

Estimating Soil-Water Characteristic Curve (SWCC) Using Machine Learning and Soil Micro-Porosity Analysis

Aida Bakhshi

University of Zanjan

Parisa Alamdari (✉ p_alamdari@znu.ac.ir)

University of Zanjan

Ahmad Heidari

University of Tehran

Mohammad Hossein Mohammadi

University of Tehran

Research Article

Keywords: Image analysis, Modeling Soil Water, Predictive accuracy, Data mining software

Posted Date: August 9th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3216469/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

The soil water characteristic curve (SWCC) reveals soil porosity and soil-water interactions at different matric suctions. Numerous methods, such as laboratory determination, CT-scan, image analysis, and predictive models, have been employed to investigate soil porosity system and their correlation with the SWCC. Image analysis techniques offer valuable insight into soil pore system, providing data that cannot be obtained by other methods. This study aims to compare the role of image analysis technique in reconstructing the SWCC with the laboratory measurement method. Eight machine learning models and algorithms, including Gradient Boosting (GB), Ada Boost (AB), Decision Tree (DT), Random Forest (RF), Artificial Neural Network (ANN), Support Vector Machine (SVM), K-Nearest Neighborhood (KNN), and Linear Regression (LR), were utilized for the reconstruction of the SWCC using the Orange-3 data mining software. The predicted SWCCs by models were compared with the measured SWCC. The models used to reconstruction of SWCC were categorized as capable and incapable to SWCC prediction. Four statistical parameters, namely RMSE, MAE, Willmott's index of agreement (d1), and R2, were utilized to assess the performance of the models using all input data. The GB, AB, DT were the top best models in correct prediction of SWCC. Among them the GB model achieved near-perfect predictions, with RMSE, MAE, d1, and R2 values of 0.016, 0.011, 0.94, and 0.982, respectively. However, when the matric suction was used as only input data, the performance of all the models declined. This study demonstrates that Orange-3 is user friendly software to predict SWCC, without labor preprocessing activities.

1. Introduction

The Soil Water Characteristic Curve (SWCC) is closely associated with soil physical properties and plays a crucial role in soil and water management (Shwetha and Varija, 2015). The SWCC provides valuable direct and indirect information about the behavior of water in unsaturated soils (Zhai and Rahardjo, 2012; van Genuchten et al., 2015). There is a need to reliable determination of the SWCC of any given soil using a combination of both measurement and predicting techniques. However, all the field, laboratory, and computer vision-based measurements of SWCC are expensive, tedious, time-consuming, and sometimes impossible due to issues related to scaling, spatial variability, and stud-site inaccessibility (Achieng, 2019) thus use of modeling procedures is a very common approach to predict SWCC (Dobarco et al., 2019).

While multiple linear regression (MLR), ANN, and SVR, have been commonly used in the development of pedo-transfer functions (PTFs) (Rani et al., 2022). There has been a significant increase in the application of machine learning (ML) algorithms such as LR, ANNs, SVMs, classification and regression tree (CART), and RF, in soil moisture researches. These ML algorithms are preferred for their non-parametric nature and ability to capture complex and non-linear relationships (Padarian et al., 2020).

Machine learning techniques for estimating SWCC fall under the category of supervised learning, where a labeled training dataset is provided with known output values. The model is trained using algorithms applied to the input dataset to predict the desired output. Training continues until the model achieves the desired accuracy on the training dataset. Supervised learning is commonly used for classification and regression tasks (Rani et al., 2022).

Achieng (2019) conducted a comparative study of several ML algorithms for modeling SWCC in loamy sand soil. They found that the RBF-based support vector regression (SVR) outperformed SVR with linear and polynomial kernels, single-layer ANN, and deep neural network (DNN) models. In another study, Araya and Ghezzehei (2019) demonstrated the superior performance of the Boosted Regression Tree (BRT) model compared to other algorithms, such as KNN, SVR, and RF, for predicting saturated hydraulic conductivity. However, the RF model closely followed the BRT model in terms of performance. These findings highlight the satisfactory performance of various ML algorithms in predicting environmental events. For instance, Hong and Pai (2007) and Hu et al., (2013) observed the effective use of techniques such as ANN, SVM, and KNN for forecasting soil water evaporation. Furthermore, Baydaroglu and Kocak (2014) observed the valuable performance of these algorithms in predicting free water surfaces, while Valipour et al., (2012; 2013) utilized these algorithms to predict water reservoir inflows. As a result of their high flexibility, accurate predictive performance, and consistent results, data mining techniques have become a preferred choice for many researchers seeking to enhance their understanding of unsaturated soil hydrological properties (Botula et al., 2013).

The capability of machine learning methods to accurately fit the SWCC is directly influenced by the availability of measured soil water content data at various soil matric potentials (Hastie et al., 2009; K. Lamorski et al., 2013). Toth et al. (2014) analyzed the SWCC using the RF model at four matric suctions (0.1, 33, 1500 kPa, and 150 MPa). The results demonstrated that the significance of soil properties in predicting soil water content varies across different soil types and matric suctions. In another study, Gunarathna et al. (2019) evaluated ML algorithms, including ANN and KNN, to estimate the volumetric water content at matric suctions of 10, 33, and 1500 kPa. Pekel (2020) applied decision tree regression, specifically the CART algorithm, to estimate soil moisture. The input variables were air temperature, time, relative humidity, and soil temperature. In other study, Cai et al. (2019) proposed the use of a Deep Learning Regression Network (DLRN) with big data fitting capability for constructing soil moisture prediction models. Numerical models like HYDRUS-2D often require a large amount of input data for simulating the time-series of soil moisture. However, if limited input data is available, ML algorithms such as SVM and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) can efficiently handle the task (Karandish & Simunek, 2016). While the accuracy of ML algorithms may be comparatively lower than numerical models, they can serve as a better alternative under limited and missing data conditions.

Although machine learning techniques have been explored in various soil moisture-related studies, the use of boosting techniques for this purpose is relatively rare (Araya and Ghezzehei, 2019). Boosting methods aim to iteratively combine weak learners to create a strong learner that can provide more accurate predictions. One popular technique in boosting is gradient boosting, which involves sequentially adding predictors to an ensemble, with each predictor correcting the errors made by its predecessor. Unlike AB, which adjusts the weights of data points, gradient boosting trains on the residual errors of the previous predictor. In this study, gradient boosting and AB were selected as the most popular boosting-based algorithms for the estimation of SWCC using ML.

Vereecken et al. (2010) concluded that incorporating soil structure information as one of the predictors in PTFs is likely to enhance their performance. Nguyen et al. (2014) found that including categorical soil structure information in point PTFs developed using the MLR technique improved the accuracy of SWCC estimation for tropical paddy soils. They also suggested further investigation to explore whether these improvements would hold true when using different

data mining techniques and for other types of PTFs. Passoni et al. (2014) utilized ImageJ software to characterize the porosity of Oxisols in the southeastern region of Brazil by relying on shape factors or pore form. ImageJ offers a convenient built-in option for analyzing soil porosity. This feature provides valuable output parameters including the number of porosity surface area (Total Area of Porous Regions, cm^2), volume (Total Number of Porous Voxels $\times$ 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, we utilized detailed soil structural properties derived from image analysis as inputs in the machine learning technique employed in this study. The aim was to assess whether the incorporation of such soil structure information would contribute to improved SWCC estimation.

Data mining techniques have shown superiority in modeling the interactions of the soil-water complex system compared to traditional MLR techniques. However, these techniques also have some drawbacks, including susceptibility to over fitting, high data demand, and expert knowledge requirements. In this study, machine learning methods were employed to analyze soil structure using selected soil properties. Therefore, our objective is to predict SWCC in soil samples with different properties using data mining algorithms. The prediction process was conducted under two conditions: 1) using matric suction as the only predefined input, and 2) using a range of input parameters obtained from laboratory and image analysis methods.

2. Material and Methods

2.1. Soil sampling

The characteristics of soil porosity and SWCC of two soil samples with textural classes of loamy sand and silty clay, taken from Arenosols (located at coordinates 35° 54' N and 50° 32' E), and Vertisols (located at coordinates 36° 22' N and 49° 35' E) of Central Iran. These soil samples were subjected to various treatments, including amendments such as CaCO_3 , Fe_2O_3 , vermin-compost, their combined treatment, cation treatment, and a structural degradation treatment. The aim of this study was to examine the effects of these treatments on the soil porosity and SWCC. The topsoil (0–10 cm) was sampled, dried and sieved to pass a 2 mm sieve and analyzed to determine typical soil characteristics. The basic soil properties that have been widely used for SWCC estimation (Wassar et al., 2016), including soil organic carbon (SOC) (Walkley & Black, 1934), particle size distribution (PSD) (Gee & Or, 2002), cation exchange capacity (CEC) (Rhoades, 1983), electrical conductivity (EC) (Rhoades, 1996), pH (Thomas, 1996) and the parameters of soil porosity (a , n , θ_s , θ_r , h_i) (Dexter et al., 2008) were determined using the standard methods. The bulk density of soil samples was determined using the core method (Grossman & Reinsch, 2002), in the undisturbed samples collected using Kopecky rings (5 cm in height and 5 cm in diameter). Machine learning (ML) models, known for their ability to predict target properties without limitations (Bell, 2022), were employed to model the data obtained from the studied soils and analyze the effects of different treatments.

2.2. Preparation of the treatments

The soil samples were subjected to various treatments, including four levels of CaCO_3 (0%, 1.5%, 3%, 5%), $\text{Fe}_2\text{O}_3 \cdot 7\text{H}_2\text{O}$ (0%, 0.5%, 1%, 2%), vermicompost (0%, 0.6%, 1%, 2%) with the composition of vermicompost (OC, N, etc.) incorporated, and a combination of these treatments (including the blank, 1st level: $\text{CaCO}_3 = 1.5\%$, $\text{Fe}_2\text{O}_3 \cdot 7\text{H}_2\text{O} = 0.5\%$, vermicompost = 1%, as well as the 2nd and 3rd levels combined). Furthermore, to assess the effects of cations on soil structure, the soil pots were irrigated with solutions containing CaCl_2 and NaCl at concentrations of 0, 5, 10, and 20 meq L^{-1} during the incubation period, as explained subsequently. In addition to the aforementioned amendment treatments, three replicates of degraded treatments were included in the study. These degraded treatments were prepared through a consolidation process specific to each soil texture. The treated samples were weighed and filled in 1 kg pots (without drain) and incubated at room temperature 24 ~ 26°C. To investigate the effect of shrink-swell on soil structure development, all treated samples (including disturbed and structure-less soils) were subjected to a series of periodic shrink-swelling and wetting and drying cycles, repeated 20 times. The soaked pots were weighed daily to track minimal variations. Rewetting until field capacity is achieved by carefully adding water to the sponge cover placed on top of the columns to avoid disturbing soil conditions. Detailed information about the studied treatments can be seen in Table 1. The total number of samples studied was 128, which consisted of $5 \times 4 \times 3 \times 2$ (amendment treatments) + 4×2 (degraded treatments). One of the replicates was utilized to prepare intact samples for measuring the soil water characteristic curve using pressure plate and pressure membrane methods, while the other two replicates were impregnated with a mixture of polyester resin, catalyst, hardener, and fluorescent dye for UV photography and subsequent image analysis.

2.3. Determination of the soil water characteristic curve

The soil water characteristic curve (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 and higher matric suctions (100, 300, 500, 1000, 3000, 5000, 9000, and 15000 cm) using pressure plate/pressure membrane apparatus. 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 15000 cm.

2.4. Preparation of samples for imaging after SWCC construction

2.4.1. Impregnation with fluorescent dye resin

A total of 128 pre-treated samples were impregnated with a mixture of polyester resin and styrene in a 5:1 ratio, along with an appropriate amount of hardener and catalyst. To enhance the visibility of soil pores for digital imaging, 2 g.L^{-1} of fluorescent dye was added as a brightener (Ringrose-Voase 1996). The mixture was added up to the middle height of the samples placed in plastic containers within a vacuum desiccator. The desiccator was connected to a compressor and evacuated to 8 psi for 2 h to ensure resin filling and air removal from the pores. Subsequently, the samples were taken out of the desiccator, refilled with the same mixture up to a height of 1 cm from the upper surface of the samples and subjected to a vacuum for an additional two hours. Following the vacuum process, the samples were refilled to their upper level and sealed to prevent rapid volatilization of styrene (Wei et al., 2019; Liu et al., 2016). After

one week, the sealed samples were opened to let the styrene to volatilize at room temperature. Approximately 75 days later, the polyester resin reached the desired hardness.

2.4.2. Cutting, polishing, and imaging

The hardened samples were cut and polished to facilitate imaging and visualization of soil pores and potential structure development in the treated soils (Wei et al., 2019). Two horizontal and two vertical cuts were made on each sample, exposing the four nearest horizontal and vertical surfaces (2 sides × 2 cuts) for digital imaging. Imaging was conducted in a dark room equipped with two UV lamps to maximize the brightness of the fluorescent dye impregnated in the pores, enabling visibility of even the smallest pores. The images were captured using a digital camera with a resolution of 12 MP and an f/1.8 aperture. Subsequently, the captured images were imported into the ImageJ software for preprocessing and analysis of soil pores.

2.4.3. Image preprocessing

The color images were imported into the ImageJ software and then scaled, calibrated, and converted into grayscale using the image conversion module. The grayscale images were further processed by applying a thresholding method to convert them into binary images, where pores were represented as white pixels and solids as black pixels. The resulting binary images were stacked together to obtain four 3D volumes (two horizontal and two vertical) for each sample. Key parameters of the pores, including 3D porosity (defined as the total volume of pore voxels divided by the number of voxels), pore sphericity (ranging from 0 for elongated pores to 1 for spherical pores), aspect ratio (calculated as the ratio of the short axis to the long axis), and object orientation, were determined using 3D and 2D plugins in ImageJ. Object orientation refers to two angles: φ (ranging from 0° to 90°), which represents the angle between the horizontal plane and the particle's long axis (pore channel), and θ (ranging from 0° to 180°), which represents the direction of the long axis (pore channel) projected onto the horizontal plane. Some of the 2-D and 3-D properties, such as pore space surface area and sphericity, were directly determined using ImageJ software. Porosity was calculated as the fraction of image volume occupied by the pore space.

2.5. Machine learning procedure

The Orange.3 data mining software, which is a visual-based version of the Python programming language, was utilized for preprocessing the raw data collected from laboratory and image analysis methods. 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 machine learning process. This software provides comprehensive data mining, modeling, and evaluation tools without requiring complex and time-consuming coding. The results of this procedure enable a quick comparison of various fitted models, including Gradient Boosting, Ada Boost, Decision Tree, Random Forest, Neural Network, Support Vector Machine, k-Nearest Neighbors, and Linear Regression, to identify the most effective machine learning algorithm for predicting SWCC from the input raw dataset. Furthermore, the Feature Importance widget was used to determine the relative importance of input features in predicting SWCC with a minimal dataset. This widget assesses the contribution of each feature to the prediction by measuring the increase in prediction error when the values of the feature are permuted. Additionally, Orange.3 software was used to analyze the impact of different values of each feature on the model output. This analysis helps identify critical values of important features in modeling the SWCC, providing researchers with insights into the most influential range of feature values. The statistical analysis included the calculation of R-squared (R^2) and root mean square error (RMSE) to assess the predictive capabilities of machine learning algorithms for SWCC prediction using features obtained from laboratory and image analysis methods.

3. Results and discussion

3.1. The properties of initial soil samples

Table 1 presents 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 ($4.8 \text{ cmol}^+ \text{ kg}^{-1}$ vs. $24.1 \text{ cmol}^+ \text{ kg}^{-1}$). The matric suction at the inflection point (h_i) of the soil water characteristic curve (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 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 ($\alpha, \text{ cm}^{-1}$), as well as in the saturation water content ($\theta_s, \text{ g.g}^{-1}$) and residual water content ($\theta_r, \text{ g.g}^{-1}$) between the two samples.

Table 1
Initial soil samples properties

Soil Texture	Particle size distribution			Structural and hydraulic properties						CEC	EC	OC	pH
	Clay (%)	Silt (%)	Sand (%)	Bulk Density (kg.m^{-3})	α (cm^{-1})	n (-)	θ_s (g.g^{-1})	θ_r (g.g^{-1})	h_i (cm)	($\text{cmol}^+.\text{kg}^{-1}$)	($\mu\text{S.cm}^{-1}$)	(%)	
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

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

Table 2 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. In the loamy sand samples, all measured properties (BD, CEC, EC, and OC except the pH values) were increased in all treatments, except for the Fe_2O_3 treatment, compared to the blank sample. On the other hand, in the silty clay samples, bulk density was decreased in most treatments compared to the blank, except for the treatments involving cations and degraded samples. In the loamy sand samples, CEC showed a slight increase in the CaCO_3 and OC treatments, while it decreased in the other treatments. On the other hand, in the silty clay samples, EC, OC, and pH values increased in all treatments compared to the blank sample. The results obtained from image analyses revealed that the properties examined, including porosity surface area, volume, elongation, compactness, sphericity, and flatness, did not exhibit consistent variations across different treatments. In the loamy sand sample, the porosity surface area increased in the CaCO_3 , Fe_2O_3 , and combined treatments, while decreasing in the other treatments. In the silty clay sample, there was a slight increase in the OC treatment, but a decrease was observed in all other treatments compared to the blank.

Table 2
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 ₂ O 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 ₂ O 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												

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 (-)
	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 ₂ O 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
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 ₂ O 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

Similar to Table 2 and Table 1, a dataset of individual treatments was prepared, which was automatically divided into model training and test datasets. Orange.3 software, based on a machine learning procedure, applied the mentioned features from Table 2 and Table 1 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.

3.2. 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 (Ahangar-asr et al., 2012). 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 machine learning models. Among these models, matric suction was identified as the most important parameter in the GB (Fig. 1.a), AB (Fig. 1.b), RF (Fig. 1.d), and SVM (Fig. 1.f) 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. 1.c), ANN (Fig. 1.e), KNN (Fig. 1.g), and LR (Fig. 1.h) 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. 1.e). 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.

After matric suction, soil pore characteristics emerged as the next important parameters in predicting accurate results, except in the case of the ANN model. At least one or two pore characteristics played a role in predicting the SWCC, with structural flatness and porosity surface area being particularly influential compared to other pore characteristics. In their study, Ahangar-Asr et al. (2012) attempted to incorporate soil void ratio as an input parameter in a model aimed at predicting SWCC and soil porosity characteristics. However, they did not specifically investigate the influence of these properties on the model's results.

3.3. Comparison of the models' predicted results

3.3.1 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, 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), as shown in Table 3. 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 machine learning 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 Im et al. (2016) reported RMSE values greater than $0.04 \text{ m}^3 \text{ m}^{-3}$, while Bai et al. (2019) achieved accurate results with an RMSE less than $0.02 \text{ m}^3 \text{ m}^{-3}$.

However, in this study (ANN, SVM, KNN, and LR) 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 machine learning method, categorizing it as a non-accurate model. Nguyen et al. 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 $0.05 \text{ m}^3 \text{ m}^{-3}$. In another study, Liu et al. (2018) observed an RMSE greater than $0.07 \text{ m}^3 \text{ m}^{-3}$ in the prediction of moisture content using the KNN algorithm with inputs derived from satellite-derived data.

Table 3
The statistics obtained for the models used to generate SWCC using all parameters

Model	RMSE	MAE	d1	R^2
Gradient Boosting	0.016	0.011	0.94	0.982
AB	0.020	0.014	0.93	0.971
Random Forest	0.021	0.015	0.93	0.968
Tree	0.028	0.018	0.93	0.944
Neural Network	0.087	0.071	0.65	0.462
SVM	0.077	0.070	0.58	0.583
KNN	0.085	0.071	0.62	0.490
Linear Regression	0.069	0.059	0.69	0.664

Table 4 presents the Pearson correlation (r) between the measured water content (θ_{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., 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 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 4, 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.

As anticipated, 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 ($r = -0.469$).

Table 4
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

3.3.2 Just Applying soil matric suction as model input feature

To assess the necessity of incorporating additional input features for improving the model outputs, we conducted an evaluation 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 5 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 6), 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.

Table 5
Statistics of models in the case where matric suction was used as the only input parameter

Model	RMSE	MAE	d1	R^2
Gradient Boosting	0.093	0.083	0.52	0.391
AB	0.093	0.083	0.50	0.386
Tree	0.093	0.083	0.52	0.391
Random Forest	0.093	0.082	0.51	0.389
Neural Network	0.104	0.083	0.46	0.229
SVM	0.095	0.083	0.46	0.356
kNN	0.097	0.087	0.53	0.332
Linear Regression	0.109	0.091	0.33	0.154

The use of soil matric suction as the sole input feature in the eight evaluating models significantly reduces the correlation between the models and the measured water content (θ_{Measured}). This, in turn, causes the correlation of the linear regression model with θ_{Measured} to be lower than the correlations between matric suction and θ_{Measured} (as shown in Table 6). Based on these findings, it can be concluded that utilizing matric suction values alone in the prediction of the Soil Water Characteristic Curve (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.

Table 6
Pearson correlation (r) when the matric suction is included as the only modeling parameter

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

3.4. Predicted SWCCs with Evaluating models based on the machine learning procedure

Figures 2 and 3 illustrate the soil water characteristic curves 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. 2 and 3, we explicitly demonstrate these differences. Specifically, for the Loamy Sand soil sample, Gradient Boosting, Ada Boost, Tree, and Random forest models (Fig. 2, a-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 machine learning 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. 2 and 3, e ~ 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 Soil Water Characteristic Curve (SWCC). Similarly, the kNN model yields inaccurate outputs for the Silty Clay soil sample.

3.5. Evaluating models uncertainty

Table 7 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.

Table 7
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
15000	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
15000	8.1	6.4	10.0	5.5	39.6	28.4	57.0	19.3

3.6. Prediction errors at two sides of the inflection point

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, 2003a; 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 suction in Figs. 2 and 3, as well as in Table 7. To further support this observation, we compared the error percentages of the evaluating models at matric suction values below and above a matric suction related to SWCCs inflection point (h_i). For the Loamy Sand soil samples, we calculate h_i equal to 70 cm, while for the Silty Clay soil samples, h_i was calculated equal to 300 cm. Figure 4 presents the results of this comparison. In both studied soil textures, we observed that 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 . Additionally, we observed that 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.

3.7. Residual contents of predicted SWCCs

To quantify the absolute differences between predicted and measured SWCCs, we have presented the difference curves for both Loamy Sand and Silty Clay soil samples. Figure 5 depicts the difference curves for Loamy Sand samples, while Fig. 6 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. 5 and 6 (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. 5 and 6 (subfigures e-h). Similar to the results of this 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. 5-f and 6-f 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

The utilization of orange.3 data mining software has facilitated a simple and efficient modeling procedure for predicting the soil water characteristic curve (SWCC) based on variations in soil properties. This software enables the seamless integration of a diverse range of measured physical soil properties into the model, without requiring extensive programming knowledge. In this approach, the training, testing, and evaluation of machine learning models were conducted to predict the SWCC. Interestingly, it was observed that models relying solely on soil matric suction as a predefined feature were unable to accurately predict SWCC. The evaluation of the models in this form revealed that the Mean Absolute Error (MAE) exceeded 0.08, and (R^2) value was below 40%. Therefore, this study examined the effects of using all possible soil properties as model features to enhance its performance of any of these features resulted in a significant decrease in model prediction accuracy. our findings indicate that for a more accurate estimation of SWCC, it is crucial to consider not only soil matric suction as a predefined feature, but also important soil properties such as soil structural (bulk density), physicochemical (organic carbon), and morphological (structural flatness or porosity surface area) properties. These properties were included as measured features in the model, demonstrating their significance in achieving a more precise estimation of SWCC. Based on the findings of this study, we can draw conclusions regarding the necessity of considering other soil properties, in addition to matric suction, for accurate prediction of SWCC using ML. It is determined that SWCC prediction requires the input of soil properties, thus, we successfully determined which properties have the most significant impact on the models' outputs. A negligible error was identified in the models mentioned above which is related to matric suctions greater than SWCC inflection point matric suction. Due to the low susceptibility of soil properties on soil water content at specified matric suction in this part of SWCC, we can regardless of this error and assume perfect prediction in mentioned models.

Declarations

Funding

No funding was received to assist with the preparation of this manuscript.

Conflicts of interest/Competing interests

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

Availability of data and material

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

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

References

1. 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>
2. 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>
3. Araya SN, Ghezzehei TA (2019) Using machine learning for prediction of saturated hydraulic conductivity and its sensitivity to soil structural perturbations. *Water Resour Res* 55(7):5715–5737
4. 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>
5. Baydaroğlu Ö, Koçak K (2014) SVR-based prediction of evaporation combined with chaotic approach. *J Hydrol* 508:356–363. <https://doi.org/10.1016/j.jhydrol.2013.11.008>
6. Bell J (2022) What is machine learning?. *Machine Learning and the City: Applications in Architecture and Urban Design*, 207–216. <https://doi.org/10.1002/9781119815075.ch18>
7. Botula YD, Nemes A, Mafuka P, Van Ranst E, Cornelis WM (2013) Prediction of Water Retention of Soils from the Humid Tropics by the Nonparametric-Nearest Neighbor Approach, *Vadose Zo. J* 12:1–17. <https://doi.org/10.2136/vzj2012.0123>
8. 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>
9. 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>
10. 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>
11. Fredlund DG, Xing A, Fredlund MD, Barbour SL (1996) The relationship of the unsaturated soil shear strength to the soil-water characteristic curve. *Can Geotech J* 33(3):440–448. <https://doi.org/10.1139/t96-065>
12. Gee GW, Or D (2002) 2.4 Particle-size analysis. *Methods of soil analysis: Part 4 physical methods*, 5, 255–293. <https://doi.org/10.2136/sssabookser5.4.c12>
13. Grossman RB, Reinsch TG (2002) 2.1 Bulk density and linear extensibility. *Methods of soil analysis: Part 4 physical methods*, 5, 201–228. <https://doi.org/10.2136/sssabookser5.4.c9>
14. 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>
15. 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>
16. Hastie T, Tibshirani R, Friedman J (2009) Springer: New York. *NY, USA*. <https://doi.org/10.1007/978-0-387-21606-5>
17. Hong WC, Pai PF (2007) Potential assessment of the support vector regression technique in rainfall forecasting. *Water Resour Manage* 21:495–513. <https://doi.org/10.1007/s11269-006-9026>
18. Hu J, Liu J, Liu Y, Gao C (2013) EMD-KNN model for annual average rainfall forecasting. *J Hydrol Eng* 18(11):1450–1457. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000481](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000481)
19. Hwang SI, Powers SE (2003) Lognormal distribution model for estimating soil water retention curves for sandy soils. *Soil science*, 168(3), 156–166. [https://doi.org/10.1002/1522-0185\(200303\)168:3<156::AID-SOIL156>3.0.CO;2-3](https://doi.org/10.1002/1522-0185(200303)168:3<156::AID-SOIL156>3.0.CO;2-3)
20. 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>
21. Karandish F, Šimůnek J (2016) A comparison of numerical and machine-learning modeling of soil water content with limited input data. *J Hydrol* 543:892–909. <https://doi.org/10.1016/j.jhydrol.2016.11.007>
22. Lamorski K, Pastuszka T, Krzyszczak J, Sławiński C, Witkowska-Walczak B (2013) Soil water dynamic modeling using the physical and support vector machine methods. *Vadose Zone J* 12(4). <https://doi.org/10.2136/vzj2013.05.0085>
23. 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>
24. 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
25. <https://doi.org/10.3390/rs10010031>

26. 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>
27. 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>
28. 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
29. <https://doi.org/10.1016/j.jhydrol.2012.10.024>
30. 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
31. <https://doi.org/10.1071/SR13256>
32. 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. *Biosyst Eng* 153:12–27
33. <https://doi.org/10.1016/j.biosystemseng.2016.10.013>
34. 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
35. <https://doi.org/10.1029/1999WR900262>
36. 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>
37. Passoni S, Borges FS, Pires LF, Saab SC, Cooper M (2014) Software Image J to study soil pore distribution. *Ciência e Agrotecnologia* 38:122–128. <https://doi.org/10.1590/S1413-70542014000200003>
38. Pekel E (2020) Estimation of soil moisture using decision tree regression. *Theoret Appl Climatol* 139(3–4):1111–1119. <https://doi.org/10.1007/s00704-019-03048-8>
39. Pereira JH, Fredlund DG (2000) Volume change behavior of collapsible compacted gneiss soil. *J Geotech Geoenviron Eng* 126(10):907–916
40. [https://doi.org/10.1061/\(ASCE\)1090-0241\(2000\)126:10\(907\)](https://doi.org/10.1061/(ASCE)1090-0241(2000)126:10(907))
41. Rani A, Kumar N, Kumar J, Sinha NK (2022) Machine learning for soil moisture assessment. In *Deep Learning for Sustainable Agriculture* (pp. 143–168). Academic Press. <https://doi.org/10.1016/B978-0-323-85214-2.00001-X>
42. Rhoades JD (1983) Soluble salts. *Methods of Soil Analysis: Part 2 Chemical and Microbiological Properties*, 9, 167–179. <https://doi.org/10.2134/agronmonogr9.2.2ed.c10>
43. Rhoades JD (1996) Salinity: Electrical conductivity and total dissolved solids. *Methods of soil analysis: Part 3 Chemical methods*, 5, 417–435. <https://doi.org/10.2136/sssabookser5.3.c14>
44. 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>
45. Rudiyanto, Sakai M, van Genuchten MT, Alazba AA, Setiawan BI, Minasny B (2015) A complete soil hydraulic model accounting for capillary and adsorptive water retention, capillary and film conductivity, and hysteresis. *Water Resour Res* 51(11):8757–8772. <https://doi.org/10.1002/2015WR017703>
46. 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>
47. Shwetha P, Varija K (2015) Soil water retention curve from saturated hydraulic conductivity for sandy loam and loamy sand textured soils. *Aquat Procedia* 4:1142–1149
48. 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 Manage* 27:3127–3144. <https://doi.org/10.1007/s11269-013-0337-9>
49. Thomas GW (1996) Soil pH and soil acidity. *Methods of soil analysis: part 3 chemical methods*, 5, 475–490. <https://doi.org/10.2136/sssabookser5.3.c16>
50. Tóth B, Makó A, Gergely TH (2014) Role of soil properties in water retention characteristics of main Hungarian soil types. *J Cent Eur Agric*. <https://doi.org/10.5513/jcea.v15i2.2634>
51. 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
52. <https://doi.org/10.1029/2000WR900328>
53. 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>
54. 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>
55. Valipour M, Banihabib ME, Behbahani SR (2012) Parameters estimate of autoregressive moving average and autoregressive integrated moving average models and compare their ability for inflow forecasting. *J Math Stat* 8(3):330–338
56. Valipour M, Banihabib ME, Behbahani SR (2013) Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. *J Hydrol* 476:433–441. <https://doi.org/10.1016/j.jhydrol.2012.11.017>
57. 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

58. <https://doi.org/10.2136/sssaj1980.03615995004400050002x>
59. 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>
60. 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
61. Wang C, Li SY, He XJ, Chen Q, Zhang H, Liu XY (2021) Improved prediction of water retention characteristic based on soil gradation and clay fraction. *Geoderma* 404:115293
62. <https://doi.org/10.1016/j.geoderma.2021.115293>
63. Wassar F, Gandolfi C, Rienzner M, Chiaradia EA, Bernardoni E (2016) Predicted and measured soil retention curve parameters in Lombardy region north of Italy. *Int Soil Water Conserv Res* 4(3):207–214. <https://doi.org/10.1016/j.iswcr.2016.05.005>
64. 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
65. <https://doi.org/10.1016/j.enggeo.2019.105265>
66. 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>
67. 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>

Figures

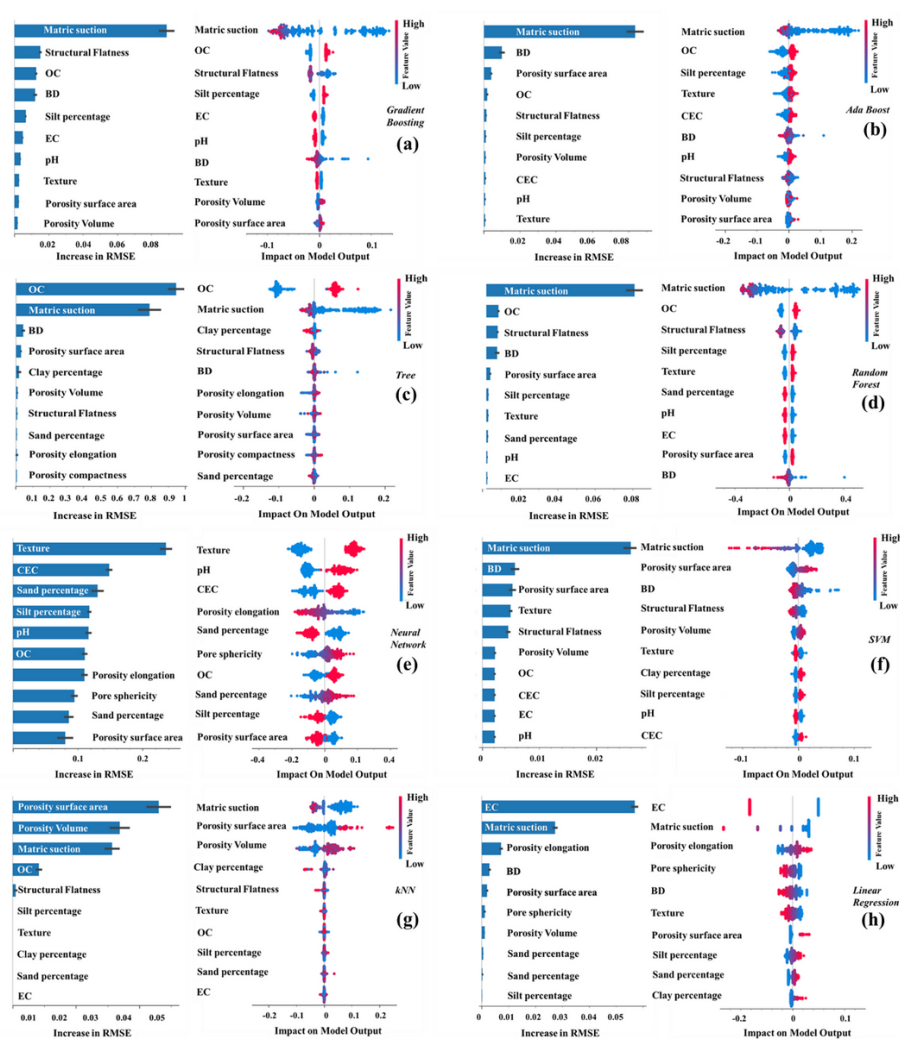


Figure 1

Input parameters and their relative importance in accurate prediction of GB, AB, RF, SVM, DT, ANN, KNN, and LR models

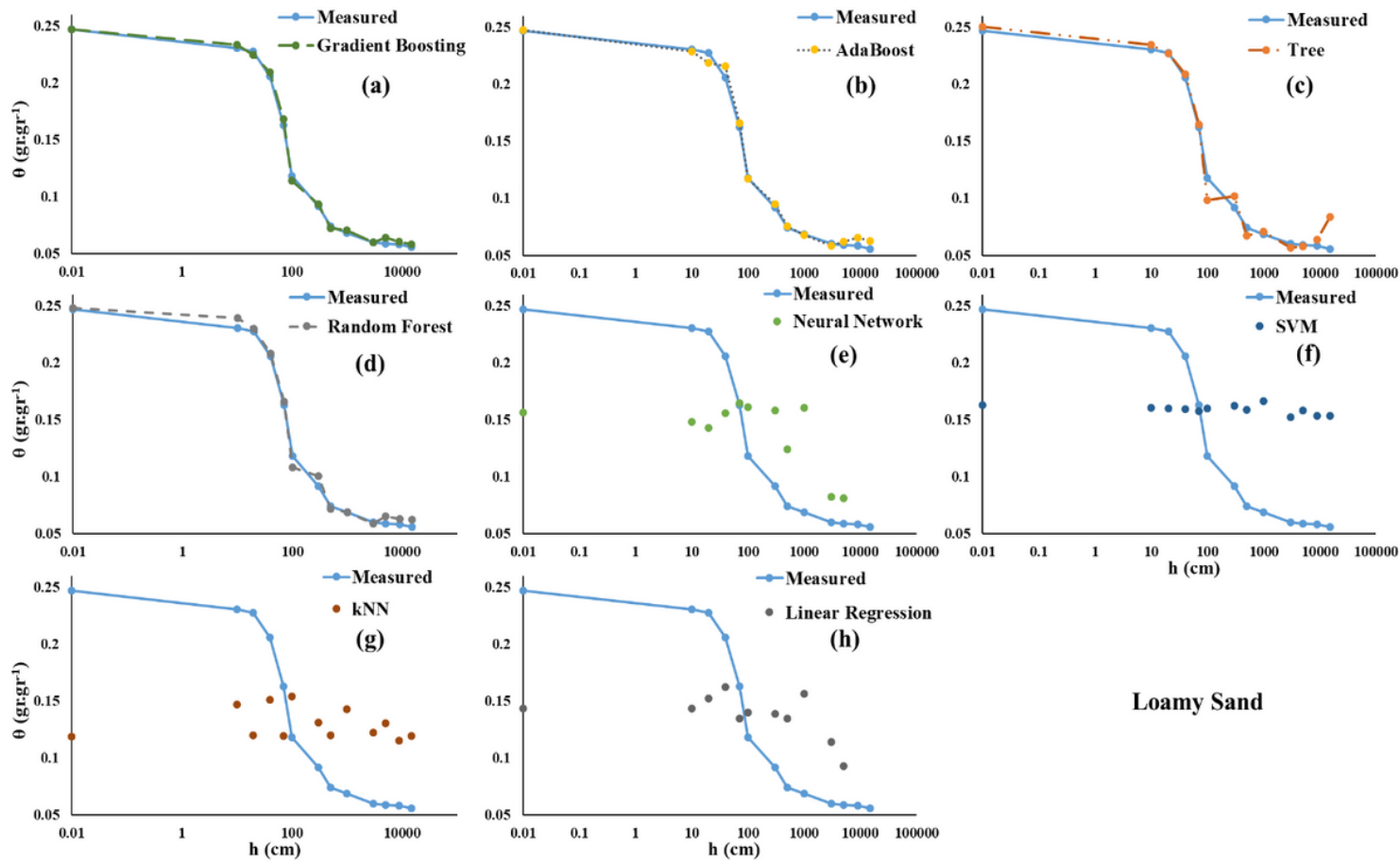


Figure 2

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

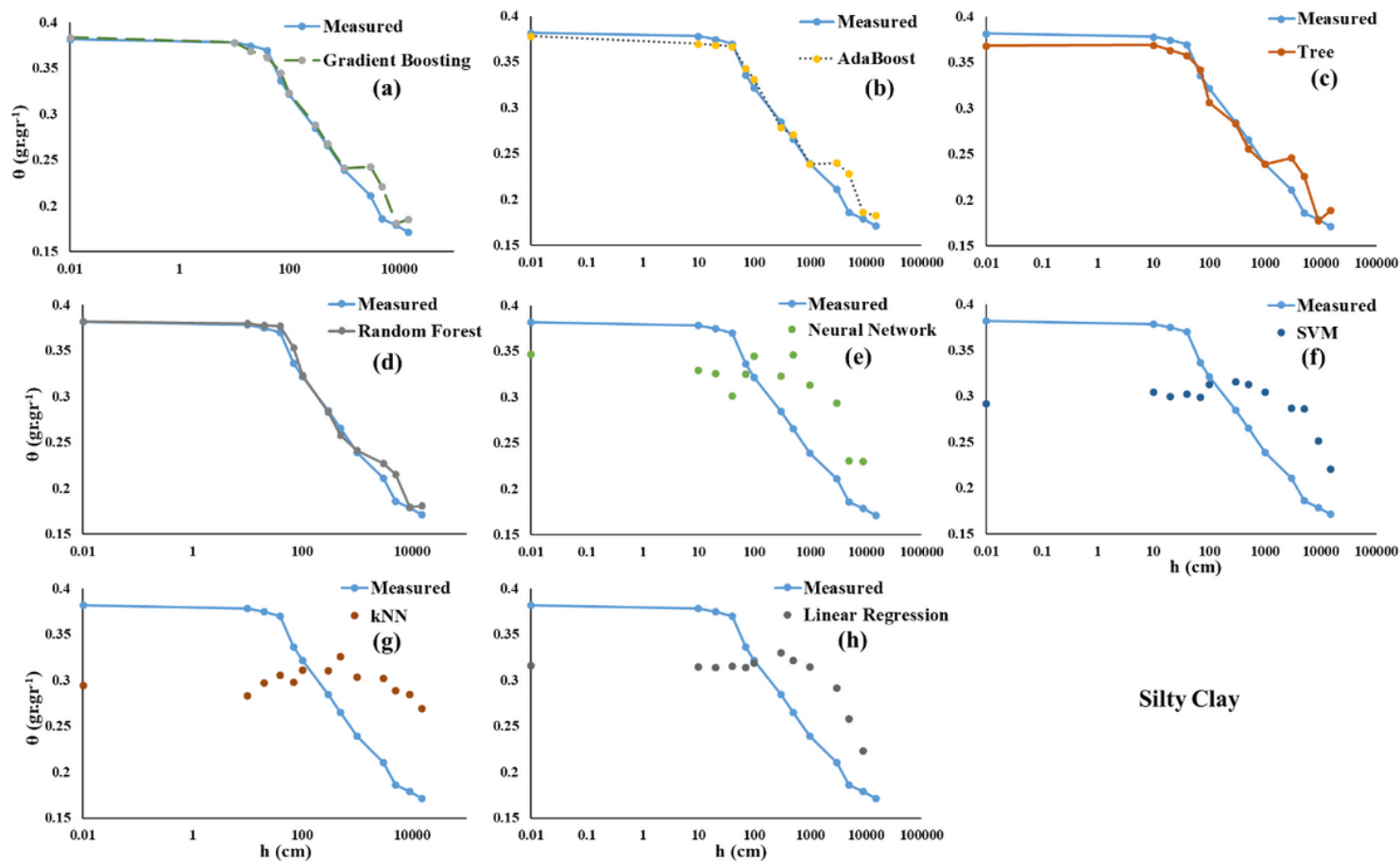


Figure 3

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

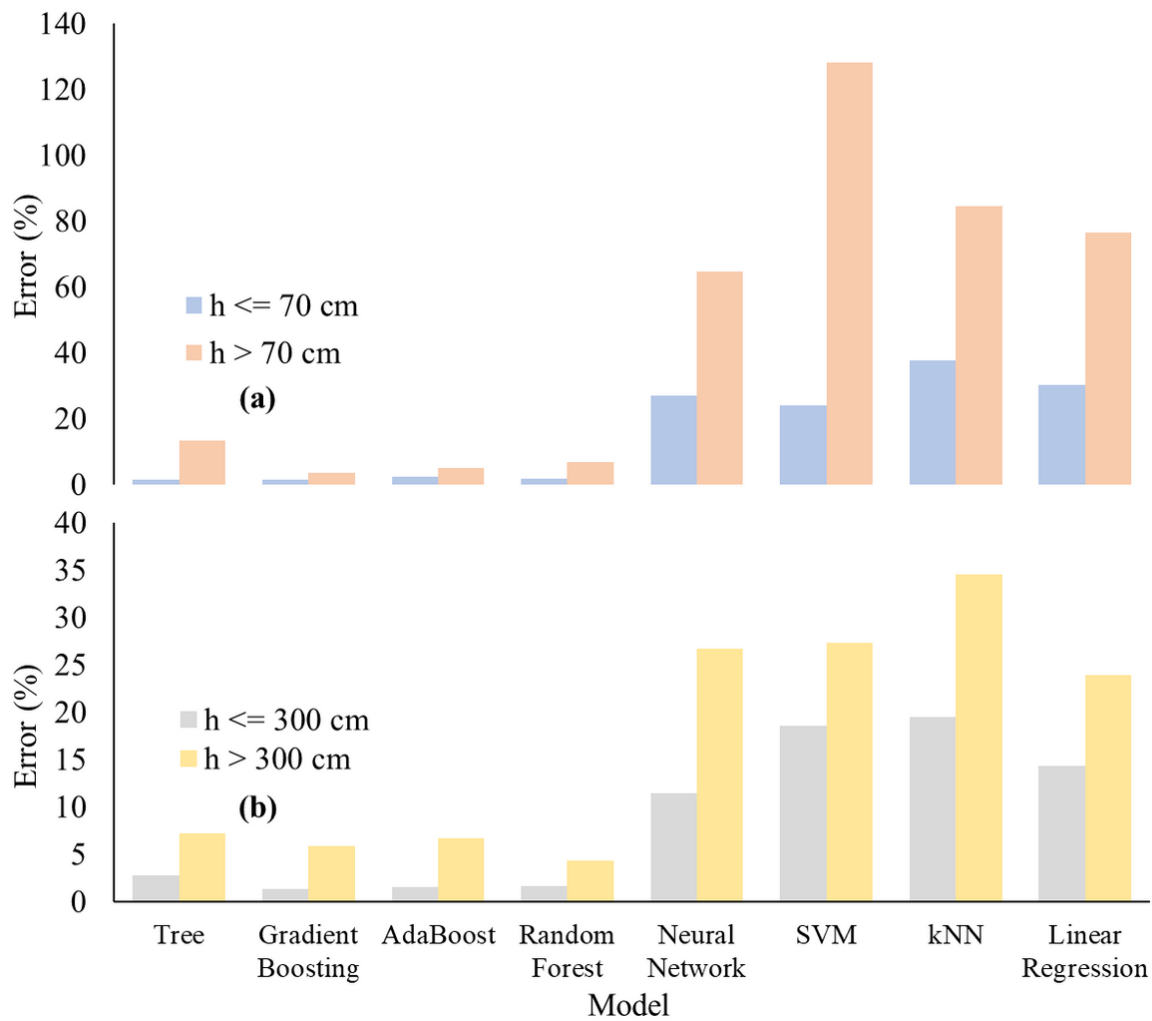


Figure 4
 Comparisons evaluating models error percentages at two side of SWCC inflection point in a) Loamy Sand and b) Silty Clay soil samples

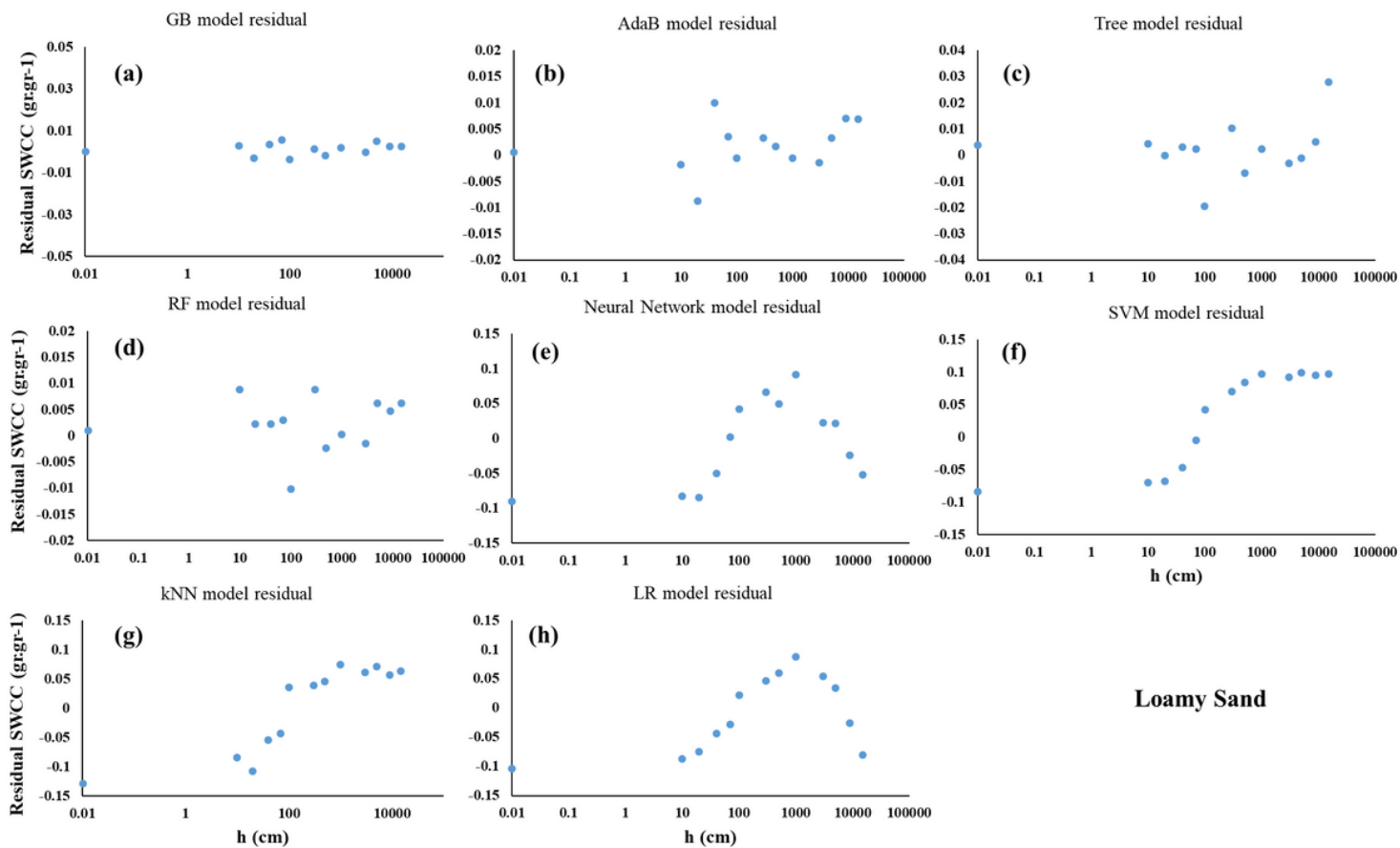


Figure 5

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

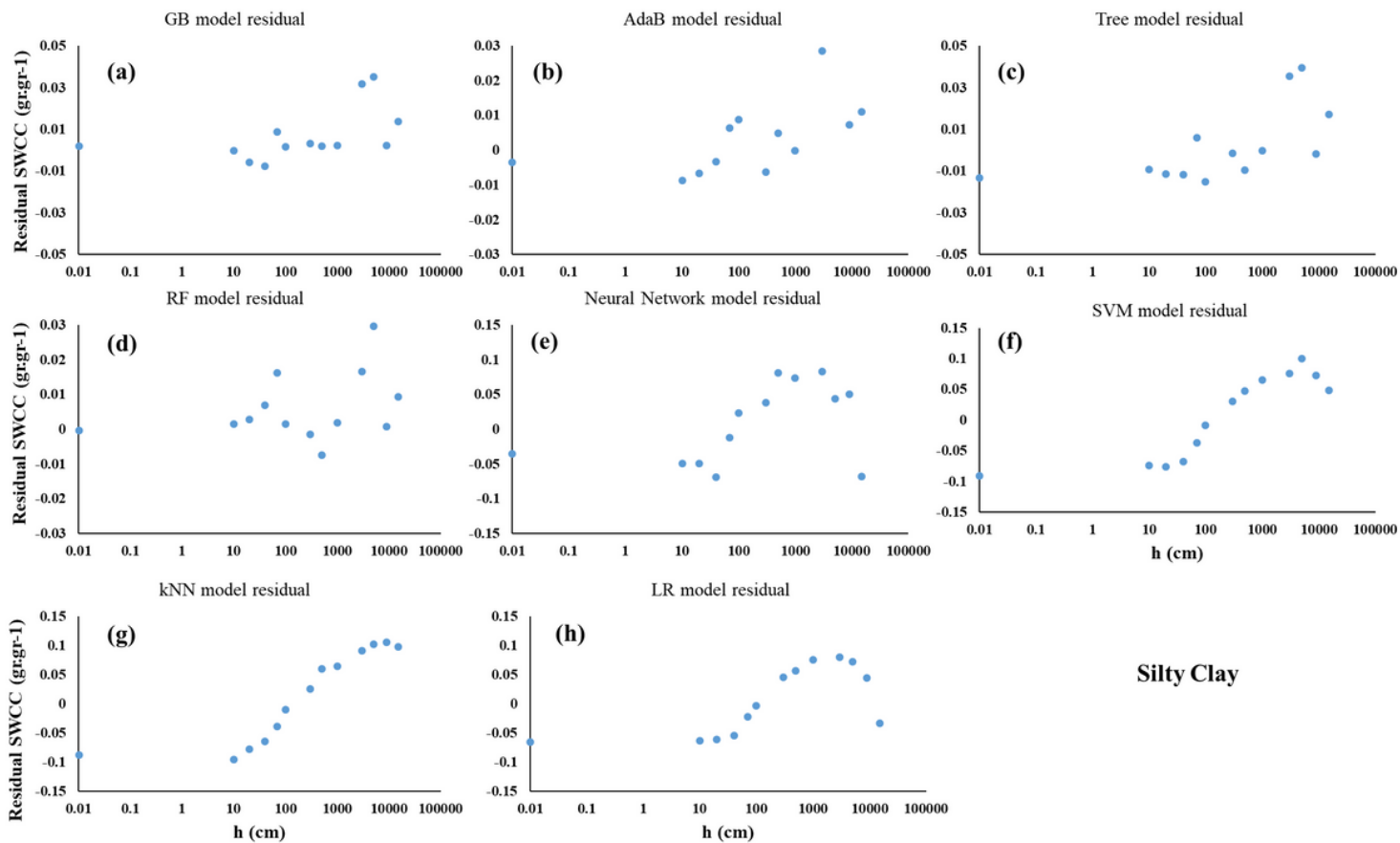


Figure 6

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