cheng et al. 2022) to unravel the intricacies of feature importance. The Shapley value (φ) embodies the average marginal contributions of features within various coalitions, as encapsulated by the equation:

$$\varphi_j = \sum_{S \subseteq X\{xj\}} \frac{|S|!(p - |S| - 1)!}{p!}(val(S \cup \{xj\}) - val(S)) \tag{7}$$

$$\sum_{i=1}^{p} \varphi_j \hat{f}(X) - E_x(\hat{f}) \tag{8}$$

Where S represents a subset of features, $X = \{X1, X2, \ldots, Xp\}$ designates the observed data point, x_j denotes the value of the feature under consideration, p signifies the total count of features, $val(S)$ reflects the model prediction marginalized over the remaining input features, $\hat{f}(X)$ stands for the model prediction for the given X, and $E_X(\hat{f})$ represents the anticipated model predictions for a given dataset. This comprehensive approach elucidates the intricate dynamics of feature importance and model contributions. All computations and feature importance analyses in this study were carried out automatically using Orange.3 software.

Results and discussion

The properties of initial soil samples

Table 2 represented the routine properties parameters obtained from the SWCCs of soil samples prior to any treatment. The selection of these two samples was done deliberately to ensure a wide range of variations in their physical, chemical, and hydraulic properties, allowing for a comprehensive evaluation. The loamy sand sample has a high sand content of 82.6% with approximately 12% clay, while the silty clay sample has a clay content exceeding 40% and a lower sand content of around 10%. Both samples are non-saline and slightly alkaline, but they differ significantly in terms of organic carbon (OC) content (0.12% vs. 0.42%) and cation exchange capacity (CEC) values (4.8 $cmol^+ kg^{-1}$ vs. 24.1 $cmol^+ kg^{-1}$). The matric suction at the inflection point (h_i) of the SWCC varies from 300 cm in the loamy sand sample to 70 cm in the silty clay sample. The shape factor (n) of the SWCC in Van Genuchten's (1980) model ranges from 2.07 in the loamy sand sample to 1.0 in the silty clay sample. The bulk density of the studied samples did not show significant differences. However, there were significant differences in the alpha coefficient, which corresponds to the inverse value of air entry into the soil (α, cm^{-1}), as well as in the saturation water content (θ_s, $g.g^{-1}$) and residual water content (θ_r, $g.g^{-1}$) between the two samples.

Changes in properties of treated samples and the results obtained from image analyses

Table 4 presents the changes in the physical and chemical properties of the treated samples after the incubation period, compared to the blank samples. Additionally, Table 2 provides the results from image analyses of the soil pores developed as a result of the treatments. The analysis of the trends in Table 4 reveals noteworthy patterns in the impact of treatments, treatment levels, and soil textures on various soil properties. Generally, there's a tendency for increasing bulk density with higher treatment levels, particularly prominent in "$CaCO_3$" and "OM" treatments. "Cations" treatments consistently lead to elevated CEC with increasing levels, while "Removed Fe_2O_3" and "Removed OM" exhibit reduced CEC after removal. "$CaCO_3$" treatments show increased electrical conductivity (EC) with higher levels, and "OM" treatments correlate with higher organic carbon (OC) content. Values of pH decreases with higher levels in "Cations" treatments, and "Removed $CaCO_3$" results in decreased pH after removal. Porosity-related attributes are influenced by treatments and levels, with "OM" treatments consistently yielding higher porosity surface area and porosity volume. Removal and degradation treatments lead to various property changes, such as decreased OC content, CEC, and porosity-related attributes, highlighting the complexity of soil responses to alterations. Overall, these trends provide valuable insights into the intricate relationships between treatments, soil properties, and textures.

Similar to Table 4 and Table 1, a dataset of individual treatments was prepared, which was automatically divided into model training and test datasets. This practice of preparing a unified database, as advocated by Wang et al. (2022) and Zhang et al. (2020), is essential in ML procedures to ensure consistency and facilitate the training, testing, evaluation, and comparison of different models, thereby yielding robust and reliable results. Then, the mentioned features from Table 2 and Table 1 applied in eight algorithms to predict the soil water content at different matric suction levels. Soil matric suction is used as a predefined input feature, while the other features are applied separately in all evaluating models. The most important features are determined based on their effects on the model output, as shown in Fig. 1.

Impacts and relative importance of the input parameters on the models

Researchers have utilized various soil properties, including the percentages of clay, silt, and sand, as well as void ratio and water content at saturation, along with soil matric suction related to gravimetric water content, for the estimation of SWCC (Pham et al. 2023; Rastgou et al. 2020). Identifying the most significant features in SWCC estimation can greatly reduce time and energy consumption while increasing accuracy. As input features of models Fig. 1 (1.a to 1.h) illustrates the effects of different input parameters on model outputs and their relative importance in terms of the model's accuracies (RMSEs) in eight ML algorithms. Similar to studies conducted in previous years (Pham et al. 2023; Gunarathna et al. 2019; Nguyen et al. 2017), our observations indicate that within these algorithms, matric suction emerged as the most pivotal parameter in the GB (Fig. 1a), AB (Fig. 1b), RF (Fig. 1d), and SVM (Fig. 1f) models. On the other hand, organic carbon percentage, soil texture, porosity surface area, and electrical conductivity emerged as the most significant parameters in the DT (Fig. 1c), ANN (Fig. 1e), kNN (Fig. 1g), and LR (Fig. 1h) models, respectively. Matric suction was identified as the most important parameter among the first three influential parameters affecting the model outputs in all models, except for the ANN model (Fig. 1e). Lower matric suction values resulted in higher prediction accuracy in the models, while higher matric suction values led to a decrease in accuracy. The results indicated that, except for the ANN model, three to five of the input characteristics were identified as the most influential parameters for prediction accuracy in different models.

Following matric suction, soil pore characteristics have emerged as the subsequent significant parameters in facilitating accurate predictions, except in the context of the ANN model. Pham et al. (2023) demonstrated that soil texture-related properties hold significant importance following soil matric suction. However, in our study, the prediction of SWCC reveals the involvement of one or two pore characteristics. Notably, attributes like structural flatness and porosity surface area exhibit notably stronger influence compared to other pore characteristics. Soil bulk density, as a other structural feature, has garnered attention from various researchers in recent years (Amanabadi et al. 2019; Gunarathna et al. 2019). However, this soil physical property serves as an average indicator of soil compaction and fails to provide insights into the detailed attributes of porosity. Some studies have endeavored to indirectly assess soil structure by incorporating soil moisture content across different matric suctions into their models (Senyurek et al. 2020; Cai et al. 2019). In one of the rare instances exploring soil porosity, Ahangar-Asr et al. (2012) integrated soil void ratio as an input parameter within a model geared towards SWCC and soil porosity characteristic prediction. Nevertheless, their investigation did not specifically delve into the impact of these properties on the outcomes of the model.

Comparison of the models' predicted results

The output of the models when all parameters used

When comparing the SWCCs generated by the models using all the studied parameters, it was found that the GB, AB,

Table 4 Properties of treatments at the end of incubation and the results obtained from image analysis

Treatments	Level	Bulk Density (kg.m^{-3})	CEC (cmol^{+}. kg^{-1})	EC (µS.cm^{-1})	OC (%)	pH	Porosity surface area (cm^{2})	Porosity Volume (cm^{3})	Structural flatness (-)	Porosity elongation (-)	Pore sphericity (-)	Porosity Compactness (-)
Loamy sand												
	Blank	1310	5.24	877	0.14	8.12	1.240	1.2×10^{-1}	4836	1.44	0.08	0.08
$CaCO_3$	1	1380	5.48	1091	0.16	7.96	1.100	1.4×10^{-1}	4115	3.5	0.21	0.13
	2	1480	5.77	1307	0.15	8.2	3.830	1.2×10^{-1}	5421	2.3	0.11	0.17
	3	1400	5.45	1030	0.16	8.05	3.430	1.1×10^{-1}	5127	2.1	0.1	0.2
Fe_2O_3	1	1390	5.36	1027	0.14	7.87	1.180	0.9×10^{-1}	4826	1.7	0.09	0.1
	2	1520	5.7	994	0.15	7.94	1.250	1.1×10^{-1}	5215	2.0	0.07	0.09
	3	1290	5.56	1076	0.14	8.06	1.430	1.3×10^{-1}	5625	1.6	0.06	0.1
OM	1	1340	5.41	915	0.17	8.21	1.070	1.1×10^{-1}	4023	4.8	0.29	0.08
	2	1440	5.10	947	0.17	8.1	1.200	1.3×10^{-1}	3745	5.6	0.33	0.07
	3	1500	5.96	917	0.18	8.07	1.000	1.0×10^{-1}	4715	3.7	0.22	0.11
Cations	1	1330	5.42	1124	0.15	7.83	1.010	1.0×10^{-1}	3965	4.9	0.14	0.11
	2	1360	5.24	1173	0.15	7.96	1.110	1.0×10^{-1}	3928	5.0	0.21	0.09
	3	1470	5.31	1271	0.14	7.86	1.030	1.0×10^{-1}	4236	4.7	0.16	0.1
Combined	1	1370	5.57	1210	0.16	8.11	4.730	1.4×10^{-1}	4830	4.4	0.13	0.2
	2	1310	5.49	1240	0.15	8.07	3.658	1.3×10^{-1}	5022	4.2	0.21	0.2
	3	1250	5.33	1320	0.17	7.91	3.354	1.4×10^{-1}	5266	4.4	0.17	0.1
Removed $CaCo_3$	–	1410	5.19	948	0.13	7.62	1.270	1.4×10^{-1}	3215	6.0	0.37	0.06
Removed Fe_2O_3	–	1270	5.34	982	0.13	8.11	0.965	1.0×10^{-1}	5118	3.1	0.17	0.13
Removed OM	–	1460	5.29	1031	0.12	7.98	0.925	1.0×10^{-1}	5084	4.2	0.28	0.1
Degraded Treatment	–	1620	5.41	916	0.14	8.16	1.000	1.0×10^{-1}	4425	1.4	0.11	0.05
Silty clay												
	Blank	1290	25.95	492	0.47	7.62	1.280	1.2×10^{-1}	5987	4.8	0.41	0.2
$CaCO_3$	1	1280	26.87	687	0.5	7.54	1.050	1.1×10^{-1}	2549	4.7	0.32	0.19
	2	1220	27.12	841	0.49	7.78	0.986	1.1×10^{-1}	2989	3.2	0.29	0.23
	3	1180	25.9	945	0.51	7.62	1.010	1.0×10^{-1}	4216	3.5	0.31	0.2
Fe_2O_3	1	1250	23.7	646	0.44	7.56	1.100	1.3×10^{-1}	6395	3.1	0.37	0.17
	2	1280	24.1	633	0.47	7.71	1.170	1.2×10^{-1}	4987	3.7	0.26	0.22
	3	1310	23.7	657	0.54	7.64	1.110	1.1×10^{-1}	4529	4.0	0.29	0.17
OM	1	1260	26.6	704	0.58	7.73	1.020	9.5×10^{-2}	3825	4.4	0.32	0.2
	2	1220	27.9	647	0.61	7.66	1.420	1.4×10^{-1}	2744	6.3	0.37	0.09
	3	1280	29.1	686	0.61	7.84	1.530	1.5×10^{-1}	4988	7.3	0.46	0.06
Cations	1	1320	26.4	724	0.49	7.76	1.340	1.2×10^{-1}	4074	5.1	0.43	0.14
	2	1440	25.6	693	0.47	7.82	1.090	1.1×10^{-1}	4765	3.6	0.35	0.19
	3	1190	24.9	686	0.5	7.72	1.220	1.1×10^{-1}	3956	4.2	0.38	0.16

Table 4 (continued)

Treatments	Level	Bulk Density (kg.m^{-3})	CEC (cmol$^+$.kg^{-1})	EC (µS.cm^{-1})	OC (%)	pH	Porosity surface area (cm^2)	Porosity Volume (cm^3)	Structural flatness (-)	Porosity elongation (-)	Pore sphericity (-)	Porosity Compactness (-)
Combined	1	1370	23.8	712	0.51	7.86	0.922	9.2×10^{-2}	3019	3.2	0.26	0.24
	2	1190	24.3	734	0.53	7.90	1.100	1.0×10^{-1}	2552	3.8	0.33	0.17
	3	1160	25.4	746	0.52	7.8	1.050	1.0×10^{-1}	3118	3.5	0.3	0.2
Removed $CaCo_3$	–	1210	23.2	634	0.47	7.73	1.200	1.5×10^{-1}	4512	5.7	0.22	0.11
Removed Fe_2O_3	–	1210	21.7	706	0.49	7.84	0.976	1.0×10^{-1}	4311	4.2	0.24	0.17
Removed OM	–	1280	23.9	691	0.48	7.93	0.942	9.8×10^{-2}	6236	5.0	0.27	0.14
Degraded Treatment	–	1550	23.4	669	0.46	7.72	0.903	8.7×10^{-2}	3721	2.7	0.31	0.26

RF, and DT models produced the most accurate results with lower RMSE (<0.028) and MAE (<0.018), and higher d1 (>0.93) and R^2 (>0.968) in test dataset (TstD), as shown in Table 5. This means that the mean difference between the predicted and measured water contents was less than 0.02 g g^{-1} for all matric suctions used to plot the SWCCs. Achieng (2019) conducted research using ML techniques, including ANN, DNN, and SVM models, to estimate SWCC. In most cases of drying SWCC, the models achieved an RMSE of less than 0.01, with R^2 and d1 values exceeding 0.99 and 0.94, respectively. The study demonstrated high accuracy in the estimation of SWCC in the studied Loamy Sand soil sample. Lamorski et al. (2017) employed various SVM models trained with physical soil properties, including SWCC drying branch, BD, Sand%, Silt%, clay%, OC, and soil specific surface, as input variables. The resulting models successfully estimated SWCC wetting branches with an R^2 greater than 0.98 and an RMSE less than 0.02. Srivastava et al. (2013) utilized the SVM algorithm, which yielded an RMSE of 0.013 and an R^2 of 0.69. In contrast, the performance of the random forest algorithm varied across different studies. Long et al. (2019) and lm et al. (2016) reported RMSE values greater than 0.04 m^3 m^{-3}, while Bai et al. (2019) achieved accurate results with an RMSE less than 0.02 m^3 m^{-3}. However, in this study ANN, SVM, kNN, and LR algorithms, showed a significant decrease in model accuracy (as indicated by higher values of RMSE, MAE, and lower values of d1 and R^2) compared to the acceptable limits of accuracy. Consequently, these models were unable to generate SWCCs that met the required level of accuracy. Similar to the findings of Hastie et al. (2009), which demonstrated that regression-based methods may yield non-accurate results in pedo-transfer function methods, the LR algorithm in this study produced an R^2 of 0.66 and an RMSE of 0.69 when applied in the ML method, categorizing it as a non-accurate model. Nguyen et al. 2017 highlighted the benefits of the kNN model, including its flexibility, simplicity, accuracy in limited data availability conditions, and the ability to incorporate new observations into training datasets without the need to redevelop the PTF models. However, Guevara and Vargas (2019) examined the performance of the kNN algorithm for predicting soil moisture content based on DEM data and found that the prediction RMSE exceeded 0.05 m^3 m^{-3}. In another study, Liu et al. (2017) observed an RMSE greater than 0.07 m^3 m^{-3} in the prediction of moisture content using the kNN algorithm with inputs derived from satellite-derived data.

Table 6 presents the Pearson correlation (r) between the measured water content ($\theta_{Measured}$) and the evaluating models, along with the identified important features. Previous studies have reported correlation coefficients greater than 0.9 between estimated and measured SWCC or soil moisture content using the random forest algorithm (Im et al.

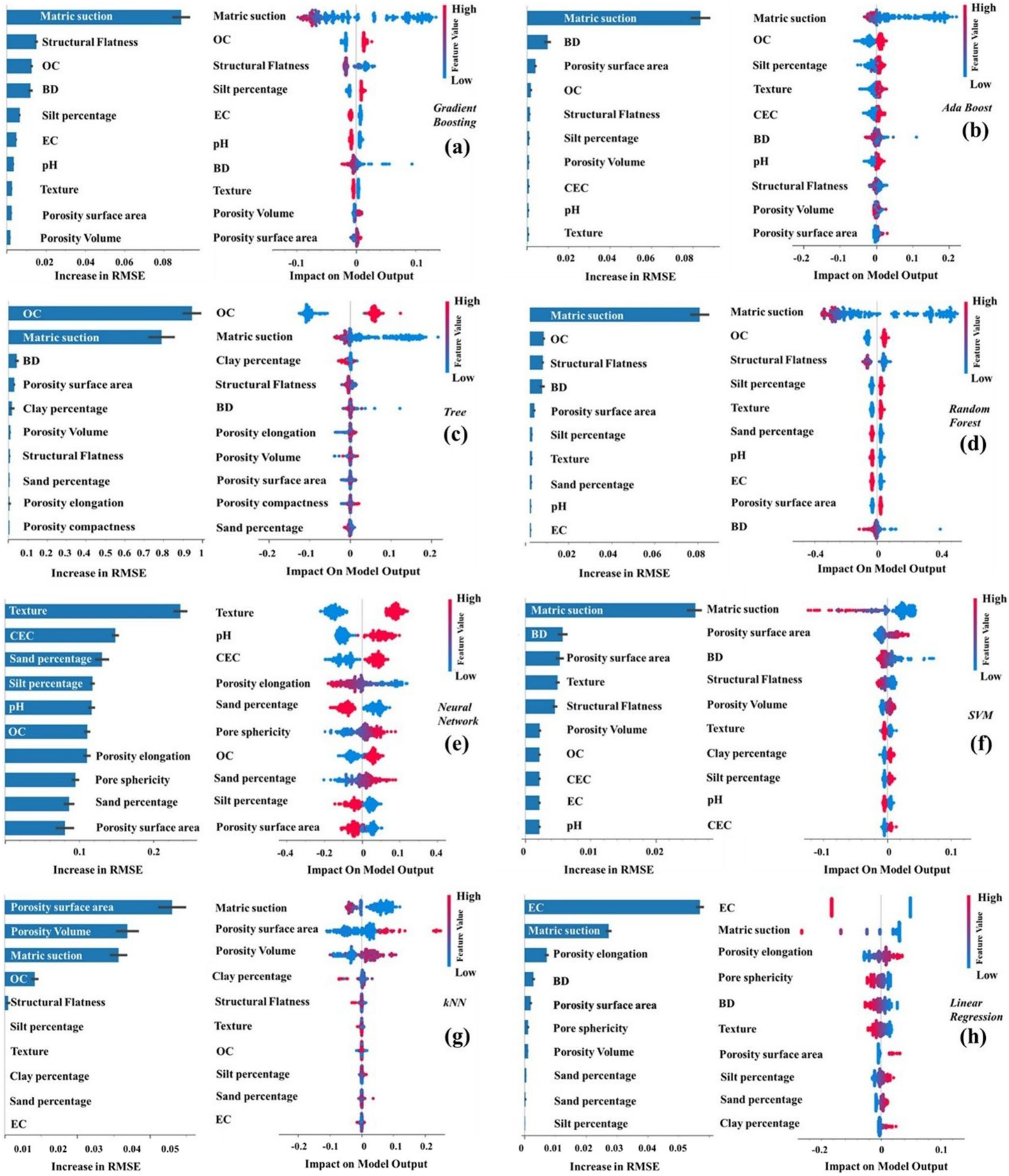

Fig. 1 Input parameters and their relative importance in accurate prediction of GB, AB, RF, SVM, DT, ANN, kNN, and LR models

2016; Bai et al. 2019; Long et al. 2019; Zappa et al. 2019). However, it is important to note that the ability of the same algorithm to estimate soil moisture content may vary depending on the input features used in the modeling procedure. For example, the aforementioned studies utilized different sets of input features, including satellite-derived

Table 5 The statistics obtained for the models used to generate SWCC using all parameters

Model	RMSE		MAE		d1		R^2		RRMSE		R		P_i	
TrD/TstD*	TrD	TstD	TrD	TstD	TrD	TstD	TrD	TstD	TrD	TstD	TrD	TstD	TrD	TstD
Gradient Boosting	0.014	0.016	0.009	0.011	0.97	0.94	0.994	0.982	0.04	0.05	0.997	0.991	0.02	0.03
AB	0.019	0.020	0.012	0.014	0.94	0.93	0.981	0.971	0.05	0.06	0.99	0.985	0.03	0.03
Random Forest	0.023	0.021	0.014	0.015	0.91	0.93	0.982	0.968	0.07	0.06	0.991	0.984	0.04	0.03
DT	0.019	0.028	0.013	0.018	0.97	0.93	0.941	0.944	0.05	0.08	0.97	0.972	0.03	0.04
Neural Network	0.064	0.087	0.055	0.071	0.67	0.65	0.434	0.462	0.18	0.25	0.659	0.68	0.11	0.15
SVM	0.061	0.077	0.063	0.070	0.61	0.58	0.659	0.583	0.17	0.22	0.812	0.764	0.09	0.12
kNN	0.071	0.085	0.066	0.071	0.67	0.62	0.612	0.490	0.2	0.24	0.782	0.7	0.11	0.14
Linear Regression	0.066	0.069	0.053	0.059	0.74	0.69	0.592	0.664	0.19	0.2	0.769	0.815	0.11	0.11

* TrD: Train and Test Dataset, respectively

Table 6 Pearson correlation (r) between model and used features with measured water content

Parameters	θ_{Measured}
$\theta_{\text{Gradient Boosting}}$	0.991
θ_{AB}	0.985
$\theta_{\text{Random Forest}}$	0.984
θ_{Tree}	0.972
$\theta_{\text{Linear regression}}$	0.817
θ_{SVM}	0.802
$\theta_{\text{Neural Network}}$	0.743
θ_{kNN}	0.710
Matric Suction	−0.704
EC	−0.640
OC	0.640
CEC	0.640
pH	−0.640
Silt percentage	0.640
Sand percentage	−0.640
Clay percentage	0.640
Structural Flatness	−0.625
Pore Sphericity	0.558
Porosity Elongation	0.547
Porosity Volume	0.536
Bulk Density	−0.469
Porosity Compactness	0.438
Porosity Surface Area	0.335

data, soil texture (Zappa et al. 2019), and leaf area index (Im et al. 2016). These variations in input features can result in different levels of correlation with the target values. As illustrated in Fig. 1 and further supported by Table 6, certain features exhibit a stronger correlation with the measured soil moisture content. Notably, matric suction has shown a strong negative correlation with θ_{Measured}, indicating its influence on soil moisture dynamics.

The reduction in soil pore size distribution resulting from increased soil compaction leads to elevated matric suction across all soil texture classes (Fredlund and Rahardjo 1993). Thus, soil bulk density and sand percentage exhibit a negative correlation with soil water content. Additionally, a negative correlation was observed between water content and structural flatness, indicating that increased soil pore compaction leads to a decrease in water content at varying matric suction levels. Notably, based on Pearson correlation coefficients, structural flatness ($r = -0.625$) demonstrates a more explicit effect on the decrease of soil water content compared to soil bulk density ($r = -0.469$).

Just appling soil matric suction as model input feature

To assess the necessity of incorporating additional input features for improving the model outputs, an evaluation conducted using only the matric suction feature as the input. While soil matric suction has a significant impact on model learning and prediction accuracy, the results presented in Table 7 demonstrate that models trained solely using matric suction and related water content data did not achieve acceptable precision. The models exhibited high error rates and low R^2 values when tested on the dataset. These findings indicate the need for additional input features to improve the accuracy and reliability of the models.

Despite the negative correlation observed between soil water content and matric suction in the evaluating models (Table 8), the calculated RMSE values revealed relatively high errors in the model outputs. The mean absolute errors further indicated significant inaccuracies in the prediction of soil water contents at different matric suction levels, with values ranging from 0.08 to 0.09. Such errors are far from acceptable in this context. Moreover, the considerably low values of R^2 highlight the inconsistency between the predicted SWCC patterns and the observed data.

The use of soil matric suction as the sole input feature in the eight evaluating models significantly reduces the

Table 7 Statistics of models in the case where matric suction was used as the only input parameter

Model	RMSE		MAE		d1		R^2		RRMSE		R		P_i	
TrD/TstD*	TrD	TstD	TrD	TstD	TrD	TstD	TrD	TstD	TrD	TstD	TrD	TstD	TrD	TstD
Gradient Boosting	0.096	0.093	0.091	0.083	0.61	0.52	0.523	0.391	0.27	0.27	0.723	0.625	0.16	0.17
AB	0.084	0.093	0.088	0.083	0.59	0.50	0.465	0.386	0.24	0.27	0.682	0.621	0.14	0.17
Random Forest	0.071	0.093	0.096	0.082	0.60	0.51	0.471	0.389	0.2	0.27	0.686	0.624	0.12	0.17
Tree	0.084	0.093	0.092	0.083	0.65	0.52	0.504	0.391	0.24	0.27	0.71	0.625	0.14	0.17
Neural Network	0.096	0.104	0.088	0.083	0.58	0.46	0.416	0.229	0.27	0.3	0.645	0.479	0.16	0.2
SVM	0.093	0.095	0.091	0.083	0.56	0.46	0.432	0.356	0.27	0.27	0.657	0.597	0.16	0.17
kNN	0.101	0.097	0.102	0.087	0.54	0.53	0.380	0.332	0.29	0.28	0.616	0.576	0.18	0.18
Linear Regression	0.124	0.109	0.101	0.091	0.39	0.33	0.294	0.154	0.35	0.31	0.542	0.392	0.23	0.22

* TrD: Train and Test Dataset, respectively

Table 8 Pearson correlation (r) when the matric suction is included as the only modeling parameter

Parameters	$\theta_{Measured}$
$\theta_{Gradient\ Boosting}$	0.627
θ_{Tree}	0.627
$\theta_{Random\ Forest}$	0.624
θ_{AB}	0.623
θ_{SVM}	0.608
θ_{kNN}	0.588
$\theta_{Neural\ Network}$	0.493
Matric Suction	-0.404
$\theta_{Linear\ regression}$	0.392

correlation between the models and the measured water content ($\theta_{Measured}$). This, in turn, causes the correlation of the linear regression model with $\theta_{Measured}$ to be lower than the correlations between matric suction and $\theta_{Measured}$ (as shown in Table 8). Based on these findings, it can be concluded that utilizing matric suction values alone in the prediction of the SWCC yields better results compared to using the Linear regression model with only matric suction values. This observation suggests that in this case the modeling process was not effective and did not produce useful outcomes. It's worth noting that Zhang et al. (2020) also observed differences in ML procedure capacity based on the number of utilized parameters, reinforcing the importance of considering parameter selection in prediction of soil thermal conductivity.

Predicted SWCCs with evaluating models based on the ML procedure

The assessment of predictive accuracy for eight ML algorithms, was undertaken by comparing their estimated results to the actual measurements. This evaluation is visualized in Fig. 2, where eight individual curves (labeled from "a" to "h") depict the performance of each algorithm. These curves provide a comprehensive representation of how well the algorithms align with the actual measurements. Notably, the 1:1 line in each segment serves as a reference for perfect agreement between predictions and measurements. Among these algorithms, Gradient Boosting (GB) showcases its remarkable predictive capabilities, reflecting its potential to closely replicate the actual SWCC.

Figures 3 and 4 illustrate the SWCC for Loamy Sand and Silty clay soil samples, respectively. As mentioned earlier, the evaluating models can be categorized into two classes based on their prediction accuracy: high and low. In Figs. 3 and 4, these differences explicitly demonstrated. Specifically, for the Loamy Sand soil sample, Gradient Boosting, Ada Boost, Tree, and Random forest models (Fig. 3a–d) exhibited almost perfect predictions of SWCC. While the high accuracy prediction of the SWCC is consistent in Silty Clay soil samples, it is worth noting that for soil matric suctions higher than 1000 cm, the error of the mentioned models shows a relatively decreased trend. Previous studies have highlighted the flexibility and reliability of ML algorithms such as ANN, kNN, and SVM in providing accurate estimations, as they do not rely on stringent assumptions about the underlying data and can adapt to various situations (Nguyen et al. 2017; Hastie et al. 2009). However, in the present study, the performance of the Neural Network, SVM, kNN, and Linear Regression models in predicting SWCC for both Sandy Loam and Silty Clay soil samples yielded errors that were deemed non-acceptable. Specific details regarding the nature and magnitude of these errors would provide further insights into the limitations of these models in the context of the study. These errors resulted in deviations between the predicted SWCC patterns and the measured SWCC pattern across the entire range of matric suctions (Figs. 3 and 4e ~ h). Specifically, the models showed underestimation at low matric suction and overestimation at high matric suction for all studied soil samples. The SVM and kNN models fail to exhibit the expected decreasing trend with respect to matric suction in the Loamy Sand sample, rendering them unable to adequately explain the SWCC.

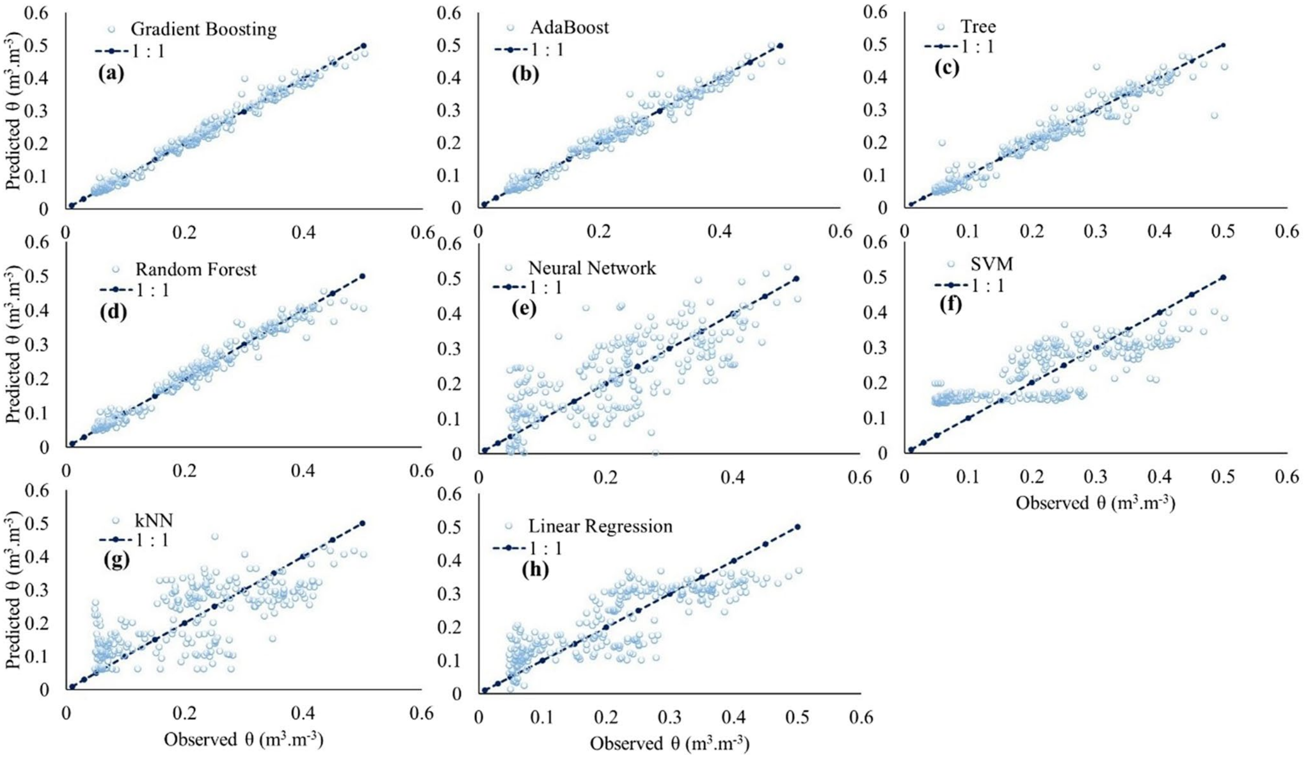

Fig. 2 Comparison of predicted and observed SWCC around the 1:1 line

Similarly, the kNN model yields inaccurate outputs for the Silty Clay soil sample.

Evaluating models uncertainty

Table 9 presents the error percentages to quantify the mean differences between the observed and predicted SWCCs in both the Loamy Sand and Silty Clay soil samples. These error percentages provide insights into the uncertainty associated with each evaluating model. Although Wang et al. (2021) demonstrated high accuracy in determining SWCC for soils with a high clay fraction, this study found that the average error of the eight models used for Loamy Sand soil samples was considerably higher at 35% compared to Silty Clay Soil samples. However, the four well-predicted models, namely Gradient Boosting, AB, Tree, and Random Forest models, exhibited an equal average error percentage of approximately 5% in both Loamy Sand and Silty Clay soil samples, and no significant difference in the estimation of SWCC was observed between the two studied soil textures. The Gradient Boosting model demonstrated superior prediction capability in both studied soil textures, and it exhibited the lowest error percentage in Loamy sand soil samples, with an average uncertainty of 2.7%. The other evaluating models, such as Neural Network, SVM, kNN, and Linear regression, exhibited unreliable outputs with error percentages exceeding 20%. In particular, the SVM model performed poorly in Loamy Sand soil samples, reaching approximately 90% errors. Interestingly, these models showed comparatively better prediction performance in Silty Clay soil samples compared to Loamy Sand soil.

Prediction errors at two sides of the inflection point (h_i)

Some researchers have observed that their models underestimated the water content of the SWCC at relatively high suction heads (Nguyen et al. 2017; Hwang and Powers 2003; Meskini-Vishkaee et al. 2014; Mohammadi and Meskini-Vishkaee 2012; Tuller and Or 2001; Tuller et al. 1999). Nguyen et al. (2017) attributed the underestimation of SWCC to the lack of measurement of input features at high matric suction situations. Other studies have shown the existence of corner water, lens water, and film water in soils, which may be one of the main causes of the underestimation phenomenon (Mohammadi and Meskini-Vishkaee 2012; Or and Tuller 1999; Shahraeeni and Or 2010; Tuller and Or 2005; Tuller et al. 1999). However, Wang et al. (2021) claim that their improved prediction model can effectively predict soil–water characteristic curves, especially for soils at high matric suctions, in contrast, in this study, we observed visual evidence of increasing model errors with higher soil matric

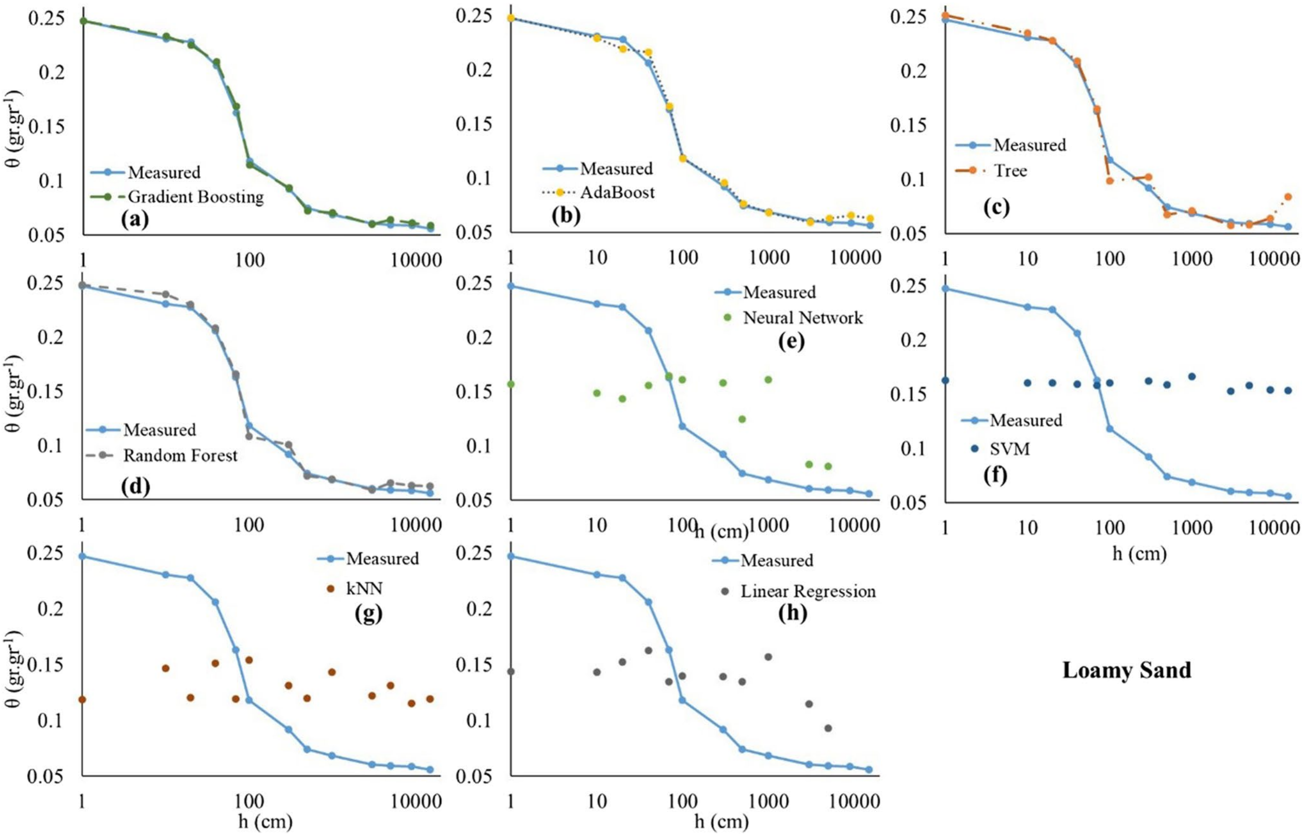

Fig. 3 Comparison of the predicted and measured SWCCs by different models in Loamy Sand soil sample

suction in Figs. 3 and 4, as well as in Table 9. To further support this observation, the error percentages of the evaluating models compared at matric suction values below and above a matric suction related to h_i. For the Loamy Sand soil samples, h_i equal to to 70 cm was calculate, while for the Silty Clay soil samples, h_i was calculated equal to to 300 cm. Figure 5 presents the results of this comparison. In both studied soil textures, the error percentages of all evaluating models are considerably higher at matric suctions greater than h_i compared to matric suctions less than h_i. Among the models, the DT model exhibited the maximum difference between the measured and predicted SWCCs at the two sides of the inflection point. Moreover, the prediction error percentages at matric suctions greater than h_i were found to be ten times higher than those at matric suctions less than h_i. Also, Bakhshi et al. (2023), employing an image analysis approach and substantiating their findings with the Laplace equation (Tuller et al. 2004) for elliptical pores, reported an overestimation of moisture contents at matric suctions exceeding h_i. Additionally, the SVM and kNN models exhibit minimal changes in prediction errors with respect to matric suction. Consequently, there is a minimum difference between the prediction errors of SWCC at the two sides of the inflection point for these models. Based on this concept, the best performance models are identified as those with lower error percentages and a minimal difference in prediction errors at the two sides of the inflection point. Models such as Gradient Boosting, AB, and Random forest exhibit these characteristics.

Residual contents of predicted SWCCs

To quantify the absolute differences between predicted and measured SWCCs, the difference curves for both Loamy Sand and Silty Clay soil samples were presented. Figure 6 depicts the difference curves for Loamy Sand samples, while Fig. 7 displays the difference curves for Silty Clay samples. Each figure includes multiple subfigures (a ~ h) representing different scenarios or conditions within each soil sample. Building upon the previous discussions regarding the high capability of the Gradient Boosting, AB, Tree, and Random forest models, it is evident from Figs. 6 and 7 (subfigures a-d) that these models exhibit minimal fluctuation relative to zero. Furthermore, the other studied models, including Neural Network, SVM, kNN, and Linear regression, demonstrate significant underestimation at low matric suction and overestimation at higher matric suctions, as depicted in Figs. 6 and 7 (subfigures e–h). Similar to the results of this

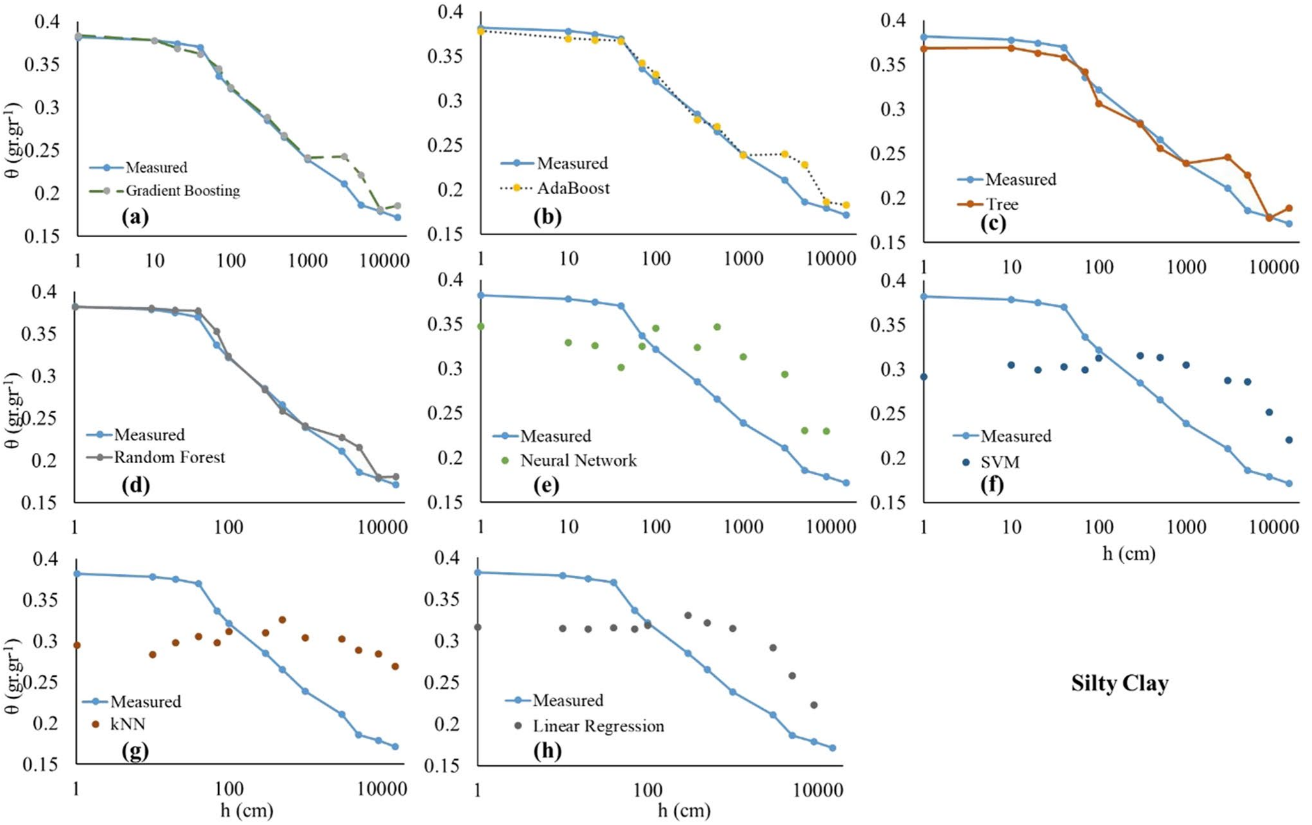

Fig. 4 Comparison of the predicted and measured SWCCs by different models in Silty Clay soil samples

study, Achieng (2019) observed residual SWCC values of about -0.1 to 0.1 g.g-1, but did not find a specific pattern for changes in errors with increasing matric suction. However, as illustrated in Figs. 6f and 7f for both the studied Loamy Sand and Silty Clay soil textures, the highest estimation errors are observed at the two ends of the SWCC. In other words, the SVM model shows the highest error in the estimation of the structural-based and textural-based sections of the SWCC, and around the inflection point, the estimation error of the SVM model diminishes to about zero.

Conclusion

- **Role of informatics in precise estimation of SWCC:** The application of ML techniques has led to the simplification of the intricate process of predicting the SWCC in this study. Additionally, the utilization of Orange.3 data mining software has enabled the incorporation of a wide range of measured physical soil properties into the predictive model, eliminating the requirement for extensive programming knowledge. Thus, through the utilization of informatics principles, we establish a connection between scientific insights and practical engineering applications, thereby facilitating a smoother transition of predictive models into real-world soil-based scenarios.
- **Shortcomings of Solely Matric Suction-based Models:** Our investigation into SWCC prediction revealed a noteworthy limitation. Models constructed solely on the basis of soil matric suction, while seemingly intuitive, exhibited inadequacy in accurately predicting SWCC behavior. This deficiency was evident in the Mean Absolute Error (MAE) exceeding 0.08 and an R-squared (R^2) value below 40% in test dataset.
- **Enhancing Accuracy through Multivariate Approach:** The quest for enhanced prediction accuracy led us to explore a more holistic approach. We embarked on a journey to comprehend the influence of various soil properties on SWCC behavior. Interestingly, our endeavors unveiled the pivotal significance of incorporating soil properties such as bulk density, organic carbon content, and micro-porosity characteristics like flatness or porosity surface area. Integrating these properties as measured features within the model yielded substantial improvements in the precision of SWCC estimation. Statistical analysis revealed that in this scenario, the Gradient Boosting algorithm led to an almost perfect estimation of SWCC, yielding RMSE and P_i values of 0.016 and

Table 9 Uncertainties of evaluating models (error percentage between observed and predicted results) in prediction of soil moisture content at different soil matric suction of Loamy Sand and Silty Clay soil samples

Matric suction	Gradient boosting	AB	Tree	Random forest	Neural network	SVM	kNN	Linear regression
Loamy sand								
0.01	0.0	0.2	1.6	0.4	36.6	34.0	51.9	41.9
10	1.2	0.8	1.9	3.8	35.6	30.5	36.3	37.8
20	1.3	3.8	0.1	1	37.2	29.6	47.2	33.1
40	1.6	4.9	1.5	1.1	24.3	22.7	26.5	21.1
70	3.5	2.1	1.5	1.8	1.1	3.1	26.7	17.3
100	3.1	0.5	16.5	8.6	36.1	35.5	30.4	18.2
300	1.5	3.5	11.1	9.6	71.8	76.2	42.4	51.0
500	2.4	2.2	9.2	3.2	67.3	113.7	61.4	81.0
1000	3.0	0.7	3.6	0.4	133.9	142.1	108.5	128.1
3000	0.4	2.3	5.2	2.4	36.9	153.0	102.6	89.6
5000	8.3	5.6	1.9	10.4	37.2	167.7	121.3	57.3
9000	4.3	11.9	8.9	8	40.5	162.4	96.9	43.4
15,000	4.6	12.2	50.0	11.1	92.5	174.0	113.3	143.2
Silty clay								
0.01	0.5	0.9	3.5	0.1	9.2	23.6	22.9	17.3
10	0.0	2.3	2.5	0.4	13.0	19.4	25.1	16.8
20	1.6	1.7	3.0	0.8	13.2	20.1	20.6	16.2
40	2.0	0.9	3.2	1.9	18.6	18.3	17.4	14.7
70	2.6	1.9	1.8	4.8	3.5	11.1	11.4	6.7
100	0.5	2.8	4.7	0.5	7.3	2.8	3.1	0.8
300	1.2	2.2	0.5	0.5	13.5	10.8	9.0	16.1
500	0.8	1.9	3.6	2.8	30.5	18.0	22.7	21.3
1000	1.0	0.1	0.0	0.8	31.0	27.6	27.0	31.7
3000	15.1	13.7	16.8	7.9	39.4	36.3	43.4	38.3
5000	18.9	22.7	21.2	16	23.8	53.8	55.3	38.8
9000	1.3	4.1	0.9	0.5	28.4	40.6	59.0	24.8
15,000	8.1	6.4	10.0	5.5	39.6	28.4	57.0	19.3

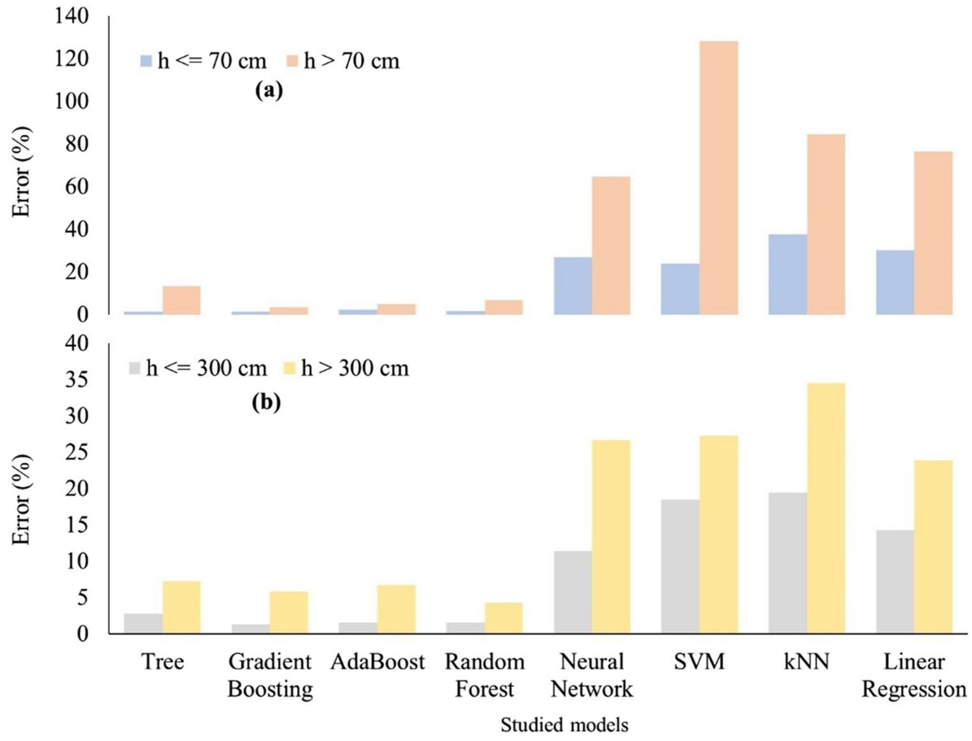

Fig. 5 Comparisons evaluating models error percentages at two side of SWCC inflection point in **a**) Loamy Sand and **b**) Silty Clay soil samples

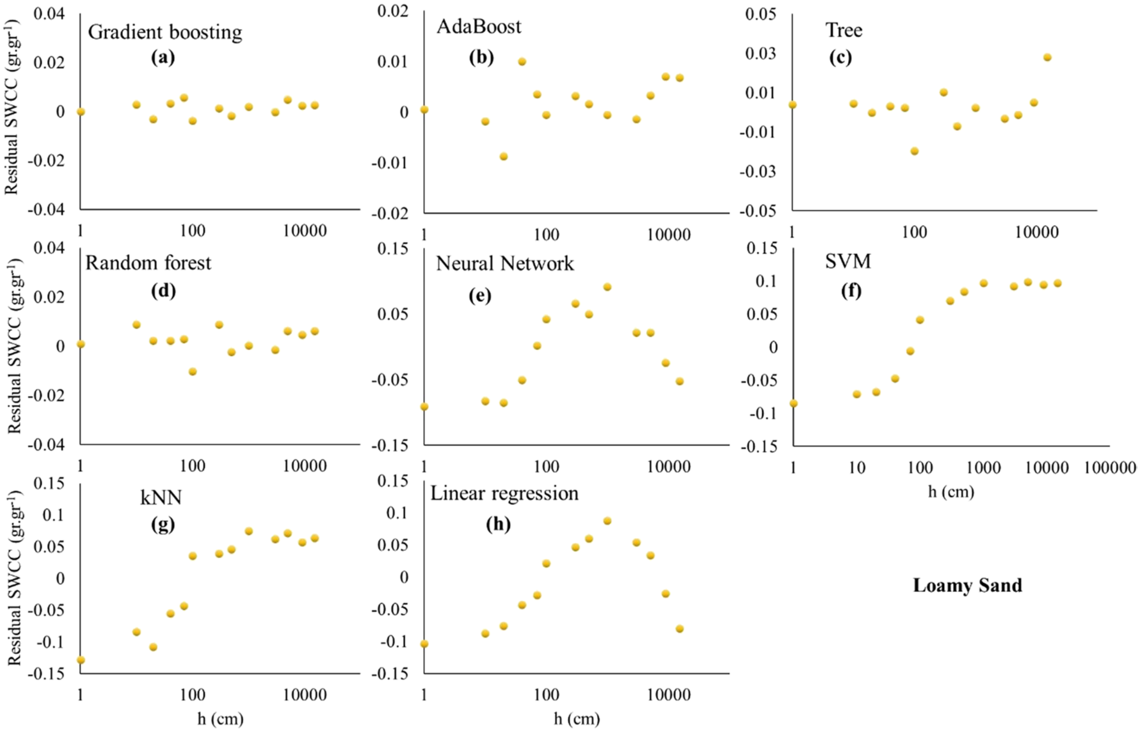

Fig. 6 Absolute difference of prediction and measured SWCC at evaluating models in Loamy Sand soil samples

0.03, respectively. Furthermore, the AB, Random Forest, and Tree models resulted in highly accurate estimations with RMSE and P_i values lower than 0.03 and 0.04, respectively. However, other evaluated models, including Neural Network, SVM, kNN, and Linear Regression, did not exhibit improvement during the training phase, even

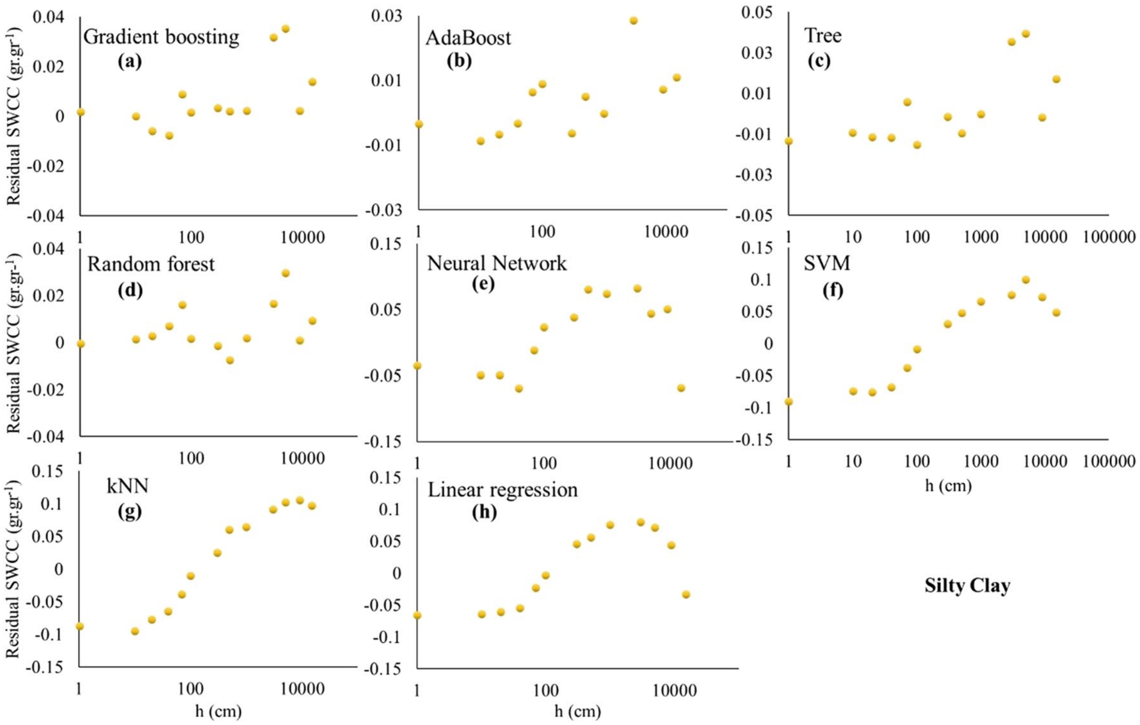

Fig. 7 Absolute difference of prediction and measured SWCC at evaluating models in Silty Clay soil samples

with the inclusion of additional properties of the studied soils.

- **Feature importance analysis:** Among the evaluated models, matric suction stood out as the most critical parameter in the GB, AB, and RF models. Its exclusion from these models resulted in a notable increase in RMSE, reaching up to 0.08. Lower matric suction values correlated with higher accuracy, while higher values reduced accuracy. Following matric suction, soil micro-porosity characteristics gained importance, lowering model RMSE by up to 0.04 in highly accurate models. Notably, structural flatness and porosity surface area played a significant role compared to other pore characteristics in predicting SWCC accurately.
- **Navigating Errors and Achieving Realistic Prediction:** Acknowledging the existence of errors in the estimated SWCC within this study is crucial, especially concerning matric suctions surpassing the SWCC inflection point. These errors were observed in proficiently recognized models, amounting to up to 8 percent in silty clay soils. However, upon analysis, these errors are relatively minor and do not substantially compromise the models' effectiveness in predicting SWCC behavior. This is attributed to the decreasing trend of variations in water content at high matric suctions.

Authors' contributions Aida Bakhshi: made substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data; or the creation of new software used in the work.

Parisa Alamdari: drafted the work or revised it critically for important intellectual content.

Ahmad Heidari: made substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data; or the creation of new software used in the work.

Mohammad Hossein Mohammadi: drafted the work or revised it critically for important intellectual content.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests The authors have no competing interests to declare that are relevant to the content of this manuscript.

References

Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H (2018) State-of-the-art in artificial neural network applications: a survey. Heliyon 4(11). https://doi.org/10.1016/j.heliyon.2018.e00938

Achieng KO (2019) Modelling of soil moisture retention curve using machine learning techniques: artificial and deep neural networks vs support vector regression models. Comput Geosci 133:104320. https://doi.org/10.1016/j.cageo.2019.104320

Ahangar-Asr A, Johari A, Javadi AA (2012) An evolutionary approach to modelling the soil–water characteristic curve in unsaturated soils. Comput Geosci 43:25–33. https://doi.org/10.1016/j.cageo.2012.02.021

Amanabadi S, Vazirinia M, Vereecken H, Vakilian KA, Mohammadi MH (2019) Comparative study of statistical, numerical and machine learning-based pedotransfer functions of water retention curve with particle size distribution data. Eurasian Soil Sci 52:1555–1571. https://doi.org/10.1134/S106422931930001X

Bai J, Cui Q, Zhang W, Meng L (2019) An approach for downscaling SMAP soil moisture by combining Sentinel-1 SAR and MODIS data. Remote Sens 11(23):2736. https://doi.org/10.3390/rs11232736

Bakhshi A, Heidari A, Mohammadi MH, Ghezelbash E (2023) Estimation of water retention at low matric suctions using the micromorphological characteristics of soil pores. Euras Soil Sci 1064–2293. https://doi.org/10.1134/S1064229323600549

Belgiu M, Drăguţ L (2016) Random forest in remote sensing: a review of applications and future directions. ISPRS J Photogramm Remote Sens 114:24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Børgesen CD, Iversen BV, Jacobsen OH, Schaap MG (2008) Pedotransfer functions estimating soil hydraulic properties using different soil parameters. Hydrol Process Int J 22(11):1630–1639. https://doi.org/10.1002/hyp.6731

Cai Y, Zheng W, Zhang X, Zhangzhong L, Xue X (2019) Research on soil moisture prediction model based on deep learning. PLoS ONE 14(4):e0214508. https://doi.org/10.1371/journal.pone.0214508

Cheng Y, Zhou WH, Xu T (2022) Tunneling-induced settlement prediction using the hybrid feature selection method for feature optimization. Transp Geotechn 36:100808. https://doi.org/10.1016/j.trgeo.2022.100808

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20:273–297. https://doi.org/10.1007/BF00994018

Cresswell HP, Green TW, McKenzie NJ (2008) The adequacy of pressure plate apparatus for determining soil water retention. Soil Sci Soc Am J 72(1):41–49. https://doi.org/10.2136/sssaj2006.0182

Dane JH, Hopmans JW, Topp GC (2002) Pressure plate extractor. Methods Soil Anal Part 4:688–690

Demir Z (2020) Alleviation of adverse effects of sodium on soil physicochemical properties by application of vermicompost. Compost Sci Util 28(2):100–116. https://doi.org/10.1080/1065657X.2020.1789011

Dexter AR, Czyż EA, Richard G, Reszkowska A (2008) A user-friendly water retention function that takes account of the textural and structural pore spaces in soil. Geoderma 143(3–4):243–253. https://doi.org/10.1016/j.geoderma.2007.11.010

Diao W, Liu G, Zhang H, Hu K, Jin X (2021) Influences of soil bulk density and texture on estimation of surface soil moisture using spectral feature parameters and an artificial neural network algorithm. Agriculture 11(8):710. https://doi.org/10.3390/agriculture11080710

Dobarco MR, Bourennane H, Arrouays D, Saby NP, Cousin I, Martin MP (2019) Uncertainty assessment of GlobalSoilMap soil available water capacity products: a French case study. Geoderma 344:14–30. https://doi.org/10.1016/j.geoderma.2019.02.036

Eben M, Cithuraj K, Justus S, Bhagavathsingh J (2020) Synthesis and characterization of stretchable IPN polymers from biodegradable resins incorporated with styrene and methyl methacrylate monomers for enhanced mechanical strength. Eur Polym J 138:109957. https://doi.org/10.1016/j.eurpolymj.2020.109957

Fredlund DG, Rahardjo H (1993) An overview of unsaturated soil behaviour. Geotechnical special publication 1–1

Freund Y, Schapire RE (1996) Experiments with a new boosting algorithm. In: icml, vol. 96, pp 148–156

Gee GW, Or D (2002) 2.4 Particle-size analysis. Methods Soil Anal: Part 4 Phys Methods 5:255–293. https://doi.org/10.2136/sssabookser5.4.c12

Grossman RB, Reinsch TG (2002) 2.1 Bulk density and linear extensibility. Methods Soil Anal: Part 4 Phys Methods 5:201–228. https://doi.org/10.2136/sssabookser5.4.c9

Guevara M, Vargas R (2019) Downscaling satellite soil moisture using geomorphometry and machine learning. PLoS ONE 14(9):e0219639. https://doi.org/10.1371/journal.pone.0219639

Gunarathna MP, Sakai K, Nakandakari T, Momii K, Kumari MN (2019) Machine learning approaches to develop pedotransfer functions for tropical Sri Lankan soils. Water 11(9):1940. https://doi.org/10.3390/w11091940

Guo G, Wang H, Bell D, Bi Y, Greer K (2003) KNN model-based approach in classification. In: On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3–7, 2003. Proceedings. Springer, Berlin Heidelberg, pp 986–996. https://doi.org/10.1007/978-3-540-39964-3_62

Hastie T, Tibshirani R, Friedman JH, Friedman JH (2009) The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1–758). Springer, New York. https://doi.org/10.1007/978-0-387-21606-5

Hopmans JW, Dane JH (1986) Temperature dependence of soil hydraulic properties. Soil Sci Soc Am J 50(1):4–9

Huang G, Su X, Rizwan MS, Zhu Y, Hu H (2016) Chemical immobilization of Pb, Cu, and Cd by phosphate materials and calcium carbonate in contaminated soils. Environ Sci Pollut Res 23:16845–16856. https://doi.org/10.1007/s11356-016-6885-9

Hwang SI, Powers SE (2003) Lognormal distribution model for estimating soil water retention curves for sandy soils. Soil Sci 168(3):156–166. https://doi.org/10.1097/01.ss.0000058888.60072.e3

Im J, Park S, Rhee J, Baik J, Choi M (2016) Downscaling of AMSR-E soil moisture with MODIS products using machine learning approaches. Environ Earth Sci 75:1–19. https://doi.org/10.1007/s12665-016-5917-6

Jalal FE, Xu Y, Iqbal M, Javed MF, Jamhiri B (2021) Predictive modeling of swell-strength of expansive soils using artificial intelligence approaches: ANN, ANFIS and GEP. J Environ Manag 289:112420. https://doi.org/10.1016/j.jenvman.2021.112420

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 Resour Res 53(2):1539–1552. https://doi.org/10.1002/2016WR019533

Leij FJ, Romano N, Palladino M, Schaap MG, Coppola A (2004) Topographical attributes to predict soil hydraulic properties along a hillslope transect. Water Resour Res 40(2). https://doi.org/10.1029/2002WR001641

Li M, Zhang P, Adeel M, Guo Z, Chetwynd AJ, Ma C, Rui Y (2021) Physiological impacts of zero valent iron, Fe3O4 and Fe2O3 nanoparticles in rice plants and their potential as Fe fertilizers. Environ Pollut 269:116134. https://doi.org/10.1016/j.envpol.2020.116134

Liu Y, Yang Y, Jing W, Yue X (2017) Comparison of different machine learning approaches for monthly satellite-based soil moisture downscaling over Northeast China. Remote Sens 10(1):31. https://doi.org/10.3390/rs10010031

Long D, Bai L, Yan L, Zhang C, Yang W, Lei H, Shi C (2019) Generation of spatially complete and daily continuous surface soil moisture of high spatial resolution. Remote Sens Environ 233:111364. https://doi.org/10.1016/j.rse.2019.111364

Meskini-Vishkaee F, Mohammadi MH, Vanclooster M (2014) Predicting the soil moisture retention curve, from soil particle size distribution and bulk density data using a packing density scaling factor. Hydrol Earth Syst Sci 18(10):4053–4063. https://doi.org/10.5194/hess-18-4053-2014

Mi W, Sun Y, Xia S, Zhao H, Mi W, Brookes PC, Wu L (2018) Effect of inorganic fertilizers with organic amendments on soil chemical properties and rice yield in a low-productivity paddy soil. Geoderma 320:23–29. https://doi.org/10.1016/j.geoderma.2018.01.016

Mohammadi MH, Meskini-Vishkaee F (2012) Predicting the film and lens water volume between soil particles using particle size distribution data. J Hydrol 475:403–414. https://doi.org/10.1016/j.jhydrol.2012.10.024

Molnar C (2020) Interpretable machine learning. Lulu. com

Myles AJ, Feudale RN, Liu Y, Woody NA, Brown SD (2004) An introduction to decision tree modeling. J Chemometr 18(6):275–285. https://doi.org/10.1002/cem.873

Natekin A, Knoll A (2013) Gradient boosting machines, a tutorial. Front Neurorobot 7:21. https://doi.org/10.3389/fnbot.2013.00021

Nguyen PM, Van Le K, Cornelis WM (2014) Using categorical soil structure information to improve soil water retention estimates of tropical delta soils. Soil Res 52(5):443–452. https://doi.org/10.1071/SR13256

Nguyen PM, Haghverdi A, De Pue J, Botula YD, Le KV, Waegeman W, Cornelis WM (2017) Comparison of statistical regression and data-mining techniques in estimating soil water retention of tropical delta soils. Biosys Eng 153:12–27. https://doi.org/10.1016/j.biosystemseng.2016.10.013

Or D, Tuller M (1999) Liquid retention and interfacial area in variably saturated porous media: Upscaling from single-pore to sample-scale model. Water Resour Res 35(12):3591–3605. https://doi.org/10.1029/1999WR900262

Pachepsky YA, Rawls WJ, Lin HS (2006) Hydropedology and pedotransfer functions. Geoderma 131(3–4):308–316. https://doi.org/10.1016/j.geoderma.2005.03.012

Padarian J, Minasny B, McBratney AB (2019) Machine learning and soil sciences: a review aided by machine learning tools. https://doi.org/10.5194/soil-6-35-2020

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

Pham K, Kim D, Le CV, Won J (2023) Machine learning-based pedotransfer functions to predict soil water characteristics curves. Transp Geotechn 101052. https://doi.org/10.1016/j.trgeo.2023.101052

Rani A, Kumar N, Kumar J, Sinha NK (2022) Machine learning for soil moisture assessment. In: Deep learning for sustainable agriculture. Academic Press, pp 143–168. https://doi.org/10.1016/B978-0-323-85214-2.00001-X

Rastgou M, Bayat H, Mansoorizadeh M, Gregory AS (2020) Estimating the soil water retention curve: comparison of multiple nonlinear regression approach and random forest data mining technique. Comput Electron Agric 174:105502. https://doi.org/10.1016/j.compag.2020.105502

Rhoades JD (1983) Soluble salts. Methods Soil Anal: Part 2 Chem Microbiol Propert 9:167–179. https://doi.org/10.2134/agronmonogr9.2.2ed.c10

Rhoades JD (1996) Salinity: electrical conductivity and total dissolved solids. Methods Soil Anal: Part 3 Chem Methods 5:417–435. https://doi.org/10.2136/sssabookser5.3.c14

Ringrose-Voase AJ (1996) Measurement of soil macropore geometry by image analysis of sections through impregnated soil. Plant Soil 183:27–47. https://doi.org/10.1007/BF02185563

Sarkar D, De DK, Das R, Mandal B (2014) Removal of organic matter and oxides of iron and manganese from soil influences boron adsorption in soil. Geoderma 214:213–216. https://doi.org/10.1016/j.geoderma.2013.09.009

Schindler U, Mueller L, da Veiga M, Zhang Y, Schlindwein S, Hu C (2012) Comparison of water-retention functions obtained from the extended evaporation method and the standard methods sand/kaolin boxes and pressure plate extractor. J Plant Nutr Soil Sci 175(4):527–534. https://doi.org/10.1002/jpln.201100325

Senyurek V, Lei F, Boyd D, Kurum M, Gurbuz AC, Moorhead R (2020) Machine learning-based CYGNSS soil moisture estimates over ISMN sites in CONUS. Remote Sens 12(7):1168. https://doi.org/10.3390/rs12071168

Sermet Y, Demir I (2019) Towards an information centric flood ontology for information management and communication. Earth Sci Inf 12(4):541–551. https://doi.org/10.1007/s12145-019-00398-9

Shahraeeni E, Or D (2010) Thermo-evaporative fluxes from heterogeneous porous surfaces resolved by infrared thermography. Water Resour Res 46(9). https://doi.org/10.1029/2009WR008455

Srivastava PK, Han D, Ramirez MR, Islam T (2013) Machine learning techniques for downscaling SMOS satellite soil moisture using MODIS land surface temperature for hydrological application. Water Resour Manag 27:3127–3144. https://doi.org/10.1007/s11269-013-0337-9

Thomas GW (1996) Soil pH and soil acidity. Methods Soil Anal: Part 3 Chem Methods 5:475–490. https://doi.org/10.2136/sssabookser5.3.c16

Tuller M, Or D (2001) Hydraulic conductivity of variably saturated porous media: film and corner flow in angular pore space. Water Resour Res 37(5):1257–1276. https://doi.org/10.1029/2000WR900328

Tuller M, Or D (2005) Water films and scaling of soil characteristic curves at low water contents. Water Resour Res 41(9). https://doi.org/10.1029/2005WR004142

Tuller M, Or D, Dudley LM (1999) Adsorption and capillary condensation in porous media: liquid retention and interfacial configurations in angular pores. Water Resour Res 35(7):1949–1964. https://doi.org/10.1029/1999WR900098

Tuller M, Or D, Hillel D (2004) Retention of water in soil and the soil water characteristic curve. Encycl Soils Environ 4:278–289

Van Genuchten MT (1980) A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci Soc Am J 44(5):892–898. https://doi.org/10.2136/sssaj1980.03615995004400050002x

Vereecken H, Weynants M, Javaux M, Pachepsky Y, Schaap MG, Genuchten MV (2010) Using pedotransfer functions to estimate the van Genuchten-Mualem soil hydraulic properties: a review. Vadose Zone J 9(4):795–820. https://doi.org/10.2136/vzj2010.0045

Wadoux AMC, Molnar C (2022) Beyond prediction: methods for interpreting complex models of soil variation. Geoderma 422:115953. https://doi.org/10.1016/j.geoderma.2022.115953

Walkley A, Black IA (1934) An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. Soil Sci 37(1):29–38. https://doi.org/10.1016/j.geoderma.2021.115293

Wang S, Fan W, Zhu Y, Zhang J (2021) The effects of fitting parameters in best fit equations in determination of soil-water characteristic curve and estimation of hydraulic conductivity function. Rhizosphere 17:100291. https://doi.org/10.1016/j.rhisph.2020.100291

Wang C, Cai G, Liu X, Wu M (2022) Prediction of soil thermal conductivity based on Intelligent computing model. Heat Mass Transf 58(10):1695–1708. https://doi.org/10.1007/s00231-022-03209-y

Wei T, Fan W, Yu N, Wei YN (2019) Three-dimensional microstructure characterization of loess based on a serial sectioning technique. Eng Geol 261:105265. https://doi.org/10.1016/j.enggeo.2019.105265

Zappa L, Forkel M, Xaver A, Dorigo W (2019) Deriving field scale soil moisture from satellite observations and ground measurements in a hilly agricultural region. Remote Sens 11(22):2596. https://doi.org/10.1016/j.compgeo.2011.11.010

Zhai Q, Rahardjo H (2012) Determination of soil–water characteristic curve variables. Comput Geotech 42:37–43. https://doi.org/10.1016/j.compgeo.2011.11.010

Zhang N, Zou H, Zhang L, Puppala AJ, Liu S, Cai G (2020) A unified soil thermal conductivity model based on artificial neural network. Int J Therm Sci 155:106414. https://doi.org/10.1016/j.ijthermalsci.2020.106414

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