Geoscience Frontiers 15 (2024) 101898

Contents lists available at ScienceDirect

Geoscience Frontiers

journal homepage: www.elsevier.com/locate/gsf

Research Paper

Towards an improved prediction of soil-freezing characteristic curve based on extreme gradient boosting model

Kai-Qi Li [a], Hai-Long He [b,c,*]

[a] Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China
[b] College of Natural Resources and Environment, Northwest A&F University, Yangling 712100, China
[c] Department of Soil Science, University of Manitoba, Winnipeg, Manitoba, R3T 2N2 Canada

A R T I C L E I N F O

Article history:
Received 7 March 2024
Revised 22 May 2024
Accepted 29 July 2024
Available online 31 July 2024
Handling Editor: B. Pradhan

Keywords:
Soil freezing characteristic curve (SFCC)
Soil temperature
Unfrozen water content
XGBoost model
Machine Learning
Feature importance

A B S T R A C T

As an essential property of frozen soils, change of unfrozen water content (UWC) with temperature, namely soil-freezing characteristic curve (SFCC), plays significant roles in numerous physical, hydraulic and mechanical processes in cold regions, including the heat and water transfer within soils and at the land–atmosphere interface, frost heave and thaw settlement, as well as the simulation of coupled thermo-hydro-mechanical interactions. Although various models have been proposed to estimate SFCC, their applicability remains limited due to their derivation from specific soil types, soil treatments, and test devices. Accordingly, this study proposes a novel data-driven model to predict the SFCC using an extreme Gradient Boosting (XGBoost) model. A systematic database for SFCC of frozen soils compiled from extensive experimental investigations via various testing methods was utilized to train the XGBoost model. The predicted soil freezing characteristic curves (SFCC, UWC as a function of temperature) from the well-trained XGBoost model were compared with original experimental data and three conventional models. The results demonstrate the superior performance of the proposed XGBoost model over the traditional models in predicting SFCC. This study provides valuable insights for future investigations regarding the SFCC of frozen soils.

1. Introduction

Over 75% of the terrestrial territory in the Northern Hemisphere is subjected to seasonally freezing and thawing cycles (Zhang et al., 1999; Li et al., 2024a; Li and Yin, 2024), which hold significant implications for the engineering, hydrology, ecology, thermodynamics, and biochemistry of soils in cold regions (Cao et al., 2021; Lu et al., 2021). During the freeze–thaw process, the liquid water within frozen soils undergoes phase changes between unfrozen water and ice at a broad temperature range. It is reported that the portion of liquid water remains unfrozen at the subfreezing temperature as low as −70 °C (Stähli et al., 2004). Consequently, as illustrated in Fig. 1, unfrozen water and ice coexist in the pores of frozen soils under the influence of freeze–thaw cycles (Yang et al., 2021; Li et al., 2023a; Feng et al., 2024).

Unfrozen water is closely relevant to soil other properties, such as strength, thermal conductivity, permeability, pore water pressure, matric suction and so on (Li et al., 2020; Li et al., 2023b). Its variation plays a vital role in the stability of engineering constructions and maintenance activities, such as railways, highways, pipelines in permafrost regions, and tunnel projects utilizing artificial ground freezing (AGF) techniques (Yuan et al., 2020; Liu et al., 2022; Li et al., 2024b). Besides, changes in water content can give rise to detrimental engineering phenomena in cold regions like frost heave and thaw settlement. These phenomena, in turn, may lead to the occurrence of geological hazards, such as slope instability, thermal slumping, soil erosion, subgrade frost boiling, differential settlement, and instability of foundation pits (Wen et al., 2011; Yu et al., 2016). Therefore, accurately predicting the unfrozen water content (UWC) at low temperatures is essential for understanding and mitigating the aforementioned engineering and geological issues.

The magnitude of UWC is a crucial hydrothermal property that quantifies frozen soils' capacity to retain and transport liquid water, solutes/salts and heat at a subzero temperature (Wang et al., 2017). The relationship between UWC and subzero temperature (T) is commonly formulated by the soil freezing characteristic curve (SFCC) (Koopmans and Miller, 1966; Spaans and Baker, 1996; Anbergen et al., 2015). The UWC can be quantitatively rep-

* Corresponding author.
E-mail address: hailong.he@nwafu.edu.cn (H.-L. He).

https://doi.org/10.1016/j.gsf.2024.101898
1674-9871/

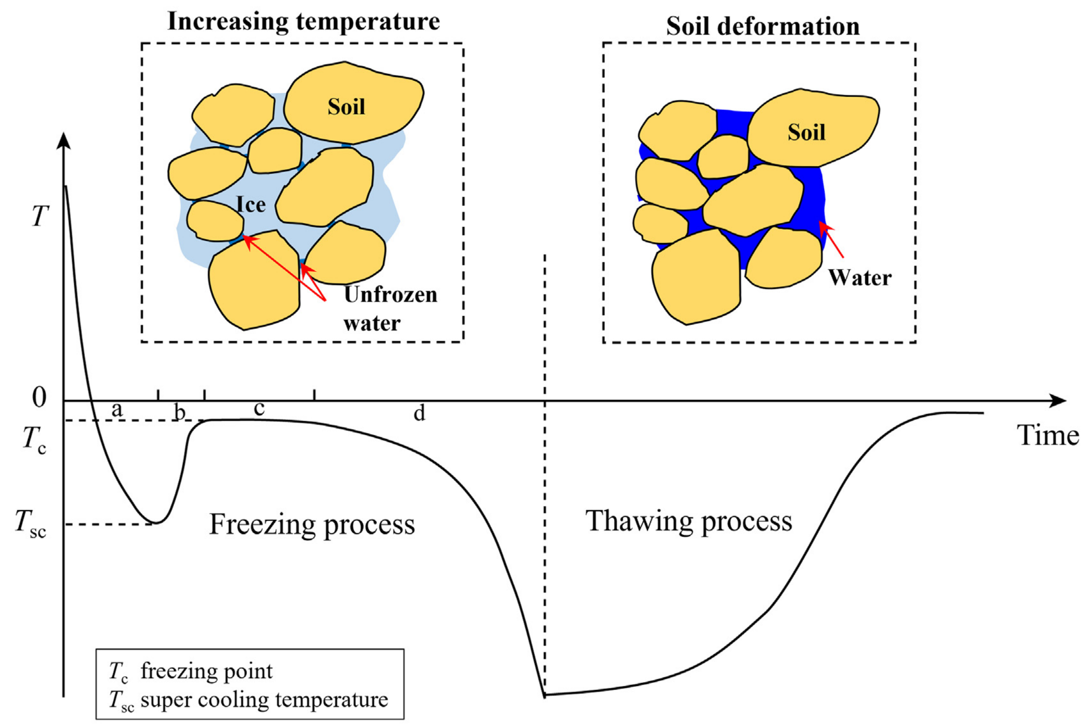

Fig. 1. Schematic diagram of frozen soil and thawed soil under increasing temperature (modified from Shastri et al., 2021; Feng et al., 2024).

resented by gravimetric or volumetric basis or with saturation. A large number of studies have been dedicated to determining the UWC of frozen soils at different subzero temperatures by experimental tests and prediction models. Methods for determining UWC can be classified into five categories, as outlined in a recent review by Feng et al. (2024). These categories are as follows: (i) a pressure-based method, i.e., gas dilatometry, that estimates UWC by measuring the pressure change resulting from approximately 9% volume expansion when water phase transitions to ice (Spaans and Baker, 1996); (ii) radioactive methods, including neutron soil moisture meter, gamma ray probes, and nuclear magnetic resonance (NMR), which rely on the decay of energy following collisions between incident nuclei and hydrogen atoms (He et al., 2023); (iii) electromagnetic methods such as time domain reflectometry (TDR) and remote sensing in L-, S-, and C-bands which rely on discrepancies in the dielectric permittivity of soil constituents (e.g., water with a permittivity of approximately 80 compared to 1 for air, 3 for ice, and approximately 5 for soil solids) (Christ and Park, 2009; Watanabe and Wake, 2009); (iv) thermal methods, including the heat pulse method and thermal dissipation method, offer an estimation of the UWC by leveraging the heat transfer characteristics within the soil (Suzuki and Kitamura, 2008; He et al., 2015); and (v) a sound-based method, specifically ultrasonic technology, relies on the propagation of sound waves for determining UWC (Christ and Park, 2009). Each testing technique possesses distinct advantages and limitations, and their applicability and accuracy are dependent on a number of factors, such as soil type, soil texture (Yoshikawa and Overduin, 2005), salt types and concentration (Robert, 1998; Peng et al., 2022), and external conditions (e.g., T and freeze–thaw history) (Mu, 2017; Zhang et al., 2019).

In addition to experimental measurements, various models have been proposed to efficiently determine the UWC based on the SFCC, which can be categorized into three groups: empirical models, semi-empirical models, and extended soil water retention characteristics curve (SWRC) or SFCC models (i.e., SWRC/SFCC-based model). More detailed introductions of these models are depicted in Section 2. Although numerous models can be applied to simulate the relationship between the UWC and T (e.g., Hu et al., 2020; Bi et al., 2023; Feng et al., 2024) and new models are also being developed (e.g., Bai et al., 2018; Zhou et al., 2018; Wu et al., 2023), most of these models are validated with limited data on UWC for specific soil samples (Table 1). Therefore, a more reliable and generalized prediction model for UWC is still lacking.

Table 1
Summary of existing typical investigations on comparison between predictions from models and experimental measurements.

References	Contrast models	Model types	Number of data points
Bai et al. (2018)	Liu and Yu (2013)	3	47
	Xu et al. (2001)	1	
	Tice et al. (1982)	2	
Chai et al. (2018)	Anderson and Tice (1972)	1	24
	Michalowski (1993)	1	
	Xu et al. (2001)	1	
	Kozlowski (2007)	2	
Wan et al. (2022)	Anderson and Tice (1972)	1	61
	Michalowski (1993)	1	
	Dall'Amico et al. (2011)	3	
	Liu and Yu (2013)	3	
	Zhang et al. (2017)	1	
Ren et al. (2023)	Anderson and Tice (1972)	2	48
	Liu and Yu (2013)	3	
Wan et al. (2023)	Anderson and Tice (1972)	1	71
	Kozlowski (2007)	2	
	Liu and Yu (2013)	3	

Notes: model type 1 is empirical model; 2 is semi-empirical model; 3 is SWRC/SFCC-based model.

Table 2
Summary of existing models for determining unfrozen water content.

No.	Model types	Features	References
1	Empirical models	Derived by fitting experimental data	e.g., Anderson and Tice (1972); Michalowski (1993); Osterkamp and Romanovsky (1997); Riseborough (2004); McKenzie et al. (2007); Zhang et al. (2008); Qin et al. (2009); Westermann et al. (2011); Kurylyk and Watanabe, (2013); Nicolsky et al. (2017); Zhang et al. (2017); Lu et al. (2021); Weng et al. (2021)
2	Semi-empirical models	Predicting UWC by other soil properties (e.g. SSA, PI, PSD or microstructure), or based on Clapeyron equation	e.g., Anderson and Tice (1972); Cahn et al. (1992); Dash et al. (1995); Kozlowski (2007); Zhang et al. (2007); Bai et al. (2018); Kong et al. (2020); Chai et al. (2018); Li et al. (2020); Teng et al. (2021); Li et al. (2023f); Li and Li (2023)
3	SWRC/SFCC-based models	Derived by combining Clapeyron equation and SWRC or modified SWRC (suction is replaced by temperature)	e.g., Shoop and Bigl (1997); Bittelli et al. (2003); Niu and Yang (2006); Nishimura et al. (2009); Dall'Amico et al. (2011); Sheshukov and Nieber (2011); Watanabe et al. (2011); Liu and Yu (2013); Ren et al. (2018); Luo et al. (2009); Wen et al. (2020); Fu et al. (2021)

Notes: specific surface area (SSA); plastic index (PI); soil–water retention curve (SWRC); particle size distribution (PSD).

The UWC can be influenced by various internal factors (e.g., particle size distribution, salinity, clay content, plastic index, specific surface area (SSA)) and external factors (e.g., freeze–thaw cycles, stress state). However, it is challenging to find abundant investigations incorporating all influencing factors. Furthermore, the complex impacts of these factors are not well understood and result in highly nonlinear relationships with UWC, which makes it challenging to identify the most relevant factors for a more reliable UWC prediction model. To overcome these limitations, machine learning (ML) methods offer a promising solution, which have been successfully applied to soil science and hydrology (Wang et al., 2023), such as soil mapping (Wadoux et al., 2020), soil organic carbon prediction (Heuvelink et al., 2021), assessment of soil properties and behaviours (Li et al., 2023c,d,e; Li et al., 2024c). ML methods are appealing for SFCC assessment due to their key benefit of not requiring prior knowledge of relations between input and output variables. Unlike (semi-)empirical or SWRC/SFCC-based models, ML methods do not necessitate simplifications or assumptions, making them more suitable for accurate UWC predictions. In addition, ML methods can continuously improve their performance by updating with new training examples as available fresh data (Bergen et al., 2019).

Among the various ML methods, the extreme Gradient Boosting (XGBoost) model stands out as an effective supervised learning model for solving regression issues due to its high prediction performance (Chen and Guestrin, 2016). Furthermore, it also possesses dominant advantages over other models, notably its exceptional ability to extract complex nonlinear relations between inputs and outputs from a large amount of data, handle the problem of missing data, and identify the most influencing factors. The XGBoost model has been successfully used to solve manifold issues, such as soil pH prediction, arsenic concentration and liquefaction potential assessment of soils (Tziachris et al., 2020; Ye et al., 2023). To the authors' knowledge, no previous investigations have been conducted on applying the XGBoost model to estimate the SFCC.

Therefore, the objective of this study was to establish an XGBoost model to predict the SFCC of frozen soils. A systematic SFCC database of frozen soils compiled from existing experimental studies by Devoie et al. (2022) was used to train and validate the XGBoost model. In addition, the grid search combined with 5-fold cross-validation is used to tune the hyperparameters and further improve the accuracy of the prediction model. To demonstrate the effectiveness of the XGBoost model, the predicted SFCC from the XGBoost model are compared with calculated results from three commonly used traditional models. It is hoped that this study will provide valuable insights for future studies related to the prediction of UWC in frozen soils via machine learning methods and contribute to a deeper understanding of the freeze–thaw process in frozen soils.

2. SFCC prediction models for comparison

Table 2 presents a comprehensive overview of different models for predicting UWC of frozen soils that are categorized into three types: (i) empirical models, which are derived by fitting experimental UWC and temperature data and often have relatively simple formulations without physically meaningful parameters; (ii) semi-empirical models, on the other hand, indirectly calculate the UWC using other soil properties such as plastic index and SSA or based on the pore size distribution and solid fabrics' microscopic structural arrangements. Besides, they can also be developed based on the Clapeyron equation quantitatively analyzing the phase equilibrium relationship (e.g., ice and water) in a single subsystem at different temperatures, which often possess complicated expressions; (iii) SWRC/SFCC-based models. The accuracy of these models depends on the similarity between the SWRC and SFCC. However, the underlying mechanism of SFCC and SWRC display significant dissimilarity, e.g., how ice replaces water in frozen soils differs from air replacing water in unsaturated soils (He and Dyck, 2013). Moreover, the presence of salinity further complicates these dissimilarities (Zhou et al., 2018; Ren and Vanapalli, 2019). Unlike existing prediction models, this work primarily establishes a data-driven model for evaluating UWC with the aid of machine learning methods, which can address the limitations of previous prediction models.

To evaluate the performance of the proposed XGBoost model against conventional models and validate its potential advantages, three commonly employed models are selected for comparison: two empirical models and one semi-empirical model. It is worth noting that SWRC/SFCC-based models are not included in this comparison due to their complex implementation and the unavailability of requisite parameters in our compiled database. The first model is a typical empirical model developed by Anderson and Tice (1972), which is given as Eq. (1).

$$\theta = a|T|^b \text{ (Model 1)} \tag{1}$$

where θ is volumetric UWC (m^3/m^3), T is temperature (°C), a and b are fitting parameters. The second empirical model is proposed by Kurylyk and Watanabe (2013) who suggested that UWC can be formulated as an exponential function of T.

$$\theta = \theta_{\min} + (\theta_{\text{total}} - \theta_{\min})exp\left(-(T/a)^2\right) \text{(Model 2)} \tag{2}$$

where $\theta_{\min}$ is the minimum volumetric UWC (m^3/m^3), θ_{total} is the total volumetric UWC. Given the report that the unfrozen water still exists in frozen soils at −70 °C (Williams, 1964), the value of θ_{thre} is equal to −70 °C herein.

The third model for comparison is a widely-used semi-empirical model proposed by Anderson and Tice (1972), which is adopted herein since it involves one input parameter (e.g., SSA). Specifically, Anderson and Tice (1972) model is characterized by a power function that involves two parameters (T and SSA) to determine the gravimetric UWC.

$$lnw_u = 0.2618 + 0.5519lnS - 1.449S^{-0.264}\ln(-T)(\text{Model 3}) \tag{3a}$$

where S is SSA (m^2/g); w is gravimetric UWC (g/g), and it can be converted to θ by

$$\theta = \rho_d w/\rho_w \tag{3b}$$

where ρ_d is dry density (g/cm^3), ρ_w is water density (=1 g/cm^3).

3. Database and methodology

3.1. Database

SFCC is a widely used proxy to represent the UWC within frozen soils, which relates to the UWC and specific subzero temperatures of soils at frozen conditions. Theoretically, the UWC is a monotonic function of temperature, but strong hysteresis exists between freezing and thawing cycles (Tian et al., 2014), especially in fine-grained soils. For simplicity, experimental data of each SFCC study were collected only along the freezing limb, i.e., the hysteretic effect is not considered during the database compilation.

Accordingly, a systematic database of the SFCC of frozen soils is compiled from existing experimental studies by Devoie et al. (2022), including UWC and other measured soil properties. The measured UWC serves as the output variable for the XGBoost model. In addition to temperature, another three variables (i.e., dry density, initial unfrozen water content and SSA) are also selected as input variables, as they are widely recognized as critical factors directly impacting the UWC in many studies (e.g., Yoshikawa and Overduin, 2005; Watanabe and Wake, 2009; Akagawa et al., 2012; Chai et al., 2018; Ren et al., 2023). The dry density plays a crucial role in the capillary behaviour of soil pore systems and influences the amount of unfrozen water (Tian et al., 2014; Ren et al., 2023), which is also helpful in identifying and categorizing soils when other information is unavailable (Devoie et al., 2022). As reported by some scholars (Nersesova and Tsytovich, 1963; Anderson and Tice, 1972), the amount of UWC is highly dependent on SSA. Fig. 2 depicts the SFCC obtained from the compiled database and provides the probability density and histograms of the data distribution.

Table 3 provides a comprehensive overview of the compiled database, including the relevant literature, soil types, testing methods and specific numbers of SFCC. The database comprises a total of 1252 data points derived from 71 SFCCs, which have been extracted from 22 scientific studies. It is important to note that only five investigations within the dataset provide information regarding SSA. Table 3 also indicates that the UWC of various types of frozen soils were obtained by measurements using NMR, pNMR, TDR, FDR, calorimetry, dilatometry, and probes. Each of these test methods has its inherent limitations and potential sources of error. Factors such as equipment variations, calibration procedures, and measurement protocols can also contribute to differences in the obtained UWC values. To establish a comprehensive database, the UWC data were collected regardless of the specific testing methods employed. Furthermore, the values of gravimetric water content were converted to volumetric water content by multiplying them with soil dry density for consistency. Therefore, it can be concluded that the compiled database is representative and systematic for training the XGBoost model to estimate the UWC.

3.2. XGBoost model

The ensemble learning methods have gained significant popularity in addressing regression problems, as they leverage the combination of multiple machine learning techniques or multiple instances of the same technique to achieve enhanced predictive accuracy by aggregating individual model results. In general, ensemble learning methods are separated into bagging and boosting. Bagging is a method that aims to improve accuracy and reduce variance by randomly sampling the training dataset and computing the final output by averaging the results of the individual models. In contrast, the boosting method iteratively combines weak learners with strong learners by emphasizing the mispredicted training cases. Various boosting models are available to be applied for regression issues, such as gradient boosting, XGBoost, adaptive boosting (AdaBoost), etc.

Initially proposed by Chen and Guestrin (2016), the XGBoost model has emerged as an advanced supervised model that has gained widespread recognition in the machine learning competitions of Kaggle and has been applied by many scholars to address the complicated prediction challenges in engineering fields, such as soil properties mapping (Zhao et al., 2022), slope stability landslide (Gu et al., 2023; Yang et al., 2024), strength evaluation prediction (Feng et al., 2021; Zhang et al., 2023). The XGBoost model offers notable advantages, including its ability to handle large datasets with high computational efficiency and its capacity to handle complex relationships.

As a decision tree (DT)-based model, the XGBoost model utilizes the parallel DT model to solve the regression issues and incorporate additional regularization terms into the loss function to avoid overfitting. Notably, the XGBoost model can effectively handle missing data by automatically learning how to handle missing values during training. The objective function of the XGBoost model consists of two parts, i.e., conventional loss function and model complexity (see Eqs. (4) and (5)), which can be employed to represent the operational efficiency of the model.

$$Obj(\theta) = \sum_{i=1}^{n}(y_{pi},\ y_{ei}) + \sum\Omega(f_k) \tag{4}$$

$$\Omega(f_k) = \gamma T + 0.5\lambda\|\ w\ \|^2 \tag{5}$$

where θ is UWC (m^3/m^3); n is the number of data points fed into the k^{th} tree; y_{pi} and y_{ei} are i^{th} prediction and original experimental value of i^{th} data point, respectively. f_k is the k^{th} tree in the XGBoost model; λ and γ are penalty coefficients, which are used for controlling the regularization. T is the number of leaf nodes, and w is the score at each node. For a more comprehensive understanding of this model, in-depth explanations can be found in the work by Chen and Guestrin (2016).

3.3. Prediction framework

This section depicted a detailed prediction framework of the data-driven model for constructing a straightforward prediction model for UWC. Specifically, the implementation of the XGBoost model in predicting the UWC includes three phases: data pre-processing, hyperparameter optimization, and model evaluation, as shown in Fig. 3.

Step 1 Data pre-processing

The compiled database on UWC includes three input parameters (temperature, dry density, initial unfrozen water content and SSA) and one output (UWC). Before performing the XGBoost model,

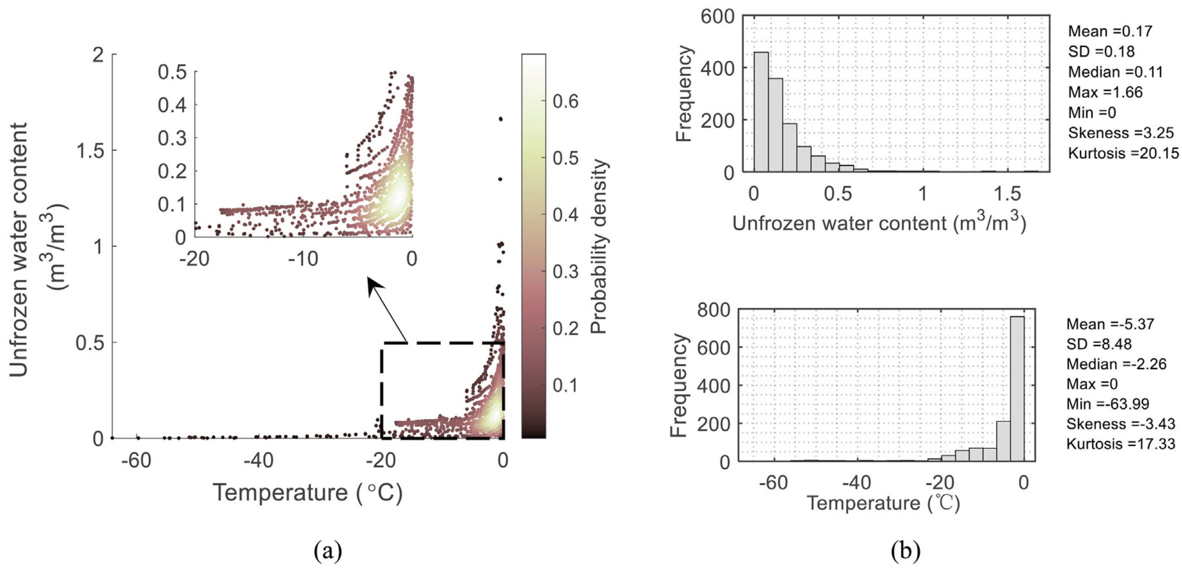

Fig. 2. (a) Soil freezing characteristic curve (SFCC) from the compiled database of Devoie et al. (2022). The color of the scatter represents the data probability density, and (b) histograms of temperature and unfrozen water content in the compiled database.

Table 3
Summary of the compiled database for measured unfrozen water content (UWC) based on Devoie et al. (2022).

No.	References	Soil types	Test methods	No. of SFCC	No. of measurements
1	Akagawa et al. (2012)	Kaolin, clay, volcanic ash, soft mudstone	pNMR	4	36
2	Anderson and Tice (1972) *	Silt, limonite, gravel, kaolinite	Isothermal calorimeter	4	52
3	Black and Tice (1989)	Sandy loam	pNMR	2	28
4	Bouyoucos (1921)	Clay loam	Empirical method	1	12
5	Buehrer and Rose (1943)	Clay, sandy loam, loam	Dillatometry	5	37
6	Christ and Park (2009)	Sand	TDR	1	12
7	Jame and Norum (1972)	Silical flour and montmorillonite	Isothermal calorimeter	1	16
8	Koopmans and Miller (1966)	Silt	Dillatometry	2	18
9	Nakano et al. (1982)*	Morin clay	–	1	17
10	Nersesova and Tsytovich (1963)	Loam, quartz, sandy loam	Calorimetry	3	13
11	Oliphant et al. (1983)*	Morin clay	–	1	14
12	Patterson and Smith (1981)	Silt loam	TDR	1	12
13	(Pusch, 1979)*	Clay	Dillatometry	4	24
14	Spaans and Baker (1996)	Silt loam	TDR	2	16
15	Stähli and Stadler (1997)	Loam, sand	TDR	4	130
16	Tice et al. (1982)	Submarine soil	NMR	1	14
17	Watanabe and Osada (2017)	Sandy loam	TDR	1	10
18	Williams (1964)	Clay, silt	Calorimetry	5	126
19	Wu et al. (2015)	Silt loam	TDR	3	107
20	Wu et al. (2021)	Clay	FDR	4	162
21	Yoshikawa and Overduin (2005)*	Silt and clay	NMR, one TDT (CS615 probe), three FDR (theta probe, Vitel probe, ECH2O probe)	9	151
22	Yuan et al. (2020)	In situ soil sample along Qinghai–Tibet Highway	CS616 water content reflectometer	11	245

Notes: In total, 1,252 data points were involved in the compiled database. Soil freezing characteristic curve (SFCC); nuclear magnetic resonance (NMR); pulsed nuclear magnetic resonance (pNMR); time domain reflectometry (TDR); frequency domain reflectometry (FDR). The references marked with asterisk indicate that the information on SSA is included in the corresponding studies.

the entire database is stochastically split into a training dataset (80%) and a testing dataset (20%). All the inputs and outputs in the compiled database are mapped to the interval (−1, 1) using a data normalization strategy to reduce the computation cost and eliminate the influence of feature scaling.

Step 2 Hyperparameter optimization

Hyperparameters are parameters in machine learning models that need to be predetermined before the model training process, which can significantly impact the prediction accuracy of the XGBoost model. Hyperparameter optimization is a crucial step in the model training process to improve the model's performance and accelerate the convergence speed, which can be realized by

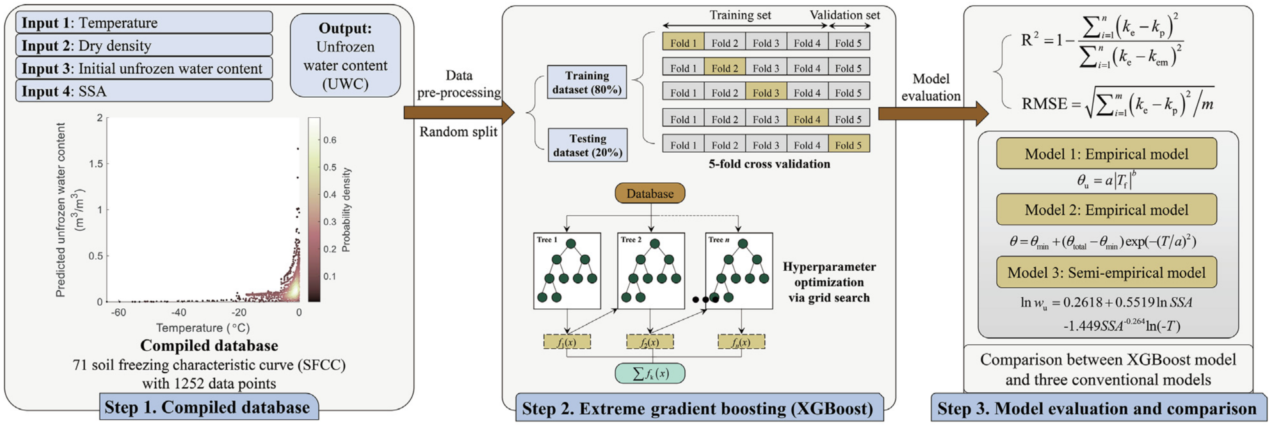

Fig. 3. Prediction framework of XGBoost model for estimating soil freezing characteristic curve (SFCC, unfrozen water content as a function of temperature) in frozen soil.

grid search. Grid search is commonly employed to tune the hyperparameters of machine learning models by searching a suitable set of hyperparameter configurations and testing each combination of all hyperparameters. The ultimate tuned hyperparameters of the XGBoost model are summarized in Section 4.1. *N*-cross validation (i.e., 5-fold cross-validation herein) is conducted to evaluate the performance of each candidate dataset and avoid the overfitting problem. In the 5-fold cross-validation, the training dataset is divided into five equal subsets, where four subsets are used as training sets, and the remaining one subset serves as a validation set. After performing the training-validation procedure five times, the average of predictions for UWC is compared among the XGBoost models with different combinations of hyperparameters. The best-performing XGBoost model is retained during the stage of model evaluation.

In order to enhance the prediction accuracy and convergence speed of the proposed XGBoost model for estimating UWC, the hyperparameters of XGBoost are first optimized by grid search. The specific tuning process involves constructing the XGBoost model to evaluate the UWC on the training set, where the mean square error (MSE) is the evaluation indicator. The particular equation of MSE is shown as follows.

$$\mathrm{MSE} = \sum_{i=1}^{m} (\theta_e - \theta_p)^2 / m \tag{6}$$

where θ_e is measured UWC (m^3/m^3); θ_p is the estimated UWC (m^3/m^3); m is the number of data points. When the MSE converges steadily to its minimum value during the training process, the optimal value for one hyperparameter can be obtained while keeping the other hyperparameters fixed. This procedure is repeated to obtain the optimal values for the remaining hyperparameters.

Step 3 Model evaluation

Once the hyperparameters are optimized, the XGBoost model with optimal configuration can be conducted on the training and testing sets. To determine the prediction performance of XGBoost model, the following model evaluation indicators are used herein, including coefficient of determination (R^2) and root mean square error (RMSE).

$$R^2 = 1 - \frac{\sum_{i=1}^{m} (\theta_e - \theta_p)^2}{\sum_{i=1}^{m} \left(\theta_e - \bar{\theta}_e\right)^2} \tag{7}$$

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{m} (\theta_e - \theta_p)^2 / m} \tag{8}$$

where $\bar{\theta}_e$ is mean value of measured UWC (m^3/m^3). A higher value of R^2 and a lower value of RMSE indicates that the model can yield more accurate predictions. To further validate the predictive effectiveness of the proposed data-driven model, the predicted SFCC from the XGBoost model are compared with those computed by some traditional models. To quantify the prediction performances of the XGBoost model and conventional models, a model rank index (i.e., model rank, MR) is also introduced.

$$\mathrm{MR} = \sum_{i=1}^{k} rank_i / k \tag{9}$$

where $rank_i$ is the ordering of each prediction model, and k refers to the number of model evaluation indicators. The model with the lowest value of MR can be regarded as the best-performing model (Li et al., 2022a,b).

4. Results and discussion

4.1. Prediction performance of XGBoost model

Table 4 provides a summary of the optimal hyperparameters of XGBoost tuned by grid search, along with the corresponding interpretations for each hyperparameter. Besides, Fig. 4 provides a comparison between the original experimental data and the predicted values of UWC for both the training set and the testing set. It is evident that the predicted UWC values align closely with the experimental data in the training set, exhibiting a strong correlation along the 1:1 line. In contrast, the predictions generated by the XGBoost model for the testing set also demonstrate a reasonable agreement with the measured UWC, albeit with a few outliers that underestimate the observed UWC values. Besides, Fig. 4 includes histograms depicting the distribution of measured and predicted UWC values for the training and testing sets. These histograms reveal that most UWC values fall below 0.5 m^3/m^3 in the training and testing sets.

To visualize the dependencies between measurements and predictions for SFCC (as outlined in Table 3) from 22 studies, Fig. 5 shows the comparison between the original SFCC from laboratory tests and the predicted SFCC from the proposed XGBoost model. Generally, the estimations derived from the XGBoost model demonstrate a strong alignment with the actual SFCC, which indicates that the proposed XGBoost model can efficiently capture the features of SFCC. Specifically, this model was trained using instances from the training set, enabling accurate predictions on instances that were included in the testing data. After completing the training phase, the model achieves a considerable level of accu-

Table 4
Optimal hyperparameters of the XGBoost model.

Hyperparameters	Ranges	Optimal values	Remarks
booster	Decision tree	Decision tree	The base learner in boosting technique
eta	[0.1, 1]	0.6	To reduce the weights and avoid overfitting
n_estimators	[20, 120]	90	Number of weaker learners
max_depth	[3, 20]	4	Maximum depth of tree. A higher value leads to the complexity of the model and adds the overfitting probability.
mini_child_weight	[1,10]	1	Minimum sum of instance weight in a child.
max_delta_step	[1, 10]	1	Maximum delta step that allows each leaf output to be.
subsample	(0, 1]	0.7	A fraction of observations will be randomly sampled for each tree.
colsample_bytree	[0.5, 1]	1	The fraction of columns is to be a random sample for each tree.
gamma	[0]	0	Minimum loss reduction required to make a split.
reg_lambda	[1]	1	L_2 regularization term on weights
reg_alpha	[0]	0	L_1 regularization term on weights

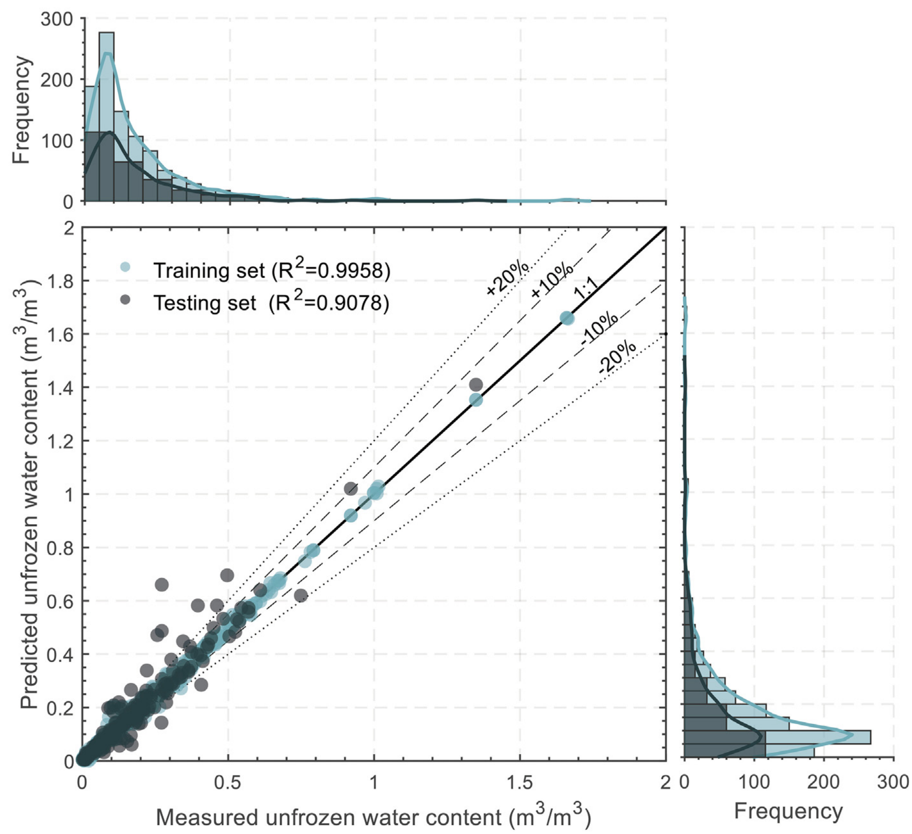

Fig. 4. Comparison between experimental data and predicted results for the training set and testing set.

racy on the testing set, as observed in Fig. 4, where the predicted UWC in the testing set generally aligns closely with the 1:1 line, with only a few outliers. Notably, it can be noted in Fig. 5 that most of the observed deviations (represented by solid symbols) primarily originate from the testing set, implying a slightly inferior prediction performance of the XGBoost model on unseen data compared to the training set. This phenomenon further underscores the limited generalization capability of the XGBoost model.

4.2. Comparison with conventional models

To validate the predictive capacity of XGBoost model for determining UWC of frozen soils., three conventional models are selected. Fig. 6 provides a summary of the prediction results obtained from the three traditional models and the XGBoost model for four different soil types across existing experimental studies (Mu, 2017; Zhou et al., 2018; Kong et al., 2020; Li et al., 2020). The SSA values for four soils are 75.3 m^2/g (loess), 380.6 m^2/g (Bentonite), 3 (fine sand) and 16.6 m^2/g (silt), respectively. It is noted that the dataset used for comparison is excluded from the dataset for training the XGBoost model. Therefore, this dataset is suitable for comparison, and this comparison can also be used to further verify the robustness and applicability of the proposed XGBoost model for the assessment of UWC. As shown in Fig. 6, apart from Bentonite, Model 1 (an empirical model developed by Anderson and Tice (1972)) can track the basic trend of UWC with decreasing

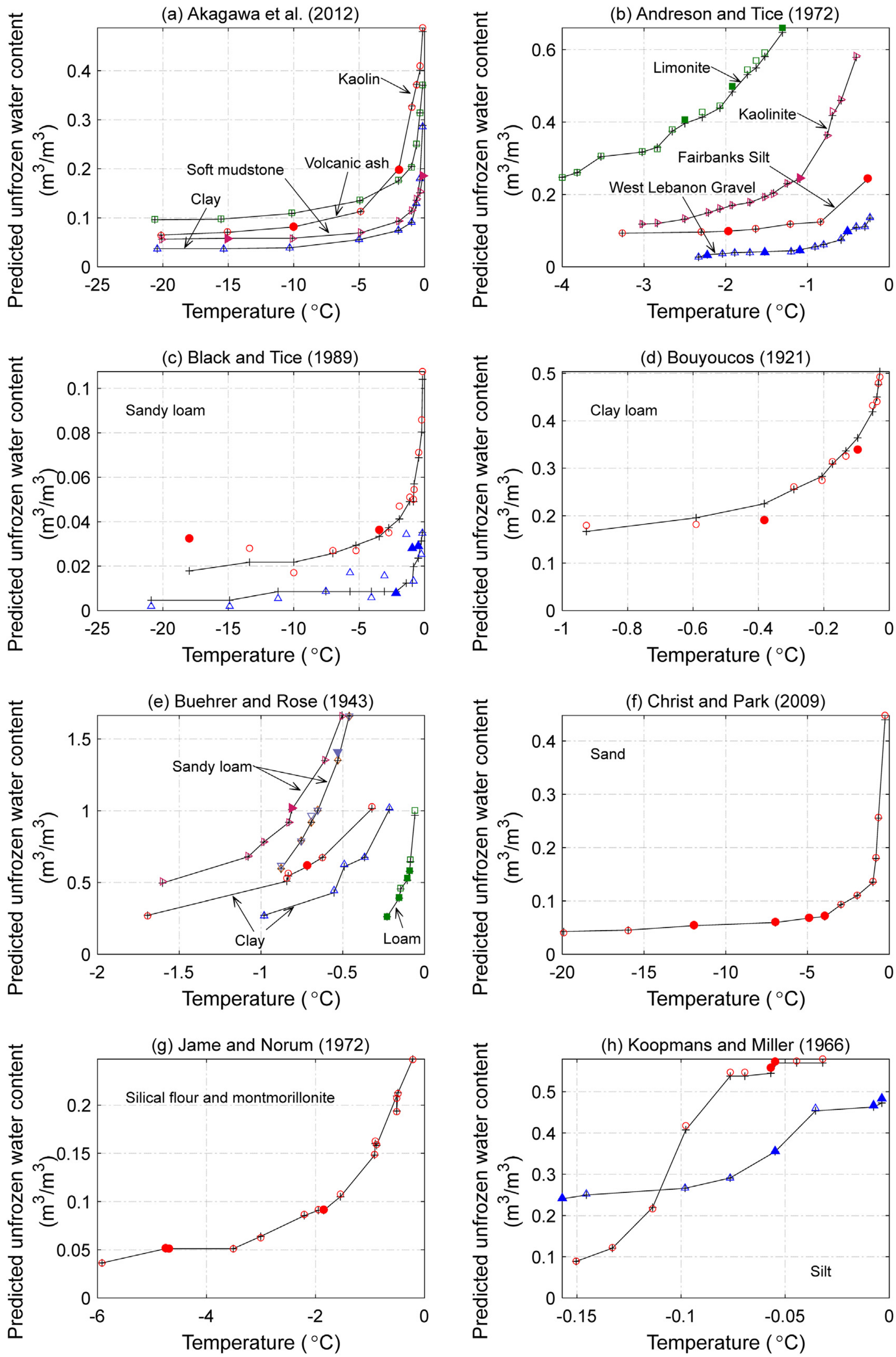

Fig. 5. Comparison between the original measured soil-freezing characteristic curve (SFCC) and predicted SFCC by the XGBoost model. The black cross symbols represent the original data from the complied database for unfrozen soils; colored hollow symbols and solid symbols denote the predicted unfrozen water content (UWC) of soil samples in the training set and testing set, respectively.

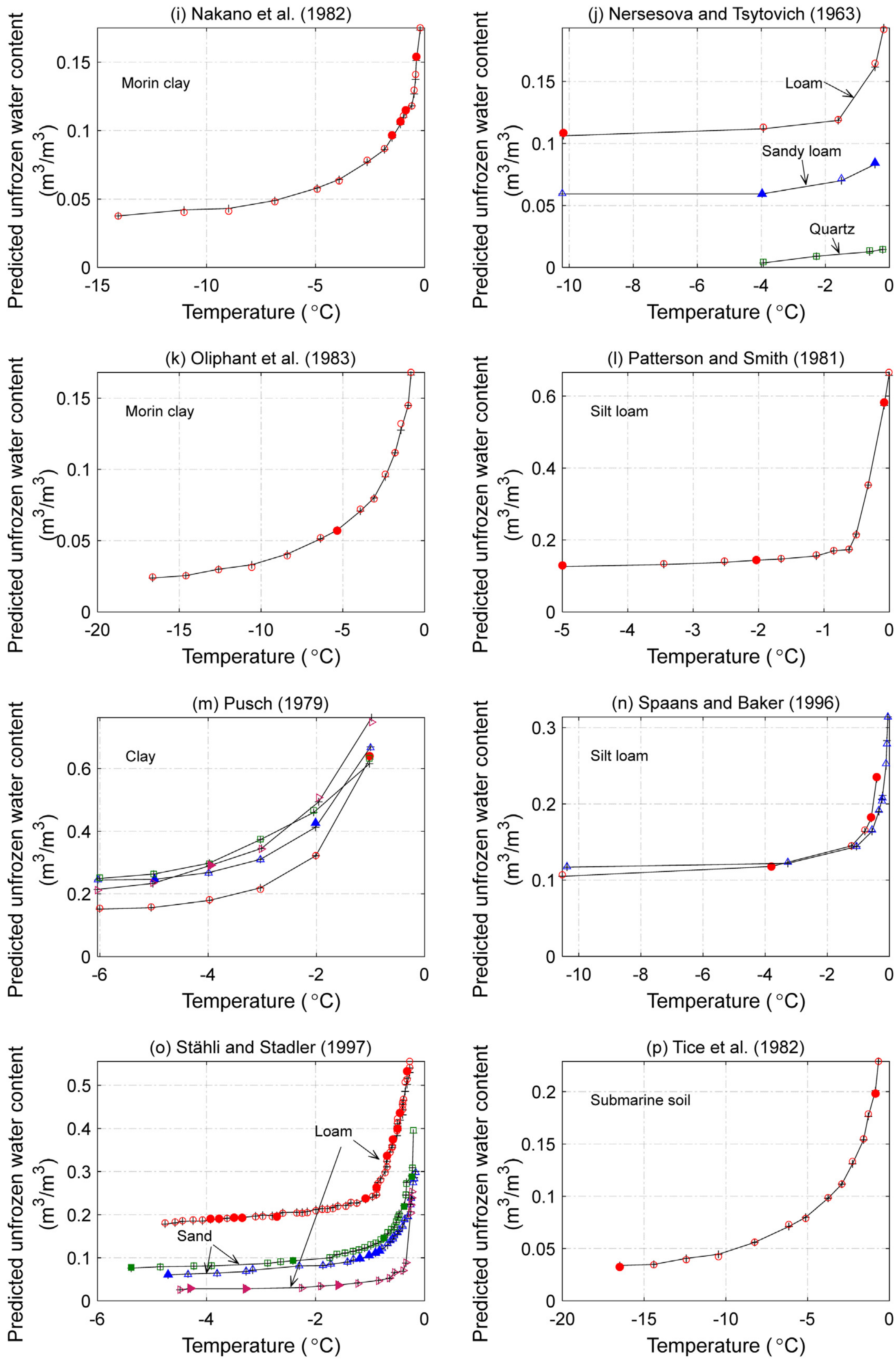

Fig. 5 (*continued*)

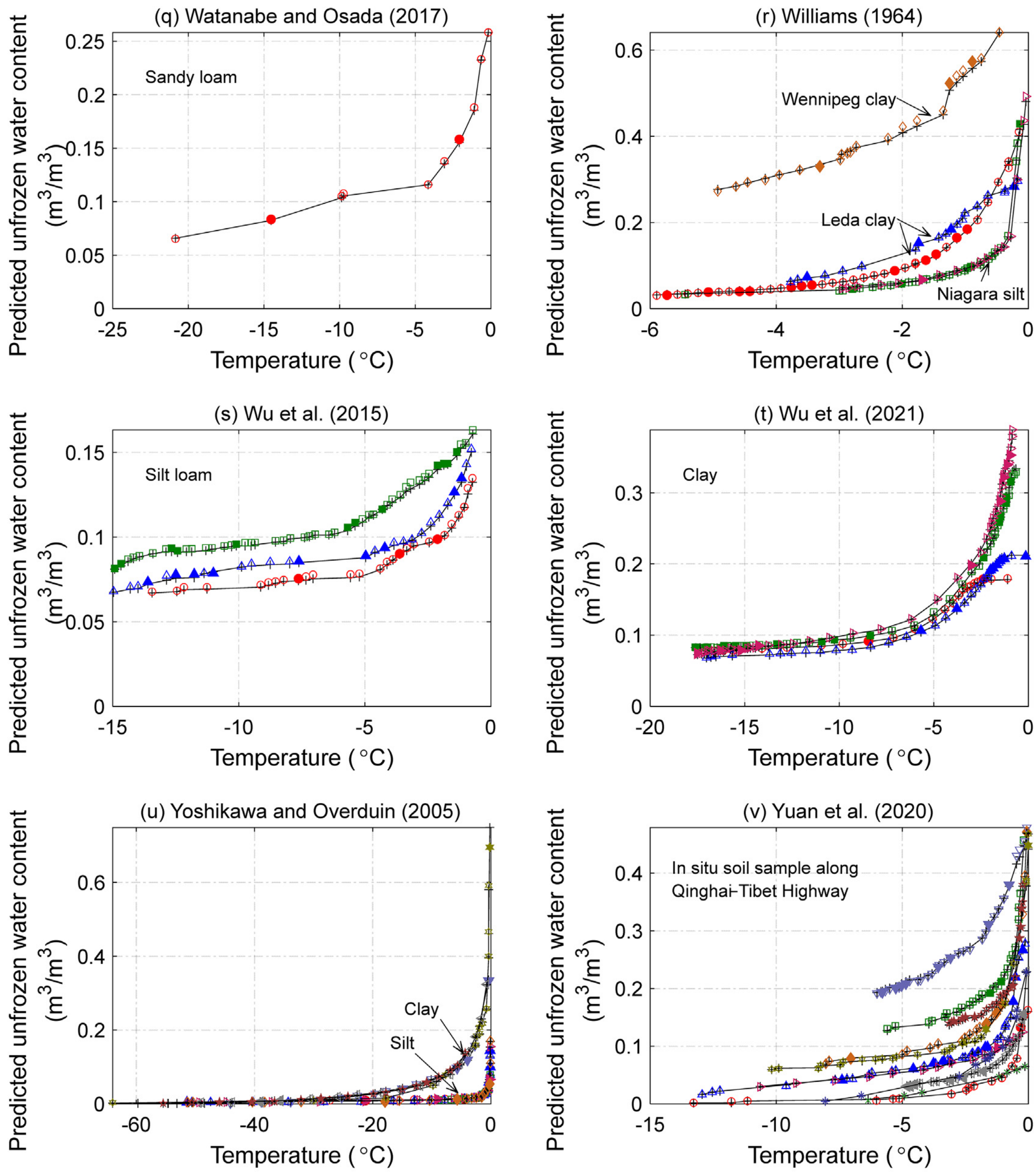

Fig. 5 (*continued*)

temperature. Although it involves another input (i.e., SSA), Model 3 (a semi-empirical model from Anderson and Tice (1972)) yields overestimations for UWC for plastic soils (i.e., loess, bentonite and silt) and underestimates the UWC for fine sand. Considering the total UWC and minimum UWC, Kurylyk and Watanabe (2013) proposed an exponential function of *T* to determine UWC. It can be noted from Fig. 6a–c, the prediction performance of Model 3 is superior to that of two other models, as its R^2 values are lower than those of Model 1 and Model 3. It is evident from Fig. 6 that the XGBoost model can approximately reproduce the relationship between temperature and UWC for loess, bentonite, silt and fine sand. The majority of predictions by the XGBoost model closely align with the measured UWC. As for the prediction performance of each model, the XGBoost model outperforms the three traditional models in simulating the SFCC, as indicated by the lowest RMSE values. Therefore, it can be concluded that the XGBoost model demonstrates promising performance in assessing UWC values compared to the traditional models.

Fig. 7 illustrates the comparison between three traditional models and the XGBoost model in predicting UWC. The results demonstrate that the three conventional models are inadequate for accurately estimating UWC for the soil types in the four studies. In general, most of the predicted UWC from Model 2 (an empirical model by Kurylyk and Watanabe (2013)) lie within the ±20% error lines. A similar comparison can be observed between predictions from Model 1 (empirical model relating to *T*) and measured UWC, as shown in Fig. 7a. However, the deviations from Model 1 are more pronounced since some UWC values predicted by this model are lower than experimental data. The UWC values predicted by Model 3 (a semi-empirical model incorporating *T* and SSA) deviate significantly from the 1:1 line, with most of the data falling outside the ±20% error lines. In other words, three tradi-

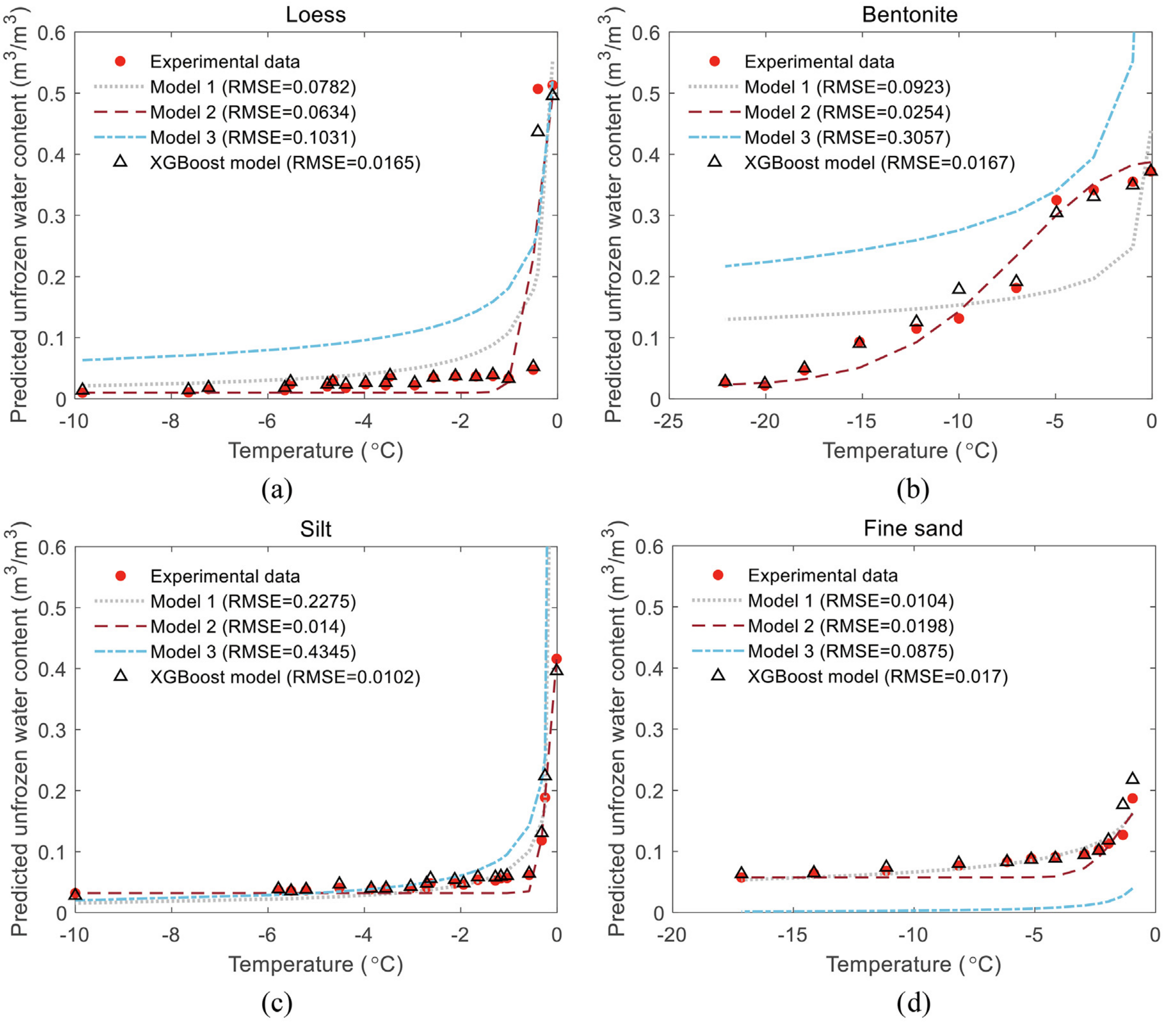

Fig. 6. Comparison between the XGBoost model and three traditional models (a) Mu, 2017; (b) Kong et al., 2020; (c) Zhou et al., 2018; (d) Li et al. (2020).

tional models either overestimate or underestimate the UWC of four types of soils. This conclusion aligns with the findings reported by Ren et al. (2023) and Wang et al. (2023), which highlight the relatively poor prediction accuracy of the Anderson and Tice (1972) model.

In contrast, the UWC values predicted by the XGBoost model generally exhibit a good agreement with the measured data, as depicted in Fig. 7d. This comparison indicates that the proposed XGBoost model outperforms the three traditional models in terms of its superior prediction accuracy. Furthermore, Table 5 presents the ultimate results of model evaluation indicators and MR (model rank index, see Eq. (9)) for the XGBoost model and three traditional models, which reinforces that the XGBoost model exhibits the best performance with the highest rank for assessing the UWC.

4.3. Features importance analysis

Feature importance is vital for feature selection and model interpretability. The trained XGBoost model provides an automated way to calculate the importance of features. The importance of the feature can be obtained using the gain criterion within XGBoost, which measures the contribution of each feature across all trees in the model. A higher value of feature importance indicates a more significant contribution to the model's predictions. Fig. 8 displays the feature relative importance of three inputs (i.e., *T*, dry density, initial unfrozen water content and SSA). Among these four features, temperature (49.87%) is recognized as the most important feature for UWC, followed by initial unfrozen water content (30.74%), dry density (16.34%) and SSA (3.05%). These findings offer valuable insights into understanding the UWC of frozen soils, and similar results were also reported in other literature (i.e., Bi et al., 2023; Ren et al., 2023).

4.4. Limitations and perspectives

It is essential to acknowledge that while machine learning models, such as the XGBoost model employed in this study, demonstrate remarkable predictive ability, they are likely to lack explicit physical interpretations. Without any information on physical laws, machine learning models inherently function as a "black box" for discerning the nonlinear relationships between inputs and outputs, which would raise concerns regarding their generalization and interpretability. Since these models rely solely on patterns and correlations within the available data, they may struggle to accurately predict unseen scenarios or provide meaningful understanding from these statistical models. The absence of physical laws limits our ability to understand the underlying mechanisms driving the observed data in SFCC. Therefore, while these models may exhibit strong fitting performance, their insufficient reliability and interpretability limit the application of

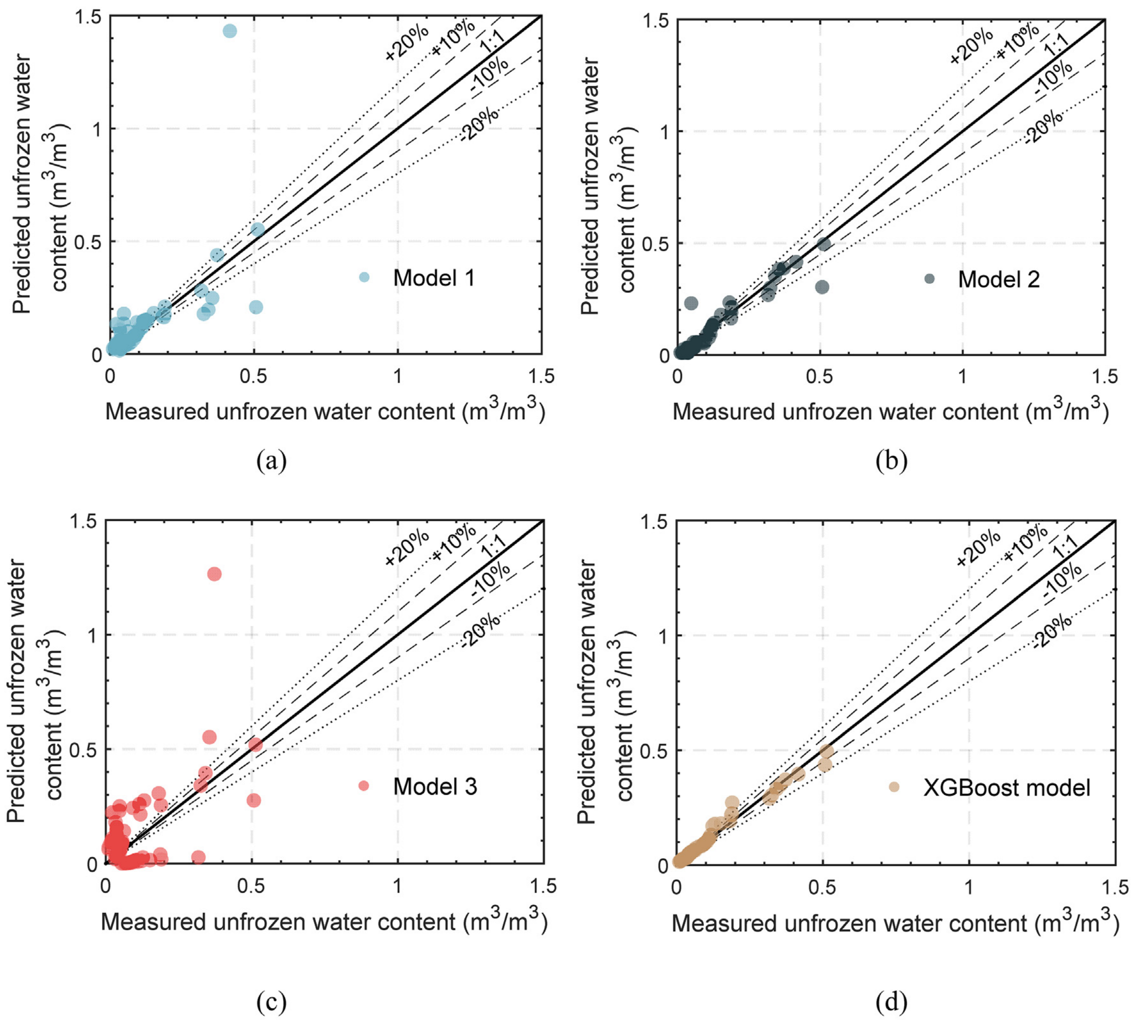

Fig. 7. Comparison of measured unfrozen water content and predicted unfrozen water content from (a) Model 1; (b) Model 2; (c) Model 3; (d) XGBoost model.

Table 5
Comparison between XGBoost and three traditional models.

Prediction Model	Empirical model		Semi-empirical model	XGBoost model
	Model 1	Model 2	Model 3	
R^2	−0.3320 (3)	0.8889 (2)	−4.6115 (4)	0.9739 (1)
RMSE	0.1303 (3)	0.0376 (2)	0.2675 (4)	0.0182 (1)
Model Rank	3	2	4	1

machine learning models. Further efforts can be dedicated to developing more advanced data-driven models that integrate the baseline machine learning model with explicit incorporation of physical laws.

For a broader perspective, this study primarily concentrated on the SFCC along the freezing limb. Specifically, it incorporated the impacts of temperature, density, initial unfrozen water content and SSA on the change of UWC. However, it is essential to note that the actual UWC in frozen soils exhibits distinct hysteretic behaviour and varies with other factors such as initial water content, freeze–thaw history and stress state. The influence of salt concentrations on the SFCC cannot be disregarded, as the presence of salts can induce freezing-point depression. During the freezing process, the salts are expelled from the ice phase, resulting in a higher concentration of the remaining solution, which in turn leads to a further depression of the freezing point. The freezing point can be affected by various factors, such as initial water content, pore structure, particle size distribution, and salinity (Wen et al., 2012; Zhou et al., 2018; Li et al., 2024a). Additionally, some frozen soils with high plasticity display a two-stage SFCC, characterized by a distinct "gap", as highlighted by Kozlowski (2004) and Kozlowski and Nartowska (2013). Therefore, further investigations will be conducted within the framework established in this study to elucidate the impacts of these additional influencing factors and to model the hysteretic effects and gap effect persisting through SFCC. Additionally, unlike the SWRC, data on the SFCC is

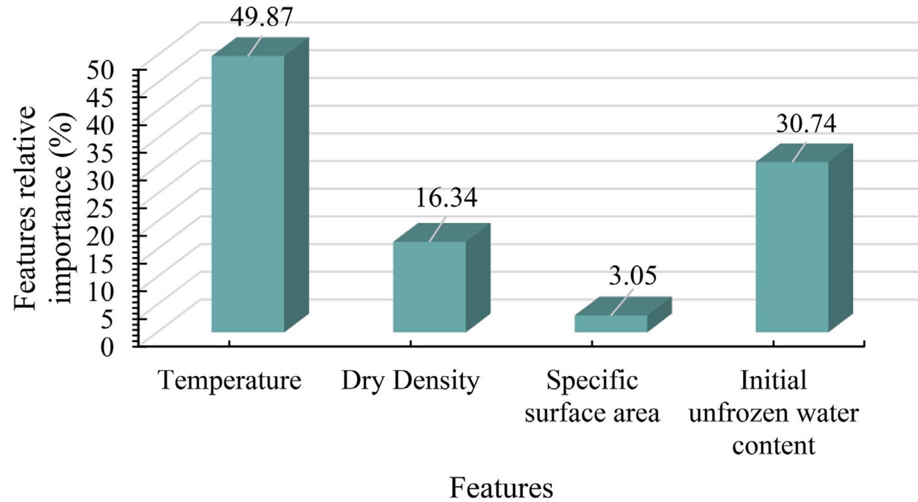

Fig. 8. Results of feature relative importance.

relatively scarce, and it is highly encouraged to undertake future efforts to obtain a more extensive collection of such measurements.

5. Conclusion

The determination of UWC in frozen soils through experimental tests requires specialized equipment and verification on specific soil types, which could be expensive, time-consuming and unavailable in all scenarios. Moreover, accurately determining the variation of UWC is challenging due to the influence of multiple factors. A comprehensive compilation of experimental data on UWC from diverse testing methods and various frozen soil types is employed to train a machine learning model, i.e., XGBoost mode, for estimating the UWC of frozen soils. This model incorporates the effects of temperature, dry density, initial unfrozen water content and SSA on the variation of UWC. The primary advantage of using XGBoost is that it offers a unified tool for modelling the SFCC of frozen soils based on the compiled systematic database.

The framework of using the XGBoost model to predict the UWC of frozen soils is verified by comparing three conventional models based on the compiled database for UWC. Two model evaluation indicators and one model ranking index are combined to quantitatively evaluate the predictive capacity of each model in estimating the UWC. The results showed the proposed XGBoost model effectively captures and reproduces the complex relationships between UWC and inputs (temperature, density and SSA) of various types of frozen soils. In addition, the predictive capacity of the three models in order of highest to lowest is the XGBoost model, model 2 (an empirical model proposed by Kurylyk and Watanabe (2013) where UWC is an exponential function of T), model 3 (a semi-empirical model proposed by Anderson and Tice (1972) where UWC is a function of T and SSA) and model 1 (an empirical model proposed by Anderson and Tice (1972) where UWC is a function of T).

The notable advantage of the new XGBoost model is its capability to directly predict UWC values based on other soil properties by recognizing the intricate relationships among temperature, dry density, SSA, initial unfrozen water content and UWC. This model offers a general approach applicable to various types of frozen soils and can provide practitioners with a more accurate and efficient tool for rapidly evaluating UWC in frozen soils that significantly influences engineering geology projects and disaster control problems. Additionally, the model serves as a valuable reference for parameterizing UWC in soils within cold regions, further enhancing its practical utility.

CRediT authorship contribution statement

Kai-Qi Li: Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing, Conceptualization. **Hai-Long He:** Conceptualization, Funding acquisition, Project administration, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was financially supported by the National Natural Science Foundation of China (Grant No. 42177291) and Innovation Capability Support Program of Shaanxi Province (2023-JC-JQ-25 and 2021KJXX-11).

References

Akagawa, S., Iwahana, G., Watanabe, K., Chuvilin, E.M., Istomin, V.A., 2012. Improvement of pulse NMR technology for determination of unfrozen water content in frozen soils. In: Proceedings of the Tenth International Conference on Permafrost, Salekhard, Russia.
Anbergen, H., Rühaak, W., Frank, J., Sass, I., 2015. Numerical simulation of a freeze–thaw testing procedure for borehole heat exchanger grouts. Can. Geotech. J. 52 (8), 1087–1100.
Anderson, D.M., Tice, A.R., 1972. Predicting unfrozen water contents in frozen soils from surface area measurements. Highw. Res. Rec. 393 (2), 12–18.
Bai, R., Lai, Y., Zhang, M., Yu, F., 2018. Theory and application of a novel soil freezing characteristic curve. Appl. Therm. Eng. 129, 1106–1114.
Bergen, K.J., Johnson, P.A., de Hoop, M.V., Beroza, G.C., 2019. Machine learning for data-driven discovery in solid Earth geoscience. Science 363 (6433), eaau0323.
Bi, J., Wang, G., Wu, Z., Wen, H., Zhang, Y., Lin, G., Sun, T., 2023. Investigation on unfrozen water content models of freezing soils. Front. Earth Sci. 10, 1039330.
Bittelli, M., Flury, M., Campbell, G.S., 2003. A thermodielectric analyzer to measure the freezing and moisture characteristic of porous media. Water Resour. Res. 39 (2), 1041.
Black, P.B., Tice, A.R., 1989. Comparison of soil freezing curve and soil water curve data for Windsor sandy loam. Water Resour. Res. 25 (10), 2205–2210.
Bouyoucos, G.J., 1921. A new classification of soil moisture. Soil Sci. 11 (1), 33–48.

Buehrer, T.F., Rose, M.S., 1943. Studies in soil structure V. Bound water in normal and puddled soils. College of Agriculture, University of Arizona (Tucson, AZ).
Cahn, J.W., Dash, J.G., Fu, H., 1992. Theory of ice premelting in monosized powders. J. Cryst. Growth 123 (1–2), 101–108.
Cao, D.F., Zhu, H.H., Wu, B., Wang, J.C., Shukla, S.K., 2021. Investigating temperature and moisture profiles of seasonally frozen soil under different land covers using actively heated fiber Bragg grating sensors. Eng. Geol. 290, 106197.
Chai, M., Zhang, J., Zhang, H., Mu, Y., Sun, G., Yin, Z., 2018. A method for calculating unfrozen water content of silty clay with consideration of freezing point. Appl. Clay Sci. 161, 474–481.
Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785-794.
Christ, M., Park, J.B., 2009. Ultrasonic technique as tool for determining physical and mechanical properties of frozen soils. Cold Reg. Sci. Technol. 58 (3), 136–142.
Dall'Amico, M., Endrizzi, S., Gruber, S., Rigon, R.J.T.C., 2011. A robust and energy-conserving model of freezing variably-saturated soil. Cryosphere 5 (2), 469–484.
Dash, J.G., Fu, H., Wettlaufer, J.S., 1995. The premelting of ice and its environmental consequences. Rep. Prog. Phys. 58 (1), 115.
Devoie, É.G., Gruber, S., McKenzie, J.M., 2022. A repository of measured soil freezing characteristic curves: 1921 to 2021. Earth Syst. Sci. Data 14 (7), 3365–3377.
Feng, S.N., Chen, J.R., Jones, S.B., Flerchinger, G., Dyck, M., Filipovic, V., Hu, Y., Si, B., Lv, J., Wu, Q., He, Hai., 2024. Miscellaneous methods for determination of unfrozen water content in frozen soils. J. Hydrol. 31, 130802.
Feng, D.C., Wang, W.J., Mangalathu, S., Taciroglu, E., 2021. Interpretable XGBoost-SHAP machine-learning model for shear strength prediction of squat RC walls. J. Struct. Eng. 147 (11), 04021173.
Fu, Z.T., Wu, Q.B., Miles, D., He, H.L., 2021. A new model to describe soil freezing and thawing characteristic curve. J. Glaciol. Geocryol. 43, 437–452.
Gu, X., Wang, L., Ou, Q., Zhang, W., 2023. Efficient stochastic analysis of unsaturated slopes subjected to various rainfall intensities and patterns. Geosci. Front. 14 (1), 101490.
He, H., Dyck, M., 2013. Application of multiphase dielectric mixing models for understanding the effective dielectric permittivity of frozen soils. Vadose Zone J. 12 (1). vzj2012-0060.
He, H., Dyck, M., Wang, J., Lv, J., 2015. Evaluation of TDR for quantifying heat-pulse-method-induced ice melting in frozen soils. Soil Sci. Soc. Am. J. 79 (5), 1275–1288.
He, Y., Xu, Y., Lv, Y., et al., 2023. Characterization of unfrozen water in highly organic turfy soil during freeze–thaw by nuclear magnetic resonance. Eng. Geol. 312, 106937.
Heuvelink, G.B., Angelini, M.E., Poggio, L., et al., 2021. Machine learning in space and time for modelling soil organic carbon change. Eur. J. Soil Sci. 72 (4), 1607–1623.
Hu, G., Zhao, L., Zhu, X., et al., 2020. Review of algorithms and parameterizations to determine unfrozen water content in frozen soil. Geoderma 368, 114277.
Jame, Y.W., Norum, D.I., 1972. Phase composition of a partially frozen soil. Res. Paper II. Division of Hydrology. College of Engineering University of Saskatchewan, Saskatoon, Canada.
Kong, L., Wang, Y., Sun, W., Qi, J., 2020. Influence of plasticity on unfrozen water content of frozen soils as determined by nuclear magnetic resonance. Cold Reg. Sci. Technol. 172, 102993.
Koopmans, R.W.R., Miller, R.D., 1966. Soil freezing and soil water characteristic curves. Soil Sci. Soc. Am. J. 30 (6), 680–685.
Kozlowski, T., 2004. Soil freezing point as obtained on melting. Cold Reg. Sci. Technol. 38 (2–3), 93–101.
Kozlowski, T., 2007. A semi-empirical model for phase composition of water in clay–water systems. Cold Reg. Sci. Technol. 49 (3), 226–236.
Kozlowski, T., Nartowska, E., 2013. Unfrozen water content in representative bentonites of different origin subjected to cyclic freezing and thawing. Vadose Zone J. 12 (1), vzj2012-0057.
Kurylyk, B.L., Watanabe, K., 2013. The mathematical representation of freezing and thawing processes in variably-saturated, non-deformable soils. Adv. Water Resour. 60, 160–177.
Li, Z., Chen, J., Sugimoto, M., 2020b. Pulsed NMR measurements of unfrozen water content in partially frozen soil. J. Cold Reg. Eng. 34 (3), 04020013.
Li, K., Horton, R., He, H., 2023a. Application of machine learning algorithms to model soil thermal diffusivity. Int. Commun. Heat Mass Transfer 149, 107092.
Li, K.Q., Li, D.Q., Liu, Y., 2020a. Meso-scale investigations on the effective thermal conductivity of multi-phase materials using the finite element method. Int. J. Heat Mass Transf. 151, 119383.
Li, K.Q., Kang, Q., Nie, J.Y., Huang, X.W., 2022a. Artificial neural network for predicting the thermal conductivity of soils based on a systematic database. Geothermics 103, 102416.
Li, X., Li, X., 2023. A soil freezing-thawing model based on thermodynamics. Cold Reg. Sci. Technol. 211, 103867.
Li, X., Li, X., Liu, J., 2023d. A dynamic soil freezing characteristic curve model for frozen soil. J. Rock Mech. Geotech. Eng. https://doi.org/10.1016/j.jrmge.2023.09.008.
Li, K.Q., Liu, Y., Kang, Q., 2022b. Estimating the thermal conductivity of soils using six machine learning algorithms. Int. Commun. Heat Mass Transfer 136, 106139.
Li, K.Q., Yin, Z.Y., Liu, Y., 2023b. A hybrid SVR-BO model for predicting the soil thermal conductivity with uncertainty. Can. Geotech. J. 61 (2), 258–274.
Li, K.Q., Yin, Z.Y., Zhang, N., Liu, Y., 2023c. A data-driven method to model stress-strain behaviour of frozen soil considering uncertainty. Cold Reg. Sci. Technol.103906
Li, K., Yin, Z., Liu, Y., 2023e. Influences of spatial variability of hydrothermal properties on the freezing process in artificial ground freezing technique. Comput. Geotech. 159, 105448.
Li, K.Q., Yin, Z.Y., Qi, J.L., Liu, Y., 2024a. State-of-the-art constitutive modelling of frozen soils. Arch. Comput. Meth. Eng. https://doi.org/10.1007/s11831-024-10102-w.
Li, KQ., Yin, ZY., 2024. State of the Art of Coupled Thermo -hydro-Mechanical -Chemical Modelling for Frozen Soils. Archives of Computational Methods in Engineering. https://doi.org/10.1007/s11831-024-10164-w.
Li, K.Q., Yin, Z.Y., Jin, Y.F., Liu, Y., 2024b. Investigation on evolution law of frozen wall thickness in artificial ground freezing under seepage conditions. Can. Geotech. J. https://doi.org/10.1139/cgj-2023-0576.
Li, K.Q., Yin, Z.Y., Zhang, N., Li, J., 2024c. A PINN-based modelling approach for hydromechanical behaviour of unsaturated expansive soils. Comput. Geotech. 169, 106174.
Li, X., Zheng, S.F., Wang, M., Liu, A.Q., 2023f. The prediction of the soil freezing characteristic curve using the soil water characteristic curve. Cold Reg. Sci. Technol. 212, 103880.
Liu, Y., Li, K.Q., Li, D.Q., Tang, X.S., Gu, S.X., 2022. Coupled thermal–hydraulic modeling of artificial ground freezing with uncertainties in pipe inclination and thermal conductivity. Acta Geotech. 17, 257–274.
Liu, Z., Yu, X., 2013. Physically based equation for phase composition curve of frozen soils. Transp. Res. Rec. 2349 (1), 93–99.
Lu, N., Likos, W.J., Luo, S., Oh, H., 2021. Is the conventional pore water pressure concept adequate for fine-grained soils in geotechnical and geoenvironmental engineering? J. Geotech. Geoenviron. Eng. 147 (10), 02521001.
Luo, S., Lü, S., Zhang, Y., 2009. Development and validation of the frozen soil parameterization scheme in Common Land Model. Cold Reg. Sci. Technol. 55 (1), 130–140.
McKenzie, J.M., Voss, C.I., Siegel, D.I., 2007. Groundwater flow with energy transport and water–ice phase change: numerical simulations, benchmarks, and application to freezing in peat bogs. Adv. Water Resour. 30 (4), 966–983.
Michalowski, R.L., 1993. A constitutive model of saturated soils for frost heave simulations. Cold Reg. Sci. Technol. 22 (1), 47–63.
Mu, Q.Y., 2017. Hydro-mechanical behaviour of loess at elevated and subzero temperatures Doctoral dissertation. The Hong Kong University of Science and Technology (HKUST), Hong Kong, China.
Nakano, Y., Tice, A., Oliphant, J., Jenkins, T., 1982. Transport of water in frozen soil: I: Experimental determination of soil-water diffusivity under isothermal conditions. Adv. Water Resour. 5 (4), 221–226.
Nersesova, Z.A., Tsytovich, N.A., 1963. Unfrozen water in frozen soils. In: In Permafrost: Proceedings of 1st International Conference, pp. 11–15.
Nicolsky, D.J., Romanovsky, V.E., Panda, S.K., Marchenko, S.S., Muskett, R.R., 2017. Applicability of the ecosystem type approach to model permafrost dynamics across the Alaska North Slope. J. Geophys. Res. Earth 122 (1), 50–75.
Nishimura, S., Gens, A., Olivella, S., Jardine, R.J., 2009. THM-coupled finite element analysis of frozen soil: formulation and application. Géotechnique 59 (3), 159–171.
Niu, G.Y., Yang, Z.L., 2006. Effects of frozen soil on snowmelt runoff and soil water storage at a continental scale. J. Hydrometeorol. 7 (5), 937–952.
Oliphant, J.L., Tice, A.R., Nakano, Y., 1983. Water migration due to a temperature gradient in frozen soil. Proc. 4th Int. Conf. on Permafrost, Fairbanks, Alaska., vol. 1. National Academy Press, Washington, D.C, 95-956.
Osterkamp, T.E., Romanovsky, V.E., 1997. Freezing of the active layer on the coastal plain of the Alaskan Arctic. Permafr. Periglac. Process. 8 (1), 23–44.
Patterson, D.E., Smith, M.W., 1981. The measurement of unfrozen water content by time domain reflectometry: Results from laboratory tests. Can. Geotech. J. 18 (1), 131–144.
Peng, W., Lu, Y., Wang, M., Ren, T., Horton, R., 2022. Determining water content and bulk density: The heat-pulse method outperforms the thermo-tdr method in high-salinity soils. Geoderma 407, 115564.
Qin, Y., Li, G., Qu, G., 2009. A formula for the unfrozen water content and temperature of frozen soils. In Cold Regions Engineering 2009: Cold Regions Impacts on Research, Design, and Construction, pp. 155–161.
Ren, J., Vanapalli, S.K., 2019. Comparison of soil-freezing and soil-water characteristic curves of two Canadian soils. Vadose Zone J. 18 (1–14), 180185.
Ren, J., Vanapalli, S.K., Han, Z., 2018. Soil freezing process and different expressions for the soil-freezing characteristic curve. Sciences in Cold and Arid Regions 9 (3), 221–228.
Pusch, R., 1979. Unfrozen water as a function of clay microstructure. Engineering Geology 13 (1–4), 157–162.
Ren, J., Fan, X., Yu, X., Vanapalli, S., Zhang, S., 2023. Use of an artificial neural network model for estimation of unfrozen water content in frozen soils. Can. Geotech. J. 60 (8), 1234–1248.
Riseborough, D.W., 2004. Exploring the parameters of a simple model of the permafrost-climate relationship. Carleton University.
Robert, A., 1998. Dielectric permittivity of concrete between 50 MHz and 1 GHz and GPR measurements for building materials evaluation. J. Appl. Geophys. 40 (1–3), 89–94.
Shastri, A., Sánchez, M., Gai, X., Lee, M.Y., Dewers, T., 2021. Mechanical behavior of frozen soils: Experimental investigation and numerical modeling. Comput. Geotech. 138, 104361.

Sheshukov, A.Y., Nieber, J.L., 2011. One-dimensional freezing of nonheaving unsaturated soils: Model formulation and similarity solution. Water Resour. Res. 47 (11), W11519.

Shoop, S.A., Bigl, S.R., 1997. Moisture migration during freeze and thaw of unsaturated soils: modeling and large scale experiments. Cold Reg. Sci. Technol. 25 (1), 33–45.

Spaans, E.J., Baker, J.M., 1996. The soil freezing characteristic: Its measurement and similarity to the soil moisture characteristic. Soil Sci. Soc. Am. J. 60 (1), 13–19.

Stähli, M., Stadler, D., 1997. Measurement of water and solute dynamics in freezing soil columns with time domain reflectometry. J. Hydrol. 195 (1–4), 352–369.

Stähli, M., Bayard, D., Wydler, H., Flühler, H., 2004. Snowmelt infiltration into alpine soils visualized by dye tracer technique. Arct. Antarct. Alp. Res. 36 (1), 128–135.

Suzuki, S., Kitamura, S., 2008. Unfrozen water in amylosic molecules is dependent on the molecular structures—a differential scanning calorimetric study. Food Hydrocoll. 22 (5), 862–867.

Teng, J., Zhong, Y., Zhang, S., Sheng, D., 2021. A mathematic model for the soil freezing characteristic curve: the roles of adsorption and capillarity. Cold Reg. Sci. Technol. 181, 103178.

Tian, H., Wei, C., Wei, H., Zhou, J., 2014. Freezing and thawing characteristics of frozen soils: Bound water content and hysteresis phenomenon. Cold Reg. Sci. Technol. 103, 74–81.

Tice, A.R., Anderson, D.M., Sterrett, K.F., 1982. Unfrozen water contents of submarine permafrost determined by nuclear magnetic resonance. Dev. Geotech. Eng. 28, 135–146.

Tziachris, P., Aschonitis, V., Chatzistathis, T., Papadopoulou, M., Doukas, I.J.D., 2020. Comparing machine learning models and hybrid geostatistical methods using environmental and soil covariates for soil pH prediction. ISPRS Int. J. Geo Inf. 9 (4), 276.

Wadoux, A.M.C., Samuel-Rosa, A., Poggio, L., Mulder, V.L., 2020. A note on knowledge discovery and machine learning in digital soil mapping. Eur. J. Soil Sci. 71 (2), 133–136.

Wan, X., Pei, W., Lu, J., Qiu, E., Yan, Z., Pirhadi, N., Zhu, J., 2022. Analytical model to predict unfrozen water content based on the probability of ice formation in soils. Permafr. Periglac. Process. 33 (4), 436–451.

Wan, X.S., Zhu, S.J., Pei, W.S., Zhou, F.X., Lu, J.G., Yan, Z.R., Wa, D., 2023. A theoretical model on unfrozen water content in soils and verification. J. Hydrol. 622, 129675.

Wang, C., Lai, Y., Zhang, M., 2017. Estimating soil freezing characteristic curve based on pore-size distribution. Appl. Therm. Eng. 124, 1049–1060.

Wang, Q., Yan, X., Dong, Y., Su, W., Meng, Y., Sun, W., 2023b. Effect of Beishan groundwater salinity on the self-sealing performance of compacted GMZ bentonite. Environ. Earth Sci. 82 (17), 391.

Wang, X., Yang, Y., Lv, J., He, H., 2023a. Past, present and future of the applications of machine learning in soil science and hydrology. Soil Water Res. 18 (2), 67–80.

Watanabe, K., Osada, Y., 2017. Simultaneous measurement of unfrozen water content and hydraulic conductivity of partially frozen soil near 0 C. Cold Reg. Sci. Technol. 142, 79–84.

Watanabe, K., Wake, T., 2009. Measurement of unfrozen water content and relative permittivity of frozen unsaturated soil using NMR and TDR. Cold Reg. Sci. Technol. 59 (1), 34–41.

Watanabe, K., Kito, T., Wake, T., Sakai, M., 2011. Freezing experiments on unsaturated sand, loam and silt loam. Ann. Glaciol. 52 (58), 37–43.

Wen, H., Bi, J., Guo, D., 2020. Evaluation of the calculated unfrozen water contents determined by different measured subzero temperature ranges. Cold Reg. Sci. Technol. 170, 102927.

Wen, Z., Zhang, T., Sheng, Y., Ma, W., Wu, Q., Feng, W., Sun, Z., 2011. Managing ice-rich permafrost exposed during cutting excavation along Qinghai-Tibetan railway: Experiences and implementation. Eng. Geol. 122 (3–4), 316–327.

Wen, Z., Ma, W., Feng, W., Deng, Y., Wang, D., Fan, Z., Zhou, C., 2012. Experimental study on unfrozen water content and soil matric potential of Qinghai-Tibetan silty clay. Environ. Earth Sci. 66, 1467–1476.

Weng, L., Wu, Z., Liu, Q., Chu, Z., Zhang, S., 2021. Evolutions of the unfrozen water content of saturated sandstones during freezing process and the freeze-induced damage characteristics. Int. J. Rock Mech. Min. Sci. 142, 104757.

Westermann, S., Boike, J., Langer, M., Schuler, T.V., Etzelmüller, B., 2011. Modeling the impact of wintertime rain events on the thermal regime of permafrost. Cryosphere 5 (4), 945–959.

Williams, P.J., 1964. Unfrozen water content of frozen soils and soil moisture suction. Géotechnique 14 (3), 231–246.

Wu, M., Tan, X., Huang, J., Wu, J., Jansson, P.E., 2015. Solute and water effects on soil freezing characteristics based on laboratory experiments. Cold Reg. Sci. Technol. 115, 22–29.

Wu, Y., Wang, Y., Hu, L., 2023. A theoretical model of the soil freezing characteristic curve for saline soil. J. Hydrol. 622, 129639.

Wu, B., Zhu, H.H., Cao, D., Xu, L., Shi, B., 2021. Feasibility study on ice content measurement of frozen soil using actively heated FBG sensors. Cold Reg. Sci. Technol. 189, 103332.

Xu, X.Z., Wang, J.C., Zhang, L.X., 2001. Frozen Soil Physics Science. Science Press, Beijing.

Yang, L., Jia, H., Han, L., Zhang, H., Tang, L., 2021. Hysteresis in the ultrasonic parameters of saturated sandstone during freezing and thawing and correlations with unfrozen water content. J. Rock Mech. Geotech. Eng. 13 (5), 1078–1092.

Yang, C., Yin, Y., Zhang, J., Ding, P., Liu, J., 2024. A graph deep learning method for landslide displacement prediction based on global navigation satellite system positioning. Geosci. Front. 15 (1), 101690.

Ye, M., Zhu, L., Li, X., et al., 2023. Estimation of the soil arsenic concentration using a geographically weighted XGBoost model based on hyperspectral data. Sci. Total Environ. 858, 159798.

Yoshikawa, K., Overduin, P.P., 2005. Comparing unfrozen water content measurements of frozen soil using recently developed commercial sensors. Cold Reg. Sci. Technol. 42 (3), 250–256.

Yu, F., Qi, J., Lai, Y., et al., 2016. Typical embankment settlement/heave patterns of the Qinghai-Tibet highway in permafrost regions: Formation and evolution. Eng. Geol. 214, 147–156.

Yuan, L., Zhao, L., Li, R., Hu, G., Du, E., Qiao, Y., Ma, L., 2020. Spatiotemporal characteristics of hydrothermal processes of the active layer on the central and northern Qinghai-Tibet plateau. Sci. Total Environ. 712, 136392.

Zhang, T., Barry, R.G., Knowles, K., Heginbottom, J.A., Brown, J., 1999. Statistics and characteristics of permafrost and ground-ice distribution in the Northern Hemisphere. Polar Geogr. 23 (2), 132–154.

Zhang, Y., Carey, S.K., Quinton, W.L., 2008. Evaluation of the algorithms and parameterizations for ground thawing and freezing simulation in permafrost regions. J. Geophys. Res. Atmos. 113, D17116.

Zhang, M., Pei, W., Li, S., Lu, J., Jin, L., 2017. Experimental and numerical analyses of the thermo-mechanical stability of an embankment with shady and sunny slopes in a permafrost region. Appl. Therm. Eng. 127, 1478–1487.

Zhang, X., Sun, S.F., Xue, Y., 2007. Development and testing of a frozen soil parameterization for cold region studies. J. Hydrometeorol. 8 (4), 690–701.

Zhang, M., Zhang, X., Lu, J., Pei, W., Wang, C., 2019. Analysis of volumetric unfrozen water contents in freezing soils. Exp. Heat Transfer 32 (5), 426–438.

Zhang, N., Zhou, A., Jin, Y.F., Yin, Z.Y., Shen, S.L., 2023. An enhanced deep learning method for accurate and robust modelling of soil stress–strain response. Acta Geotechnica 18 (8), 4405–4427.

Zhao, D., Wang, J., Zhao, X., Triantafilis, J., 2022. Clay content mapping and uncertainty estimation using weighted model averaging. Catena 209, 105791.

Zhou, J., Wei, C., Lai, Y., Wei, H., Tian, H., 2018. Application of the generalized Clapeyron equation to freezing point depression and unfrozen water content. Water Resour. Res. 54 (11), 9412–9431.