Use of an artificial neural network model for estimation of unfrozen water content in frozen soils

Junping Ren[1], Xudong Fan[2], Xiong (Bill) Yu[2], Sai K Vanapalli[3], and Shoulong Zhang[4]

[1]Lanzhou University
[2]Case Western Reserve University
[3]University of Ottawa
[4]Hokkaido University

November 22, 2022

Abstract

A portion of pore water is typically in a state of unfrozen condition in frozen soils due to the complex soil-water interactions. The variation of the amount of unfrozen water and ice has a significant influence on the physical and mechanical behaviors of the frozen soils. Several empirical, semi-empirical, physical and theoretical models are available in the literature to estimate the unfrozen water content (UWC) in frozen soils. However, these models have limitations due to the complex interactions of various influencing factors that are not well understood or fully established. For this reason, in the present study, an artificial neural network (ANN) modeling framework is proposed and the PyTorch package is used for predicting the UWC in soils. For achieving this objective, extensive UWC data of various types of soils tested under various conditions were collected through an extensive search of the literature. The developed ANN model showed good performance for the test dataset. In addition, the model performance was compared with two traditional statistical models for UWC prediction on four additional types of soils and found to outperform these traditional models. Detailed discussions on the developed ANN model, and its strengths and limitations in comparison to different other models are provided. The study demonstrates that the proposed ANN model is simple yet reliable for estimating the UWC of various soils. In addition, the summarized UWC data and the proposed machine learning modeling framework are valuable for future studies related to frozen soils.

Use of an artificial neural network model for estimation of unfrozen water content in frozen soils

Junping Ren[1], Xudong Fan[2*], Xiong (Bill) Yu[2], Sai K. Vanapalli[3], Shoulong Zhang[4]

[1] MOE Key Laboratory of Mechanics on Disaster and Environment in Western China, College of Civil Engineering and Mechanics, Lanzhou University, Lanzhou 730000, China

[2] Department of Civil and Environmental Engineering, Case Western Reserve University, Cleveland 44106-7201, USA

[3] Department of Civil Engineering, University of Ottawa, Ottawa K1N6N5, Canada

[4] Division of Field Engineering for the Environment, Hokkaido University, Sapporo 060-8628, Japan

*Corresponding author: Xudong Fan

Email: xxf121@case.edu

Abstract

A portion of pore water is typically in a state of unfrozen condition in frozen soils due to the complex soil-water interactions. The variation of the amount of unfrozen water and ice has a significant influence on the physical and mechanical behaviors of the frozen soils. Several empirical, semi-empirical, physical and theoretical models are available in the literature to estimate the unfrozen water content (UWC) in frozen soils. However, these models have limitations due to the complex interactions of various influencing factors that are not well understood or fully established. For this reason, in the present study, an artificial neural network (ANN) modeling framework is proposed and the PyTorch package is used for predicting the UWC in soils. For achieving this objective, extensive UWC data of various types of soils tested under various conditions were collected through an extensive search of the literature. The developed ANN model showed good performance for the test dataset. In addition, the model performance was compared with two traditional statistical models for UWC prediction on four additional types of soils and found to outperform these traditional models. Detailed discussions on the developed ANN model, and its strengths and limitations in comparison to different other models are provided. The study demonstrates that the proposed ANN model is simple yet reliable for estimating the UWC of various soils. In addition, the summarized UWC data and the proposed machine learning modeling framework are valuable for future studies related to frozen soils.

Keywords: *Frozen soils, unfrozen water, artificial neural network, modeling framework, prediction*

1. Introduction

The freezing of water to form ice is one of the most common phase transformations in the natural environment (Wettlaufer, 1999). Nearly one-third of the land surface of the Earth experiences freezing and thawing annually (Lu et al., 2021). In these permafrost and seasonally frozen regions, unfrozen water and pore ice coexist within a frozen soil, due to the complex soil-water interactions. The unfrozen water exists in small pore spaces by capillarity and as thin films adsorbed on the surfaces of soil particles, in equilibrium with the pore ice at subzero temperatures. The relationship between the amount of unfrozen water and its energy state in a frozen soil is generally referred to as the soil-freezing characteristic curve (SFCC) in the literature (Ren et al., 2021). The quantity of unfrozen water in the frozen soil can be represented by either the gravimetric water content, or volumetric water content, or degree of unfrozen water saturation. The energy state of unfrozen water is typically represented by the subzero temperature of the frozen soil.

The SFCC links the degree of water-ice phase transition to the subzero temperature in a frozen soil. Since the constitutive relationships for hydraulic, thermal, and mechanical fields of frozen soils are functions of the quantity of unfrozen water, the SFCC is essential for modeling the transport mechanism of water, heat, and solutes in frozen soils (e.g., Lai et al., 2014; Yu et al., 2020a, b; Saberi and Meschke, 2021). For example, reliable determination of the unfrozen water in frozen soils is valuable for predicting their hydraulic properties which are vital for models of flood forecasting during spring thawing, and their mechanical properties that determine the stability of the ground for infrastructure in cold regions (Amankwah et al., 2021). In other words, a sound understanding of SFCC is critical for broad engineering applications and for understanding the likely impacts associated with climate change (Lara et al., 2021).

Due to its essential role in cold regions science and engineering, an accurate description of the unfrozen water content (UWC) is crucial to achieve a realistic representation of the behavior of frozen soils. In addition, the increasing use of permafrost regions for civil infrastructure constructions and the effects of global warming on these regions has further stimulated research on the behavior of frozen soils (Shastri, 2014; Saberi and Meschke, 2021), among which the UWC is a key property. Many models have been proposed to estimate the soil UWC or SFCC during the last few decades. These proposed models are generally based on using soil physical properties, the similarity between SFCC and soil-water characteristic curve (SWCC), and / or physical and theoretical mechanisms (Ren, 2019). Amongst these models, the empirical models were generally put forward by earlier researchers (e.g., Dillon and Andersland, 1966; Anderson and Tice, 1972; Xu et al., 1985; Michalowski, 1993; Mckenzie et al., 2007). Most of the empirical models are based on fitting experimental results, with a connection to the basic physical properties of frozen soils (Ming et al., 2020) and subzero temperature. In recent years, there has been significant interest in proposing physical, theoretical and thermodynamic models for estimating UWC (e.g., Liu and Yu, 2013, 2014; Wang et al., 2017a; Amiri et al., 2018; Bai et al., 2018; Chai et al., 2018; Mu et al., 2018; Teng et al., 2020; Zhou et al., 2020; Jin et al., 2020; Xiao et al., 2020; Saberi and Meschke, 2021), that may be attributed to the better understanding of physical mechanisms underlying the ice-water transition in porous media. Some investigators have summarized these models in their research studies (e.g., Kurylyk and Watanabe, 2013; Mu, 2017; Ren et al., 2017; Lu et al., 2019; Hu et al., 2020).

It is widely acknowledged that many factors influence the UWC in frozen soils. These factors mainly include the soil physical and chemical properties, stress sate, and temperature. The complex effects of these factors result in a highly nonlinear relationship between these factors and the UWC. In addition, the relative contribution of each factor on UWC is not well-understood. This causes difficulties in selecting the most relevant factors for establishing a reliable UWC model. Such difficulties can be effectively addressed by using machine learning (ML) algorithms, such as the artificial neural network (ANN) models. The ANN is an adaptive information-processing technique, which allows the correlations between input and output variables to be established through inter-connected neurons (Saha et al., 2018). The key advantage of an ANN model in comparison to empirical and statistical methods is that it does not require any prior knowledge about the nature of the relationship between the input and output variables (Shahin et al., 2001; Pham et al., 2019). In addition, it is able to take account of various influencing factors that have weak or nonlinear relationships with the outcomes (Zhang et al., 2021b; Zhong et al., 2021). For this reason, there is no need to either simplify the problem or introduce simplified assumptions (Shahin et al., 2008). Moreover, ANN models can always be updated to obtain better results by presenting new training examples as new data become available (Ismeik and AI-Rawi, 2014; Zhong et al., 2021). These features make ANN suitable for predicting soil behaviors affected by various factors.

The ANN has been widely employed in geotechnical and geo-environmental engineering fields that include predicting soil stress–strain behavior (Habibagahi and Bamdad, 2003), resilient modulus (Ren et al., 2019), and thermal conductivity / resistivity (e.g., Erzin et al., 2008; Wen et al., 2020). Wang et al. (2020b) employed three ML models to estimate the UWC of a frozen saline soil. Three influencing factors (i.e., temperature, sodium bicarbonate content, and initial water content) were considered in their models. One limitation of their models, however, is that the models were developed based on limited experimental data of a specific soil. This largely restricts the use of their models for other applications. For this reason, in the present study, UWC data of various types of soils tested under various conditions are collected, through an extensive literature search. An ANN model is developed for estimating the UWC in frozen soils, based on the collected large amount of experimental data. A modelling framework is proposed and followed, and the ANN model is built by PyTorch package (Paszke et al., 2017). The developed ANN model is further compared with two traditional statistical models for UWC prediction. Detailed discussions on the developed ANN model and model comparison are also presented. The present study is one of the earliest attempts to modeling UWC in frozen soils by ML algorithms. It can provide good reference (e.g., collected data, modeling framework, and programming scripts) for future studies related to the UWC prediction, and may be incorporated in numerical codes for solving the coupled thermal–hydraulic–mechanical–chemical process in frozen soils.

2. Modeling framework and data sources

Figure 1 represents the proposed framework for the prediction of UWC in frozen soils. The main framework can be divided into data preparation (left part of Fig. 1), model optimization (middle part of Fig. 1), and model application (right part of Fig. 1). The collected datasets are prepared as a tabular dataset where the final column is the prediction target (i.e., volumetric UWC). The first four columns of the prepared dataset are the specific surface area, dry density, initial volumetric water content and temperature, respectively. With the prepared dataset, the features' values are firstly normalized by scaling each factor into a distribution with zero mean value and unit variance. This process is conducted to mitigate computational burden during the model optimization and application processes, as well as to increase the model performance. In the model optimization process, at each iteration, the normalized dataset will be randomly divided into 80%:20%. The 80% samples are used to train the ANN model with given hyperparameters, and the rest 20% samples are used for independent evaluation of the trained model. Based on the evaluation results, Bayesian optimization algorithm is used to find the optimal hyperparameters of the ANN model with better performance. The Bayesian optimization process is iterated 50 times in the present study. After obtaining the optimal hyperparameters, the ANN model is evaluated again with the k-folder cross validation. The folder with best performance is used for Shapley Additive exPlanations (SHAP) interpretation to determine the influence of considered factors on the prediction target.

The details about data collection, ANN model, Bayesian optimization, and k-folder cross validation are discussed in the following sections from Section 2.1 to 2.4.

2.1 Data collection

In the present study, soil physical properties and the UWC data were obtained from the literature. For the UWC, only data points which can be clearly identified (e.g., scattered data points in figures or tabular data) were included. Those with only unfrozen water content curves shown were not considered since it is not possible to identify the real measured UWC data points. This avoids obtaining arbitrary data from the continuous UWC curves. The raw data points were extracted from the original plots using GetData Graph Digitizer.

Factors that influence the UWC of frozen soils can be categorized into the internal and external factors. The internal factors are typical soil physical properties, such as the particle size distribution (PSD), sand/silt/clay content, plasticity indices, specific surface area (SSA), dry density, void ratio (or porosity), initial water content and salinity. The external factors can include temperature, stress state, freeze-thaw and wet-dry cycles, etc. The influencing factors that were considered in various studies in the literature are different and sometimes arbitrary. For example, Smith and Tice (1988), in their study considered four factors that include three internal factors (SSA, initial water content and dry density) and one external factor (temperature). In another study, Kruse and Darrow (2017) considered more factors such as soil cation exchange capacity and cation treatment. Besides temperature, which typically has the most significant effect on UWC, only a few studies considered other external factors such as freeze-thaw cycles and stress state (e.g., Mu, 2017; Ren and Vanapalli, 2020). Therefore, it is difficult to find abundant data or studies that took into account exact the same types of influencing factors. As a result, a search of more than 100 articles from the literature resulted in identifying 20 articles that can be used in the present study, as listed in Table 1.

In this study, the following factors were selected: SSA, dry density (ρ_d), initial volumetric water content (θ_{init}) and temperature (*Temp*). This is because these four factors were considered in all the 20 articles and the UWC data of a variety of soils are available (73 soils in Table 1). It should be noted that the soil specimens used for UWC measurement were not necessarily initially saturated. Table 1 also indicates that the UWC data were mostly measured by nuclear magnetic resonance (NMR) and time domain reflectometry (TDR), while some of them were measured by other methods such as frequency domain reflectometry (FDR), time domain transmissometry (TDT), etc. In order to increase the database, the UWC data was collected regardless of the testing methods. The gravimetric water content was converted to volumetric water content by multiplying by soil dry density. The thawing or freezing SFCC branch was generally measured in the selected studies, while several studies measured both the thawing and freezing branches. In addition, the supercooling portion on the freezing branch was abandoned when collecting UWC data, since it does not represent a real unfrozen water portion.

Special attention was paid to the SSA which is not available for a few soils. In this case, the SSA of these soils were either estimated or assumed in the present study. Several estimation methods have been proposed in the literature. For example, Ismeik and AI-Rawi (2014) suggested using equivalent diameter from the PSD to estimate SSA. Ersahin et al. (2006) highlighted that fractal dimensions for PSD can be used as an integrating index in estimating SSA. However, the soils collected in the present study do not necessarily have a PSD information, making these two methods not applicable. On the other hand, according to Yukselen-Aksoy and Kaya (2010), there is high correlation between the soil SSA and its liquid limit or plasticity index. As soil consistency limit values are generally available, the SSA of several soils was estimated by the relationship between SSA and plasticity index, suggested by Yukselen-Aksoy and Kaya (2010) (Eq. (7) in their study). However, for those soils that do not have a plasticity index, such as sand, their SSA were assumed according to typical values for those types of soils.

2.2 Artificial neural network model

The ANN is one of the supervised ML models (Fan et al., 2021). Figure 2 shows a typical structural ANN model that contains one input layer, three hidden layers and one output layer. In the input layer, the number of neurons equals the number of input variables. The number of neurons in the hidden layers determines the nonlinear degree of the designed model. In the present study, only one neuron is used in the output layer as a regression model, which represents the predicted volumetric UWC. For each neuron in the ANN, the output vector can be determined by Eq. (1) (Dongare et al., 2012),

$$y_k = f\left(\sum_{r=1}^{I} \omega_{r,k} x_r\right) \tag{1}$$

where, y_k is the output of neuron k; x_r is the input values from neurons of previous layer; $\omega_{r,k}$ is the weight of each input value. The weight will be optimized in the forward and backward propagation process. $f(\bullet)$ is the activation function used to increase the nonlinear property during the propagation. In the present study, the 'ReLu' function is used as the activation function for the hidden layers and the 'Linear' activation function is used for the output layer.

As can be seen from the architecture of ANN model, comparing to traditional statistical models, the advantage of using ANN model is that the model releases the fixed mathematical equation by combining the linear equation and activation function at each neuron. Therefore, no prior knowledge is required to predefine the relationship between the input variables and prediction target.

2.3 Bayesian optimization

The ANN model does not need any predefined relationship between the input and output variables; however, the final performance is heavily influenced by the architecture of the ANN model. A few hyperparameters inside the ANN model may influence its final prediction performance, such as the batch size, number of hidden layers, number of neurons in each layer, the type of optimizer and corresponding learning rate. In the present study, a Bayesian optimization method is used for tuning these hyperparameters to maximize the model's performance.

The Bayesian optimization used in this study is adopted from the scikit-optimization package (Head et al., 2018). In particular, the Bayesian optimization process aims to solve the optimization problem as shown in Eq. (2). As the target function $f(x)$ represents the loss value of ANN model that cannot get the gradient directly, a surrogate function is used to approximate the objective function. This surrogate function is represented by the Gaussian Processes in the present study. The next optimal hyperparameters are found by this surrogate function. After that, the surrogate function will be updated with the corresponding loss value. After repeating the inference and updating process for a certain number of iterations, the most optimal hyperparameters (x^*) can be finally determined (Wu et al., 2019).

$$x^* = \arg\min_{x} f(x) \tag{2}$$

where, x is the hyperparameters of the ANN model; $f(x)$ is the loss value of the ANN model applying on the test set; arg min is the objective function that aims to find the hyperparameters x to make the function $f(x)$ minimum.

2.4 k-folder cross validation

k-folder cross validation is a statistical method that used for ANN model's performance evaluation. The k-fold validation technique guarantees all samples in the dataset to be considered for both training and validation processes. The process of k-fold cross validation is illustrated in Fig. 3. The original dataset is randomly shuffled before the splitting. The shuffled dataset is split into five folds. After that, each fold is sequentially treated as test set and the rest folds are used to train the ANN model. For example, Fig. 3 shows the first cross validation which uses the fold 1 data set as test set and the rest folds data as training set. In the present study, the 5-folder cross validation is adopted. Therefore, the ANN model is trained and evaluated five times.

3. Data analysis and modelling results

3.1 Data distribution and correlation

As discussed earlier, in the present study four influencing factors are used for the prediction of volumetric UWC (θ_u) in frozen soils (i.e., the SSA, ρ_d, θ_{init}, and *Temp*). Table 2 summarizes the statistical properties of the considered features and θ_u, including the mean, standard deviation, minimum and maximum values. The preliminary data analysis shows the data range of the collected data, which also provides a reference for the application range of the final prediction model.

Figure 4 presents the histogram plots of the considered variables as well as the prediction target. As can be seen from Fig. 4(a), most of the collected samples have a SSA lower than 200 m^2/g. However, there are a few samples whose SSA is larger than 600 m^2/g. The distribution of initial volumetric water content is denser than that of SSA. Most samples' θ_{init} values are within 0.1 to 0.6 m^3/m^3. The dry density value ranges from 0.26 to 1.93 g/cm^3 with a mean value at 1.41 g/cm^3. Although the lowest temperature in the collected dataset is -64 °C, most samples were tested in the temperature range of 0 to -30 °C. Only a few samples whose testing temperature below -30 °C were collected. These samples are reserved in the model to fully utilize the collected data. In the end, the final UWC of the collected samples ranges from 0.00 to 0.91 m^3/m^3.

Figure 5 shows the correlation relationship among the input variables and the output. The highest correlation among the input variables is between the dry density and initial volumetric water content (-0.79), followed by the SSA (-0.62). However, well defined correlations were not observed between any input variable and the volumetric UWC. This demonstrates that the UWC prediction cannot rely on any single factor.

3.2 Bayesian optimization and training results

50 iterations were conducted for the Bayesian optimization as illustrated in Fig. 6. It represents the optimization process that the ANN was trained multiple times with different inferred hyperparameters. The objective of the optimization process is to increase the *R*-squared value when predicting the samples' UWC in the test set. The *R*-squared value is defined in Eq. (3). The closer of the predicted UWC to its measured counterpart, the higher *R*-squared value would be.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{3}$$

in which,

$$SS_{res} = \sum_{i} (y_i - y_i)^2 \tag{4}$$

$$SS_{tot} = \sum_{i} (y_i - \bar{y})^2 \tag{5}$$

where, SS_{res} is the sum of squares of residual, and SS_{tot} is the sum of squares of the original dataset; y_i is the measured UWC of each sample, and $\bar{y}$ is the corresponding average value; $\hat{y}_i$ is the predicted UWC.

The R-squared value of the test set increased from around 0.61 to 0.82. In the end, an ANN model with 64 batch size, 2 hidden layers with 128 neurons in each layer, and a learning rate at 0.004 was determined.

The optimal hyperparameters that obtained from Bayesian optimization were used to determine the final ANN model. The model was then evaluated with the original dataset by using 5-folder cross validation. The final performance of the ANN model at each fold can be seen in Fig. 7(a) and one of the prediction results is shown in Fig. 7(b). The results indicate an overall good performance of the ANN model, considering that the collected UWC data were determined under different experimental scenarios. In particular, folders 1, 2 and 4 achieve a R-squared value for the test set around 0.8. The average R-squared value among the five folders is 0.76.

3.3 Factor importance to the ANN model

Although the ANN model performs well, it is often criticized as a 'black box' since it cannot reveal the internal relationships among the input variables and prediction target. To solve this issue, a ML model interpreter, the SHAP interpretation (Lundberg and Lee, 2017) is adopted together with the LightGBM model trained with the balanced training dataset, to interpret the contribution of each influencing factor. SHAP presents a way to calculate the additive feature importance score for each factor (Strumbelj and Kononenko, 2010). The higher the importance score, the more important is the factor towards the final ML model prediction. The SHAP interpretation method together with decision-tree based ML algorithms have been widely used in the civil engineering applications, including some scenarios where highly correlated variables exist, such as the explanation of the failure of reinforced concrete (Mangalathu et al., 2020) and the roadway segment crashes (Wen et al., 2021).

Figure 8 shows the overall importance of the influencing factors considered on the UWC. It can be inferred that temperature has the largest impact, followed by the SSA. Furthermore, the initial water content has the lowest influence on the final UWC. The specific individual influence of each variable can also be analyzed by the adopted SHAP technique. As shown in Fig. 9, a positive influence of temperature on the UWC can be observed. This means that the UWC values are higher at higher temperature. Similar trends can also be observed for the SSA, i.e., larger SSA is related to higher UWC. These are consistent with general observations. However, the influence of dry density and initial water content is more controversial than other variables, which indicates that their influence is also dependent on other variables.

4. Comparison of the ANN model and two traditional models

4.1 Models description

Many models for the estimation of the UWC in frozen soils have been proposed in the literature. They can be generally classified into three types; namely, (i) empirical models, e.g., linear, power, and exponential relationships between UWC and subzero temperature and soil physical properties; (ii) models that employ SWCC expressions to represent the relation between UWC and subzero temperature, based on the similarity between frozen soils and unsaturated soils; (iii) physical and theoretical models, which take advantage of soil particle/pore size distribution, capillarity, adsorption, salt exclusion, and thermodynamic theories. In the present study, two models from the first two categories are selected and compared with the above developed ANN model. The physical and theoretical models are complex for use. Therefore, no model from this category is selected for comparison.

The first model is empirical and was proposed by Anderson and Tice (1972). They suggested that the UWC can be conveniently expressed as a function of subzero temperature by a simple power curve with two constants, which can be estimated from soil SSA. This empirical power relationship is one of the most widely used model in the literature. The model is expressed in terms of volumetric UWC (θ_u) below,

$$\begin{cases} \ln w_u = 0.2618 + 0.5519 \ln SSA - 1.449 SSA^{-0.264} \ln(-T) \\ \theta_u = \rho_d w_u / 100 \end{cases} \tag{6}$$

where, w_u is the gravimetric UWC; T is the subzero temperature, °C; ρ_d is soil dry density, g/cm^3.

The second model is based on the similarity between the SFCC and SWCC. This concept has been used in many studies (e.g., Nishimura et al., 2009; Liu and Yu, 2014; Ren et al., 2017; Teng et al., 2020). For example, Liu and Yu (2014) employed the Fredlund and Xing (1994) SWCC expression (Eq. (7)) to represent SFCC. The cryogenic suction in frozen soils is correlated with the subzero temperature through the Clapeyron equation, as shown in Eq. (8). In addition, there are many empirical relationships between the Fredlund and Xing model parameters (i.e., a, n, m, and ψ_{res}) and soil physical properties in the literature (e.g., Zapata et al., 2000; Witczak et al., 2006; Chin et al., 2010). For example, the relationships proposed by Zapata et al. (2000) are summarized in Table 3. Although these empirical relationships were developed on unsaturated soils, they are used to calculate the model parameters (a, n, m, and ψ_{res} in Eq. (7)) for frozen soils in the present study, assuming that there is exact similarity between unsaturated soils and frozen soils. After then, the volumetric UWC can be determined through Eq. (7),

$$\theta_u = \left[1 - \frac{\ln\left(1 + \frac{\psi_{cryo}}{\psi_{res}}\right)}{\ln\left(1 + \frac{10^6}{\psi_{res}}\right)}\right] * \left\{ \frac{\theta_{init}}{\left\{\ln\left[\exp(1) + \left(\frac{\psi_{cryo}}{a}\right)^n\right]\right\}^m} \right\} \tag{7}$$

$$\psi_{cryo} = -L\rho_w \ln \frac{T + 273.15}{T_0 + 273.15} \tag{8}$$

where, ψ_{cryo} is the cryogenic suction in kPa; ψ_{res} is the cryogenic suction at the residual state, kPa; θ_{init} is the initial volumetric UWC, m^3/m^3; L is the latent heat of fusion of water (L = 334 kJ/kg); ρ_w is the density of water (ρ_w = 1000 kg/m^3); T_0 is the normal freezing temperature of water (T_0 = 0 °C). The calculated cryogenic suction versus subzero temperature relationship by Eq. (8) is approximately linear with a slope of about 1225 kPa/°C, when the subzero temperature is not too low (Ren, 2019).

For comparing the above two traditional UWC models with the developed ANN model in the present study, four different types of soils were selected. It should be noted that the UWC datasets of these four soils were not included in the training or validating process when developing the ANN model. They are only used in the comparison of ANN prediction versus traditional models. In other words, these four soils provide completely independent datasets for comparing the three models, and therefore providing objective assessment of the model performance. Among the four soils, three soils are plastic and the last soil is non-plastic. It covers a variety of soil types, such as sand, silt, and high plastic bentonite. Their SSA is in a wide range, with the minimum of 3 m^2/g for fine sand and the maximum of 380.6 m^2/g for bentonite, as shown in Table 4. Therefore, the selected four soils are good representatives for model comparison. In addition, the soil physical properties that are essential for employing the three models are summarized in Table 4.

4.2 Comparison results

Figure 10 summarizes the prediction results by the two traditional models and the ANN model on four different soils with significantly different physical properties. It can be seen that the Fredlund and Xing (1994) model (with its parameters calculated by the empirical relationships suggested by Zapata et al. (2000)) is not able to provide accurate estimation of the UWC for these four soils. This approach typically overpredicts the UWC for plastic soils but underestimates those for non-plastic soils. The Anderson and Tice (1972) model provides reasonable predictions for silt; however, the estimations for the other three soils are poor. In addition, this model results in unreasonably high UWC value when the subzero temperature is close to 0 °C (see Fig. 10(c)). Meanwhile, the UWC values of the four soils predicted by the ANN model are close to the measured data points, suggesting that the performance of the pre-trained ANN model is good.

Figure 11 presents the comparison between the ANN model and the two traditional models; the measured UWC values are plotted on the abscissa. This figure clearly shows that the UWC values predicted by the two traditional models deviate from the 1:1 line, with most of the data outside the ±15% absolute percentage error lines. In other words, the two models either over- or under-estimate the UWC of the four soils. However, the prediction results by the ANN model are closer to the 1:1 line, compared with the two traditional models. This suggests that the ANN model outperforms the two traditional models, and has higher prediction accuracy.

For better illustration, the root mean squared error (RMSE) for the three models is shown in Fig. 12. It clearly shows that the ANN model generally has much smaller RMSE values for the four soils, compared with the other two models. The Anderson and Tice (1972) model provides fair estimations for three types of soils but fails on the bentonite. The Fredlund and Xing (1994) approach had the worst overall performance among the three models, which means soil specific calibrations are crucial for the performance of this model.

5. Discussion

5.1 Concerns regarding the ANN model development

There are a variety of internal and external factors that influence the UWC in frozen soils. Therefore, estimating UWC is ideally suited by ML models such as ANN, which is good at learning the highly nonlinear relationships among complex factors. In the present study, an ANN model was established and trained based on the UWC data collected from the literature. The amount of UWC data used in this study, however, is still limited. This is because the influencing factors that were considered in various published studies in the literature are different and sometimes arbitrary. This limitation contributes to the discrepancies among the collected data. Therefore, there is a need to set up large and reliable database for UWC, which can facilitate the establishment of robust and widely applicable ML models for UWC estimation.

A search of more than 100 articles in the literature resulted in 20 articles (and 73 soils in total) that contain the proper types of data. In order to obtain enough amount of data for developing the ANN model, UWC data were selected regardless of the testing methods, hysteresis effect, freeze-thaw cycles, or salt concentrations. This on one hand highlights the versatility of the developed ANN model. On the other hand, ignoring hysteresis means that both freezing and thawing UWC data were used. This partially contributes to prediction error. For example, the data point (i.e., 0.181 m^3/m^3) in Fig. 10(b) is on the freezing branch, which is at higher position than many other data points that are on the thawing branch. However, this limitation can be alleviated if the experimental UWC data on freezing and thawing branches are separately collected and used for establishing ANN models. Another issue that influences the predicting accuracy of the developed ANN model is that the experimental data (used for training and validation) themselves have some fluctuations or discrepancies. For example, the discrepancy originated from the fact that different measurement techniques yield different UWC values even for the same soil sample.

It is possible to use part of the data collected from the 20 articles, such that more influencing factors (e.g., salinity and sand/silt/clay fraction) can be included to develop ANN models. However, the present study limited its goal to use as much data as possible to ensure a stable and reliable ANN model. A smaller range of data used for model development would also limit its application scope and yield less reliable estimation results. In addition, Pham et al. (2019) opinioned that including additional specific information to input features could affect the representative capacity of the model because such information, in some cases, could not be easily obtained in practice. The way to develop an ANN model with more influencing factors essentially follows the same framework highlighted in the present study. Once more data are available, the present ANN model can be easily extended in the future for improving its capacity and performance.

Géron (2017) pointed out that in ANN modeling several hyperparameters, such as the ANN structure, number of training steps and regularization coefficient, should be aligned. Determining the most suitable combination of hyperparameters for a given task can be challenging. The developed ANN model shows good performance on the test dataset. The model performance may be further improved by developing ensemble or stacked models, applying transfer learning, or performing domain knowledge modification (Zhong et al., 2021). In addition, according to Zhong et al. (2021), the first step for developing a sound ANN model is to build a large, consistent source dataset. Unfortunately, such a large dataset is currently not available for the UWC data in the literature.

In the present study, four influencing factors (i.e., SSA, dry density, initial volumetric water content and temperature) were employed as the input variables for estimating UWC. The SHAP analysis shows that temperature and SSA are the two factors that significantly influence the UWC in frozen soils, which is in agreement with general observations. It also indicates that the initial water content does not have significant effect on UWC. In addition, the effect of density (or void ratio) on UWC is not predominant, which is consistent with the study by Wang et al. (2017b).

5.2 The strengths and limitations of different models

The Anderson and Tice (1972) model is empirical and simple. It uses the SSA and subzero temperature as two independent variables for the calculation of UWC. Although this model was established based on several soils with a variety of SSA values, it was not able to accurately predict the UWC of three of the four selected soils. Therefore, this model should be further improved using additional experimental data on different types of soils. It is likely a robust correlation could be achieved between the UWC and SSA, and its parameters by including additional experimental results. Another limitation of this model is that it yields a UWC value of infinity when the subzero temperature approaches to 0 °C. This problem has also been observed by other researchers (e.g., Michalowski, 1993; Qin et al., 2008).

Using the Fredlund and Xing (1994) SWCC expression in the estimation of the UWC in frozen soils is a semi-empirical approach and lacks theoretical foundation. This approach employs the similarity between the SFCC and SWCC, and directly replaces the suction in unsaturated soils by the cryogenic suction in frozen soils, which is calculated from subzero temperature by using the Clapeyron equation. The validity of the Clapeyron equation generally involves two assumptions; (i) thermodynamic equilibrium at the pore ice–water interface in the frozen soil, and (ii) the pore ice pressure is equal to the atmospheric pressure. In spite that these assumptions have been widely accepted as reasonable working hypotheses by many studies, some aspects of the underlying theory have been recently disputed in the literature (Vogel et al., 2019; Zhang et al., 2021a). For example, it is likely that the thermodynamic process in freezing soil is non-equilibrium, and pore ice pressure may deviate from the atmospheric pressure in unsaturated frozen soil or when overburden pressure is present. More discussions related to the similarity between freezing and drying processes are available in Ren and Vanapalli (2019). It should also be noted that for this model, its parameters were determined based on empirical relationships, which were derived from unsaturated soils. The failure of using this model in the reliable prediction of UWC data suggests that the similarity and differences between the SFCC and SWCC deserves more rigorous investigations.

Mu (2017) suggests that the empirical and SWCC-derived models may not provide reliable UWC values over a wide temperature range due to lack of consideration of the influence of both capillarity and adsorption. Furthermore, the effect of initial soil void ratio (which influences the capillarity) on the UWC was not explicitly considered in these models. On the other hand, the ANN model considered the effect of void ratio by incorporating the dry density as an input variable. In addition, the empirical models lack a theoretical basis in terms of continuum thermodynamics (Qin et al., 2008). Furthermore, although some of these models have been successfully employed to best-fit the measured UWC data, they are not readily to be used since the fitting parameters are generally based on a limited number of soils data. As a result, it is not surprising that these fitting parameters cannot be used for estimating the UWC of other soils such as the four types of soils analyzed in this study.

The comparison between the above two traditional models and developed ANN model shows better performance of the latter. The ANN model has good applicability in frozen soils. It can be applied to estimate the UWC of a variety of soils that were not employed for developing the ANN model, and that of the soils used for training the model. However, one limitation of the ANN model is that monotonic estimation of UWC cannot be guaranteed. For example, it can be seen from Fig. 10(d) that a spike exists and the predicted UWC does not strictly monotonically decrease with the decrement of temperature. The reason is that while ANN model uses thousands of neurons to free from a fixed statistical model, there is no strict equation to guarantee its output to be monotonic versus the temperature. Hence, the ANN model predicts the UWC at each temperature separately. Making the ANN model realizing the monotonicity in datasets requires more studies (Bandai and Ghezzehei, 2021).

The model from the third category (i.e., physical and theoretical model) was not selected for comparisons. This is because such models generally involve several theories, assumptions, parameters and approximations, resulting in inconvenient use of these models. Compared with the macroscopic empirical and semi-empirical models from the first two categories, the physical and theoretical models consider microscopic perspectives including in certain models at molecular levels. For example, the theoretical model proposed by Watanabe and Mizoguchi (2002) separately calculate the UWC in soil pores and that exists on particle surfaces as film water. The former is based on pore size distribution and Gibbs-Thomson effect, and the latter takes advantage of the specific surface area and thickness of the water film. The sum of the two is the total UWC in the frozen soil. Similar concepts have been widely employed by recent studies. However, as pointed by Fisher et al. (2019) that in order to use such models on natural soils, detailed information of the soil properties is needed. They include such as the pore size diameters and distribution, specific surface area, surface energy of the ice–water interface, dielectric permittivity, and Hamaker constant, which would own multiple values since soil is a complex and heterogeneous porous system (Watanabe and Mizoguchi, 2002). As a result, the application of such models can be challenging.

6. Summary

The effects of climate change on the permafrost and seasonally frozen regions and the increasing civil infrastructure construction in these regions have stimulated extensive research studies related to the behaviors of frozen soils in recent years. It is well-known that unfrozen water and pore ice coexist in the frozen soil as a result of complex soil-water interactions. The relative quantity of the unfrozen water and ice has paramount influence on the physical and mechanical properties of frozen soils, as well as on the transport of energy, water and solutes in cold regions. Due to this reason, a variety of techniques have been developed and employed to measure the unfrozen water and ice contents in frozen soils, and many models have also been proposed for the estimation of UWC in the past several decades. These proposed models are generally based on using soil physical properties, the similarity between frozen soils and unsaturated soils, and / or physical and theoretical mechanisms.

Many factors influence the UWC in frozen soils. These factors include such as soil physical and chemical properties, stress sate, and temperature. The complex effects of these factors result in a highly nonlinear relationship between these factors and UWC. In addition, the relative contribution of each factor on UWC is not well-understood. Furthermore, the previously developed statistical models generally can only incorporate a few influencing factors and therefore have limited predicting capability. Such limitations, however, can be effectively addressed by using ML algorithms, such as the ANN models.

In the present study, extensive UWC data of various types of soils tested under various conditions were collected through a comprehensive search of the literature. An ANN model for estimating the UWC in frozen soils was developed following the proposed modeling framework. The ANN model was established by using the PyTorch package and its hyperparameters were optimized with Bayesian optimization. The developed ANN model showed good performance on the test dataset. In addition, it was compared with two traditional statistical models for UWC prediction on four independent types of soils. The results indicated that the ANN model achieved better UWC prediction performance than its counterparts, which include the empirical model and semi-empirical model exploiting the similarity between frozen soils and unsaturated soils. Detailed discussions on the developed ANN model, and the strengths and limitations of different types of models were also presented. The present study demonstrates the potential of ML model to provide reliable prediction of the UWC in frozen soils. In addition, the large amount of UWC data collected and the developed ANN model will be great assets for future studies.

Acknowledgements

The first author gratefully acknowledges the financial support from the Fundamental Research Funds for the Central Universities (lzujbky-2021-kb03).

References

Akagawa, S., Iwahana, G., Watanabe, K., Chuvilin, E.M., Istomin, V.A. and Hinkel, K.M., 2012, June. 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 (Vol. 1, pp. 21-26).

Amankwah, S.K., Ireson, A.M., Maulé, C., Brannen, R., and Mathias, S.A., 2021. A Model for the Soil Freezing Characteristic Curve That Represents the Dominant Role of Salt Exclusion. Water Resources Research, 57(8), e2021WR030070.

Amiri, E.A., Craig, J.R. and Kurylyk, B.L., 2018. A theoretical extension of the soil freezing curve paradigm. Advances in Water Resources, 111: 319-328.

Anderson, D.M. and Tice, A.R., 1972. Predicting unfrozen water contents in frozen soils from surface area measurements. Highway research record, 393(2), pp.12-18.

Bai, R., Lai, Y., Zhang, M. and Yu, F., 2018. Theory and application of a novel soil freezing characteristic curve. Applied Thermal Engineering, 129: 1106-1114.

Bandai, T., and Ghezzehei, T.A., 2021. Physics-informed neural networks with monotonicity constraints for Richardson-Richards equation: Estimation of constitutive relationships and soil water flux density from volumetric water content measurements. Water Resources Research, 57(2), e2020WR027642.

Chai, M., Zhang, J., Zhang, H., Mu, Y., Sun, G. and Yin, Z., 2018. A method for calculating unfrozen water content of silty clay with consideration of freezing point. Applied Clay Science, 161, pp.474-481.

Chin, K. B., Leong, E. C., and Rahardjo, H., 2010. A simplified method to estimate the soil-water characteristic curve. Canadian Geotechnical Journal, 47(12), 1382-1400.

Dillon, H.B. and Andersland, O.B., 1966. Predicting unfrozen water contents in frozen soils. Canadian geotechnical journal, 3(2): 53-60.

Dongare, A.D., Kharde, R. R., and Kachare, A.D., 2012. Introduction to artificial neural network. International Journal of Engineering and Innovative Technology (IJEIT), 2(1), 189-194.

Ersahin, S., Gunal, H., Kutlu, T., Yetgin, B., and Coban, S., 2006. Estimating specific surface area and cation exchange capacity in soils using fractal dimension of particle-size distribution. Geoderma, 136(3-4), 588-597.

Erzin, Y., Rao, B.H., and Singh, D.N., 2008. Artificial neural network models for predicting soil thermal resistivity. International Journal of Thermal Sciences, 47(10), 1347-1358.

Fan, X., Zhang, X., and Yu, X.B., 2021. Machine learning model and strategy for fast and accurate detection of leaks in water supply network. Journal of Infrastructure Preservation and Resilience, 2(1), 1-21.

Fisher, D. A., Lacelle, D., and Pollard, W., 2019. A model of unfrozen water content and its transport in icy permafrost soils: effects on ground ice content and permafrost stability. Permafrost and Periglacial Processes, 1-16.

Fredlund, D.G., and Xing, A., 1994. Equations for the soil-water characteristic curve. Canadian geotechnical journal, 31(4), 521-532.

Géron, Aurélien, 2017. Hands-on Machine Learning with Scikit-Learn & TensorFlow.

Habibagahi, G., and Bamdad, A., 2003. A neural network framework for mechanical behavior of unsaturated soils. Canadian Geotechnical Journal, 40(3), 684-693.

Head, T., G. L. MechCoder and I. Shcherbatyi, 2018. scikit-optimize/scikit-optimize: v0. 5.2. Zenodo.

Hu, G., Zhao, L., Zhu, X., Wu, X., Wu, T., Li, R., ... and Hao, J., 2020. Review of algorithms and parameterizations to determine unfrozen water content in frozen soil. Geoderma, 368, 114277.

Ismeik, M. and Al-Rawi, O., 2014. Modeling soil specific surface area with artificial neural networks. Geotechnical Testing Journal, 37(4), pp.678-688.

Jin, X., Yang, W., Gao, X., Zhao, J. Q., Li, Z., and Jiang, J., 2020. Modeling the unfrozen water content of frozen soil based on the absorption effects of clay surfaces. Water Resources Research, 56(12), e2020WR027482.

Kong, L., Wang, Y., Sun, W. and Qi, J., 2020. Influence of plasticity on unfrozen water content of frozen soils as determined by nuclear magnetic resonance. Cold Regions Science and Technology, 172, p.102993.

Kruse, A.M. and Darrow, M.M., 2017. Adsorbed cation effects on unfrozen water in fine-grained frozen soil measured using pulsed nuclear magnetic resonance. Cold Regions Science and Technology, 142, pp.42-54.

Kurylyk, B. L., and Watanabe, K., 2013. The mathematical representation of freezing and thawing processes in variably-saturated, non-deformable soils. Advances in Water Resources, 60, 160-177.

Lai, Y., Pei, W., Zhang, M., and Zhou, J., 2014. Study on theory model of hydro-thermal–mechanical interaction process in saturated freezing silty soil. International Journal of Heat and Mass Transfer, 78, 805-819.

Lara, R.P., Berg, A.A., Warland, J., & Parkin, G., 2021. Implications of measurement metrics on soil freezing curves: A simulation of freeze–thaw hysteresis. Hydrological Processes, 35(7), e14269.

Li, Z.M., Chen, J., and Sugimoto, M., 2020. Pulsed NMR Measurements of Unfrozen Water Content in Partially Frozen Soil. Journal of Cold Regions Engineering, 34(3), 04020013.

Liu, Z. and Yu, X., 2013. Physically based equation for phase composition curve of frozen soils. Transportation Research Record: Journal of the Transportation Research Board, (2349): 93-99.

Liu, Z. and Yu, X. 2014. Predicting the phase composition curve in frozen soils using index properties: A physico-empirical approach. Cold Regions Science and Technology, 108: 10-17.

Lovell Jr, C.W., 1957. Temperature effects on phase composition and strength of partially-frozen soil. Highway Research Board Bulletin, (168).

Lu, J., Pei, W., Zhang, X., Bi, J., and Zhao, T., 2019. Evaluation of calculation models for the unfrozen water content of freezing soils. Journal of Hydrology, 575, 976-985.

Lu, N., Likos, W. J., Luo, S., and Oh, H., 2021. Is the Conventional Pore Water Pressure Concept Adequate for Fine-Grained Soils in Geotechnical and Geoenvironmental Engineering?. Journal of Geotechnical and Geoenvironmental Engineering, 147(10), 02521001.

Lundberg, S. and Lee, S.I., 2017. A unified approach to interpreting model predictions. arXiv preprint arXiv:1705.07874.

Ma, T., Wei, C., Xia, X., Zhou, J. and Chen, P., 2015. Soil freezing and soil water retention characteristics: Connection and solute effects. Journal of performance of constructed facilities, 31(1), p.D4015001.

Mangalathu, S., Hwang, S.H. and Jeon, J.S., 2020. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. Engineering Structures 219: 110927.

Mao, Y., Romero Morales, E.E. and Gens Solé, A., 2018. Ice formation in unsaturated frozen soils. In Unsaturated Soils: UNSAT 2018: The 7th International Conference on Unsaturated Soils (pp. 597-602). The Hong Kong University of Science and Technology (HKUST).

McKenzie, J.M., Voss, C.I., and Siegel, D.I., 2007. Groundwater flow with energy transport and water–ice phase change: numerical simulations, benchmarks, and application to freezing in peat bogs. Advances in water resources, 30(4), 966-983.

Michalowski, R.L., 1993. A constitutive model of saturated soils for frost heave simulations. Cold regions science and technology, 22(1), 47-63.

Ming, F., Li, D.Q., and Liu, Y.H., 2020. A predictive model of unfrozen water content including the influence of pressure. Permafrost and Periglacial Processes, 31(1), 213-222.

Mu, Q.Y., 2017. Hydro-mechanical behaviour of loess at elevated and sub-zero temperatures. Doctoral dissertation, The Hong Kong University of Science and Technology (HKUST), Hong Kong, China.

Mu, Q.Y., Ng, C.W.W., Zhou, C., Zhou, G.G.D. and Liao, H.J., 2018. A new model for capturing void ratio-dependent unfrozen water characteristics curves. Computers and Geotechnics, 101, pp.95-99.

Nishimura, S., Gens, A., Olivella, S., and Jardine, R.J., 2009. THM-coupled finite element analysis of frozen soil: formulation and application. Géotechnique, 59(3), 159-171.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., ... and Lerer, A., 2017. Automatic differentiation in pytorch. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

Pham, K., Kim, D., Yoon, Y., and Choi, H., 2019. Analysis of neural network based pedotransfer function for predicting soil water characteristic curve. Geoderma, 351, 92-102.

Qin, Y.H., Zhang, J.M., Zheng, B., Qu, G.Z., 2008. The relationship between unfrozen water content and temperature based on continuum thermodynamics. Journal of Qingdao University: Engineering and Technology, 23(1), pp.77-82. (in Chinese)

Ren, J., 2019. Interpretation of the frozen soils behavior extending the mechanics of unsaturated soils. Doctoral dissertation, University of Ottawa, Ottawa, Canada.

Ren, J., and Vanapalli, S.K., 2019. Comparison of soil-freezing and soil-water characteristic curves of two Canadian soils. Vadose Zone Journal, 18(1), 1-14.

Ren, J., and Vanapalli, S.K., 2020. Effect of freeze–thaw cycling on the soil-freezing characteristic curve of five Canadian soils. Vadose Zone Journal, 19(1), e20039.

Ren, J., Vanapalli, S. K., and Han, Z., 2017. Soil freezing process and different expressions for the soil-freezing characteristic curve. Sciences in Cold and Arid Regions, 9(3), 221-228.

Ren, J., Vanapalli, S. K., Han, Z., Omenogor, K. O., and Bai, Y., 2019. The resilient moduli of five Canadian soils under wetting and freeze-thaw conditions and their estimation by using an artificial neural network model. Cold Regions Science and Technology, 168, 102894.

Ren, J., Zhang, S., Wang, C., Ishikawa, T. and Vanapalli, S.K., 2021. The Measurement of Unfrozen Water Content and SFCC of a Coarse-Grained Volcanic Soil. Journal of Testing and Evaluation, 51(1).

Saberi, P.S. and Meschke, G., 2021. A hysteresis model for the unfrozen liquid content in freezing porous media. Computers and Geotechnics, 134, p.104048.

Saha, S., Gu, F., Luo, X., and Lytton, R.L., 2018. Use of an artificial neural network approach for the prediction of resilient modulus for unbound granular material. Transportation Research Record, 2672(52), 23-33.

Shahin, M.A., Jaksa, M. B., and Maier, H.R., 2001. Artificial neural network applications in geotechnical engineering. Australian geomechanics, 36(1), 49-62.

Shahin, M.A., Jaksa, M. B., and Maier, H.R., 2008. State of the art of artificial neural networks in geotechnical engineering. Electronic Journal of Geotechnical Engineering, 8(1), 1-26.

Shastri, A., 2014. Advanced coupled THM analysis in geomechanics. Doctoral dissertation, Texas A&M University, College Station, USA.

Smith, M.W. and Tice, A.R., 1988. Measurement of the unfrozen water content of soils. comparison of NMR (Nuclear Magnetic Resonance) and TDR (Time Domain Reflectometry) methods (No. CRREL-88-18). COLD REGIONS RESEARCH AND ENGINEERING LAB HANOVER NH.

Strumbelj, E. and I. Kononenko, 2010. An efficient explanation of individual classifications using game theory. The Journal of Machine Learning Research 11: 1-18.

Suzuki, S., 2004. Dependence of unfrozen water content in unsaturated frozen clay soil on initial soil moisture content. Soil science and plant nutrition, 50(4), 603-606.

Teng, J., Kou, J., Yan, X., Zhang, S., and Sheng, D., 2020. Parameterization of soil freezing characteristic curve for unsaturated soils. Cold Regions Science and Technology, 170, 102928.

Vogel, T., Dohnal, M., Votrubova, J., and Dusek, J., 2019. Soil water freezing model with non-iterative energy balance accounting. Journal of Hydrology, 578, 124071.

Wang, C., Lai, Y. and Zhang, M., 2017a. Estimating soil freezing characteristic curve based on pore-size distribution. Applied Thermal Engineering, 124: 1049-1060.

Wang, J., Nishimura, S., and Tokoro, T., 2017b. Laboratory study and interpretation of mechanical behavior of frozen clay through state concept. Soils and Foundations, 57(2), 194-210.

Wang, M., Li, X., Liu, Z., Liu, J. and Chang, D., 2020a. Application of PIV Technique in Model Test of Frost Heave of Unsaturated Soil. Journal of Cold Regions Engineering, 34(3), p.04020014.

Wang, M., Li, X., and Xu, X., 2021. An implicit Heat-Pulse-Probe method for measuring the soil ice content. Applied Thermal Engineering, 117186.

Wang, Q., Liu, Y., Zhang, X., Fu, H., Lin, S., Song, S. and Niu, C., 2020b. Study on an AHP-entropy-ANFIS model for the prediction of the unfrozen water content of sodium-bicarbonate-type salinization frozen soil. Mathematics, 8(8), p.1209.

Watanabe, K. and Mizoguchi, M., 2002. Amount of unfrozen water in frozen porous media saturated with solution. Cold Regions Science and Technology, 34(2), pp.103-110.

Watanabe, K. and Wake, T., 2009. Measurement of unfrozen water content and relative permittivity of frozen unsaturated soil using NMR and TDR. Cold Regions Science and Technology, 59(1), pp.34-41.

Wen, H., Bi, J., and Guo, D., 2020. Calculation of the thermal conductivities of fine-textured soils based on multiple linear regression and artificial neural networks. European Journal of Soil Science, 71(4), 568-579.

Wen, X., Y. Xie, L. Wu and L. Jiang, 2021. Quantifying and comparing the effects of key risk factors on various types of roadway segment crashes with LightGBM and SHAP. Accident Analysis & Prevention 159: 106261.

Wen, Z., Ma, W., Feng, W., Deng, Y., Wang, D., Fan, Z., and Zhou, C., 2012. Experimental study on unfrozen water content and soil matric potential of Qinghai-Tibetan silty clay. Environmental earth sciences, 66(5), 1467-1476.

Wettlaufer, J.S., 1999. Ice surfaces: macroscopic effects of microscopic structure. Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 357(1763), 3403-3425.

Witczak, M., Zapata, C., Houston, W., 2006. Models incorporated into the current enhanced integrated climatic model: NCHRP 9–23 project findings and additional changes after version 0.7. Final Report, Project NCHRP.

Wu, J., Chen, X.Y., Zhang, H., Xiong, L.D., Lei, H. and Deng, S.H., 2019. Hyperparameter optimization for machine learning models based on Bayesian optimization. Journal of Electronic Science and Technology 17(1): 26-40.

Xiao, Z., Lai, Y., and Zhang, J., 2020. A thermodynamic model for calculating the unfrozen water content of frozen soil. Cold Regions Science and Technology, 172, 103011.

Xu, X.Z., Oliphant, J.L. and Tice, A.R., 1985. Soil-water potential and unfrozen water content and temperature. Journal of Glaciology and Geocryology, 7(1), pp. 1-14. (in Chinese)

Yoshikawa, K. and Overduin, P.P., 2005. Comparing unfrozen water content measurements of frozen soil using recently developed commercial sensors. Cold Regions Science and Technology, 42(3), pp.250-256.

Yu, F., Guo, P., Lai, Y., and Stolle, D., 2020a. Frost heave and thaw consolidation modelling. Part 1: A water flux function for frost heaving. Canadian Geotechnical Journal, 57(10), 1581-1594.

Yu, F., Guo, P., Lai, Y., and Stolle, D., 2020b. Frost heave and thaw consolidation modelling. Part 2: One-dimensional thermohydromechanical (THM) framework. Canadian Geotechnical Journal, 57(10), 1595-1610.

Yukselen-Aksoy, Y. and Kaya, A., 2010. Method dependency of relationships between specific surface area and soil physicochemical properties. Applied Clay Science, 50(2), pp.182-190.

Zapata, C.E., Houston, W.N., Houston, S.L., Walsh, K.D., 2000. Soil–water characteristic curve variability. Advances in Unsaturated Geotechnics. American Society of Civil Engineers, pp. 84–124.

Zhang, L., Zhuang, Q., Wen, Z., Zhang, P., Ma, W., Wu, Q., and Yun, H., 2021a. Spatial state distribution and phase transition of non-uniform water in soils: Implications for engineering and environmental sciences. Advances in Colloid and Interface Science, 102465.

Zhang, P., Yin, Z.Y., and Jin, Y.F., 2021b. State-of-the-art review of machine learning applications in constitutive modeling of soils. Archives of Computational Methods in Engineering, 1-26.

Zhong, S., Zhang, K., Bagheri, M., Burken, J.G., Gu, A., Li, B., Ma, X., Marrone, B.L., Ren, Z.J., Schrier, J. and Shi, W., 2021. Machine Learning: New Ideas and Tools in Environmental Science and Engineering. Environmental Science & Technology.

Zhou, J.Z., Meng, X., Wei, C., and Pei, W., 2020. Unified soil freezing characteristic for variably-saturated saline soils. Water Resources Research, 56(7), e2019WR026648.

Zhou, J.Z., Tan, L., Wei, C.F. and Wei, H.Z., 2015. Experimental research on freezing temperature and super-cooling temperature of soil. Rock and Soil Mechanics, 36(3), pp.777-785. (in Chinese)

List of Tables

Table 1. Unfrozen water content testing information from the literature

Data origin	No. of Soils	Internal factors (Soil main physical information)				External factors	SFCC branches	Testing methods
Smith and Tice (1988)	25	SSA	θ_{init}	ρ_d	/	T	Thawing	NMR, TDR
Suzuki (2004)	1	SSA	θ_{init}	ρ_d	Organic content, EC, Fraction	T	Thawing	NMR, TDR
Yoshikawa and Overduin (2005)	2	SSA	θ_{init}	ρ_d	/	T	Freezing	NMR, FDR, TDT
Watanabe and Wake (2009)	4	SSA	θ_{init}	ρ_d	Porosity, C_u, EC, Ignition loss	T	Thawing	NMR
Ma et al. (2015)	2	SSA	θ_{init}	ρ_d	LL, PL, Fraction	T	Thawing	NMR
Kruse and Darrow (2017)	6	SSA	θ_{init}	ρ_d	Cation treatment, CEC, PSD	T	Both	NMR
Wang et al. (2020a)	1	SSA	θ_{init}	ρ_d	LL, PL, Fraction	T	Thawing	NMR
Zhou et al. (2020)	1	SSA	θ_{init}	ρ_d	LL, PL, Fraction, Salinity	T	Thawing	NMR
Lovell (1957)	3	SSA*	θ_{init}	ρ_d	LL, PI, PSD	T	//	Calorimetry
Akagawa et al. (2012)	4	SSA*	θ_{init}	ρ_d	LL, PL, PI	T	Both	NMR
Wen et al. (2012)	1	SSA*	θ_{init}	ρ_d	LL, PL, Fraction	T	//	NMR
Zhou et al. (2015)	1	SSA*	θ_{init}	ρ_d	LL, PL, Fraction	T	Both	NMR
Mu (2017)	1	SSA*	θ_{init}	ρ_d	LL, PL, PI, Fraction	T, F-T, Stress state	Both	TDR
Chai et al. (2018)	1	SSA*	θ_{init}	ρ_d	LL, PL, PI, PSD, Salinity, pH	T	Thawing	NMR
Mao et al. (2018)	1	SSA*	θ_{init}	ρ_d	LL, PL, Porosity, Fraction	T	Freezing	EC measurement
Kong et al. (2020)	5	SSA*	θ_{init}	ρ_d	LL, PL, PI, Fraction	T	Freezing	NMR
Li et al. (2020)	3	SSA*	θ_{init}	ρ_d	PSD	T	Both	NMR
Ren and Vanapalli (2020)	5	SSA*	θ_{init}	ρ_d	LL, PL, PSD, Porosity	T, F-T	Both	FDR
Teng et al. (2020)	3	SSA*	θ_{init}	ρ_d	LL, PL, Fraction	T	Both	NMR
Wang et al. (2021)	3	SSA*	θ_{init}	ρ_d	LL, PL, PSD	T	Thawing	NMR

Note: SSA: Specific surface area; SSA*: Calculated SSA based on soil plasticity index (PI). The SSA values of a few soils are assumed, since their PI are not available; θ_{init}: Initial volumetric water content; ρ_d: Dry density (For the soil in Chai et al. (2018), its ρ_d value was obtained by personal communication); T: Temperature; EC: Electrical conductivity; Fraction: Sand/Silt/Clay fraction by weight; C_u: Uniformity coefficient; LL: Liquid limit; PL: Plastic limit; CEC: Cation exchange capacity; PSD: Particle size distribution curve; F-T: Freeze-thaw cycles; /: Not available; //: Unknow.

Table 2. Statistical properties of the collected data

Item	Unit	Mean	Std	Min	Max
SSA	m^2/g	91.33	145.00	0.90	714.00
$\boldsymbol{\theta_{init}}$	m^3/m^3	0.39	0.16	0.07	0.83
$\boldsymbol{\rho_d}$	g/cm^3	1.41	0.36	0.26	1.93
Temp	°C	-6.15	7.31	-64.00	0.00
$\boldsymbol{\theta_u}$	m^3/m^3	0.12	0.11	0.00	0.91

Table 3. Correlations between the Fredlund and Xing model parameters and soil index properties

FX model parameter	Plastic soils (*PI* > 0)	Non-plastic soils (*PI* = 0)
a **(kPa)**	$a = 0.00364*(wPI)^{3.35} + 4*(wPI) + 11$	$a = 0.8627*(D_{60})^{-0.751}$
n	$n = [-2.313*(wPI)^{0.14} + 5]*m$	$n = 7.5$
m	$m = 0.0514*(wPI)^{0.465} + 0.5$	$m = 0.1772*\ln(D_{60}) + 0.7734$
$\boldsymbol{\psi_{res}}$ **(kPa)**	$\psi_{res} = 32.44*\exp(0.0186*(wPI))*a$	$\psi_{res} = a/(D_{60} + 9.7*\exp(-4))$
$wPI = P_{200} * PI$ where, P_{200} is the percentage passing the #200 U.S. standard sieve, as a decimal; *PI* is plasticity index, as a percentage; D_{60} is the particle size corresponding to 60% passing by weight, mm.		

Table 4. The four soils selected for model comparison

Soil ID	P_{200}	PI (%)	SSA (m^2/g)	θ_{init} (m^3/m^3)	ρ_d (g/cm^3)	Data origin
Silt	0.854	11.7	16.6	0.416	1.60	**Zhou et al. (2020)**
Loess	1	19.0	75.3	0.512	1.29	**Mu (2017)**
Bentonite	1	127.9	380.6	0.387	1.60	**Kong et al. (2020)**
Fine sand	0.23†		3.0	0.331	1.57	**Li et al. (2020)**

†: This value is the D_{60} (Unit: mm).

List of Figures

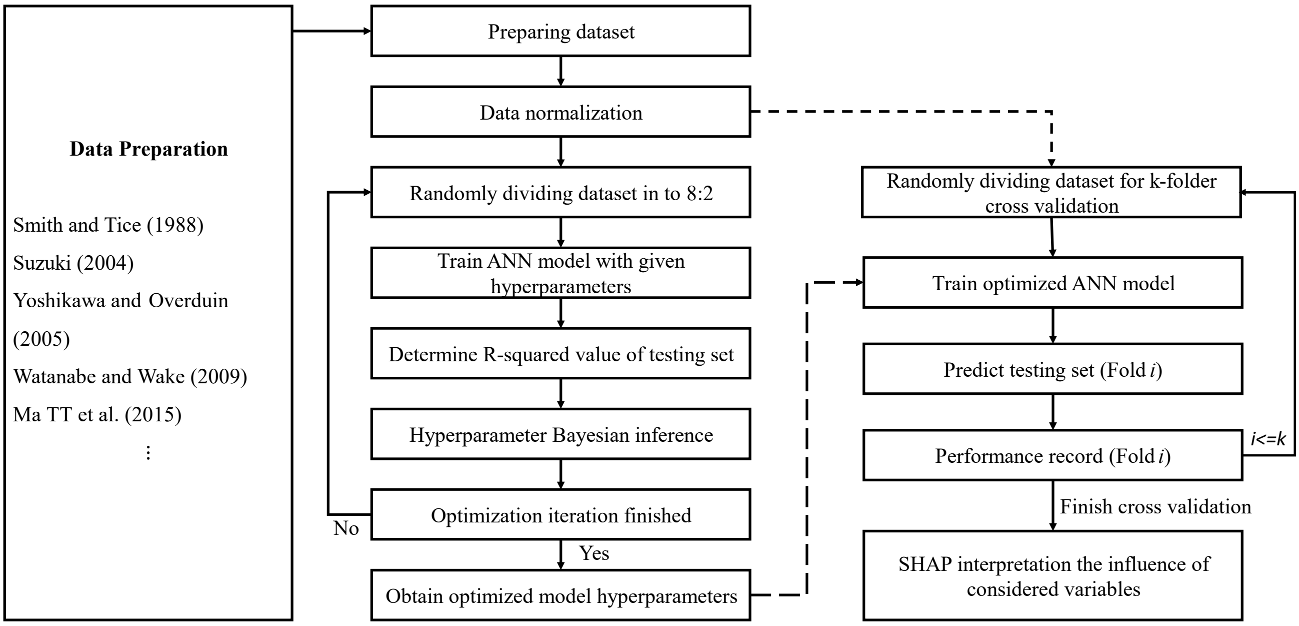


Fig. 1. Framework for unfrozen water content prediction

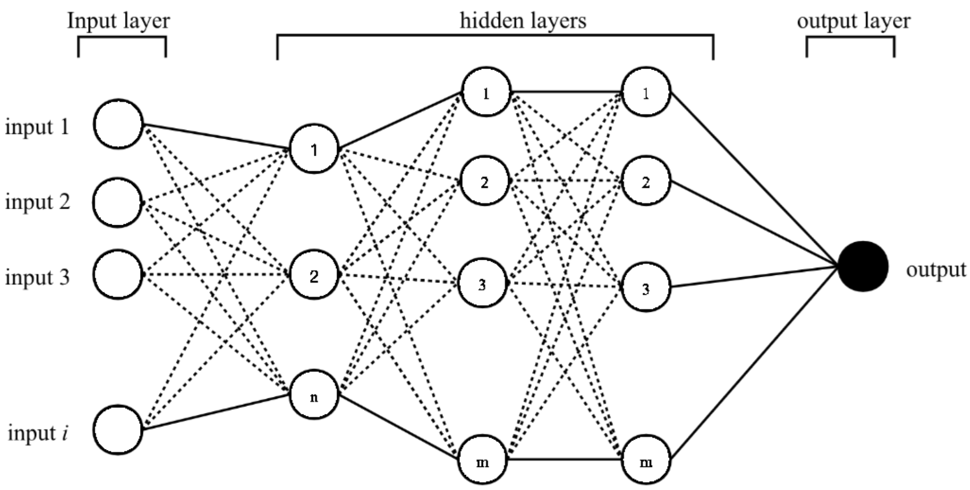


Fig. 2. Structure of ANN model with input layer, hidden layer, and output layer

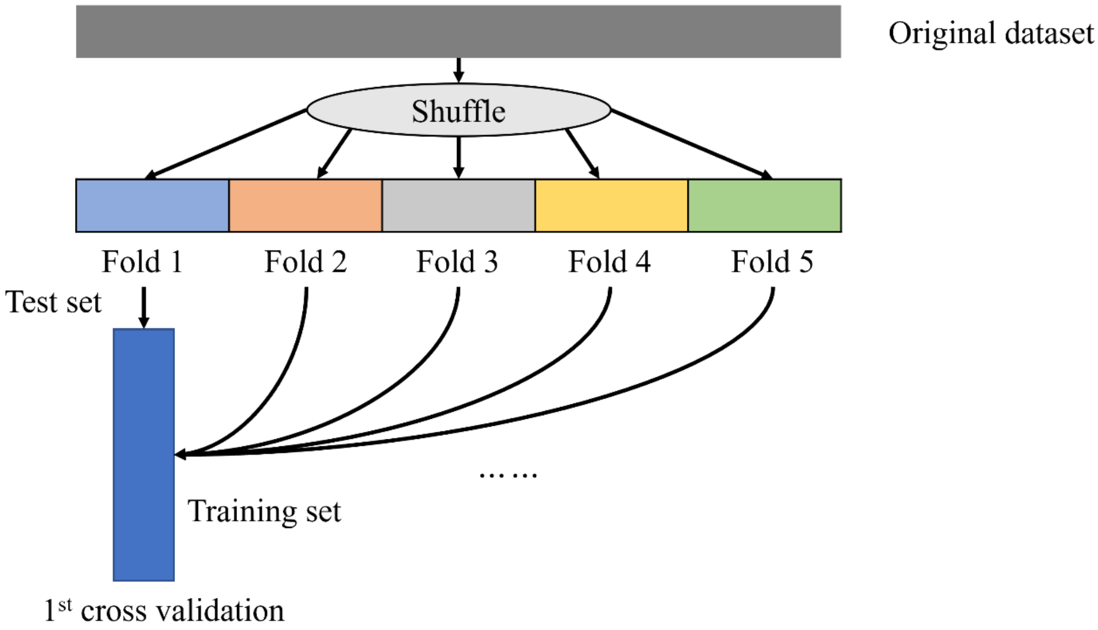
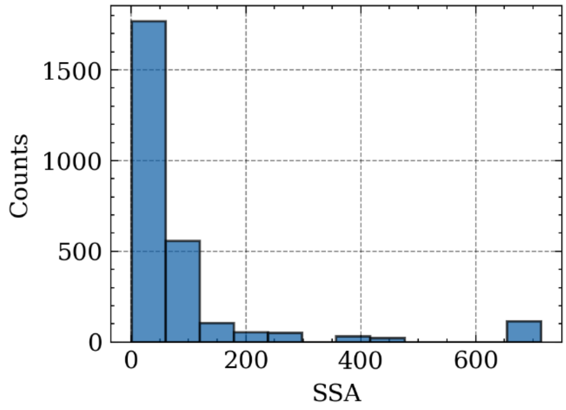
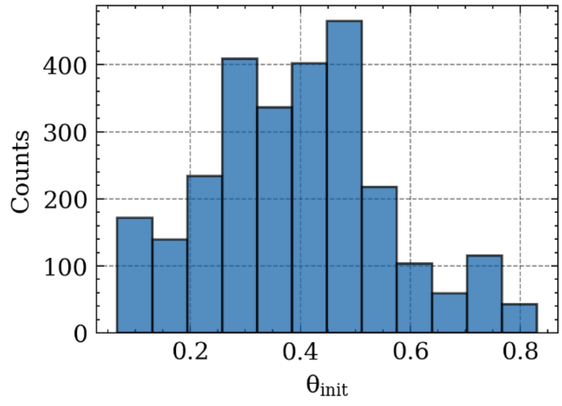


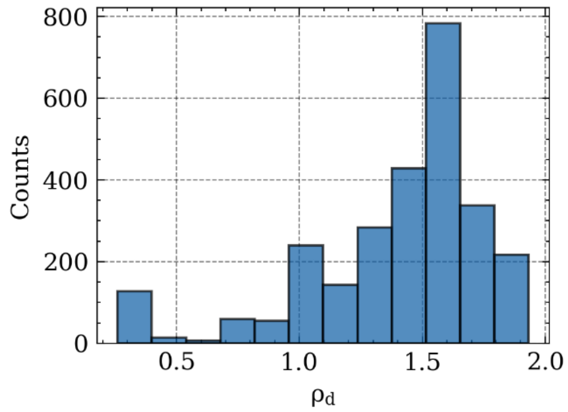
Fig. 3. Illustration of k-folder cross validation



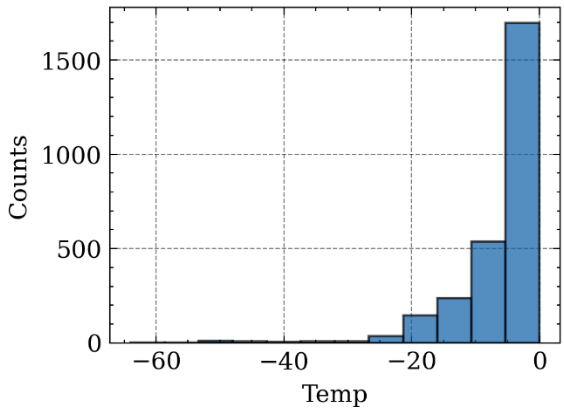
(a) Specific surface area



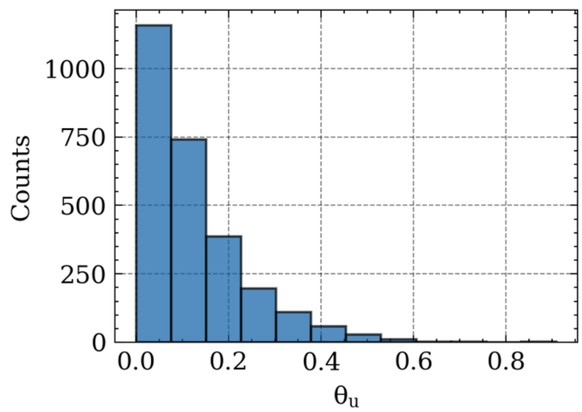
(b) Initial volumetric water content



(c) Dry density



(d) Temperature



(e) Unfrozen water content

Fig. 4. Histogram plot of the input variables and prediction target

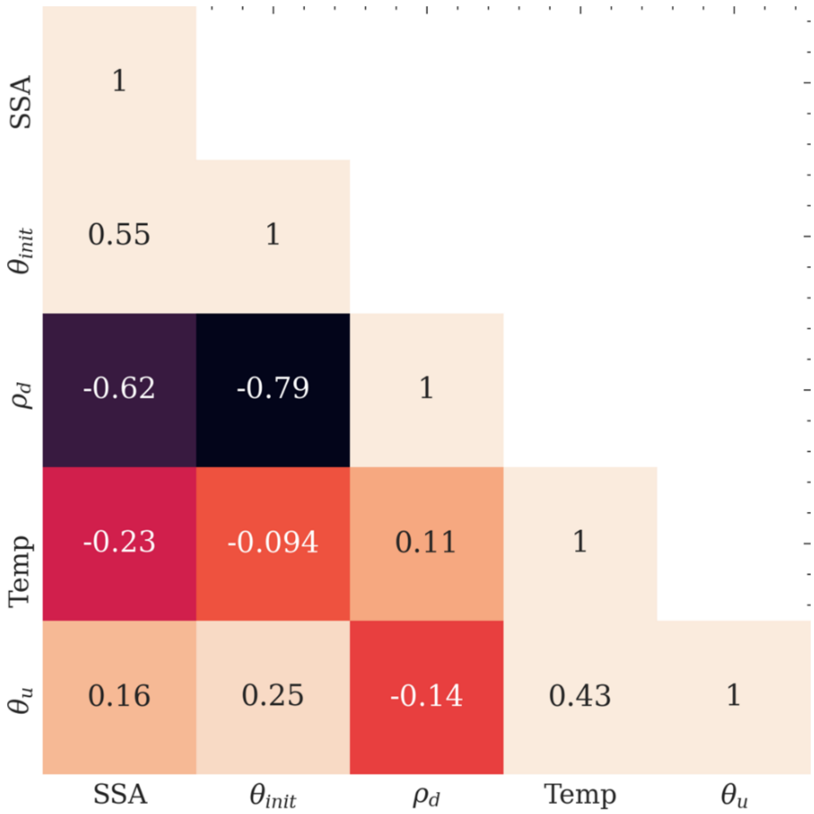


Fig. 5. Correlation map among the input variables and prediction target

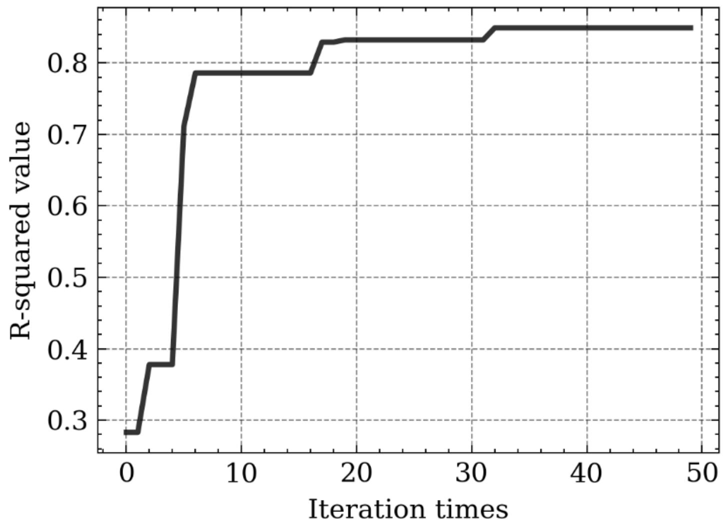
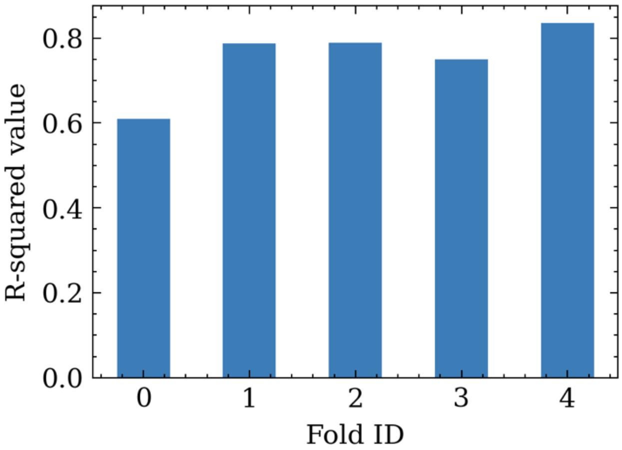
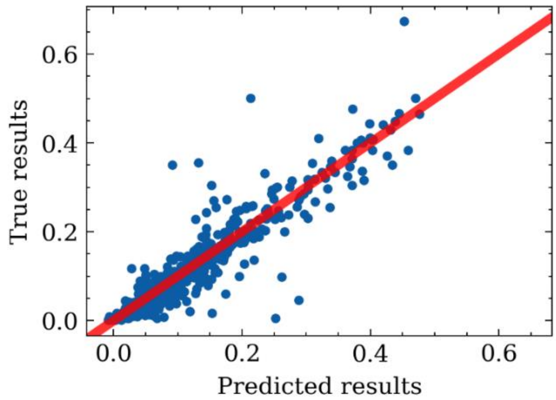


Fig. 6. Bayesian optimization process





(a) Prediction results of five folds **(b) Predicted water content in fold 2**

Fig. 7. Results of ANN model

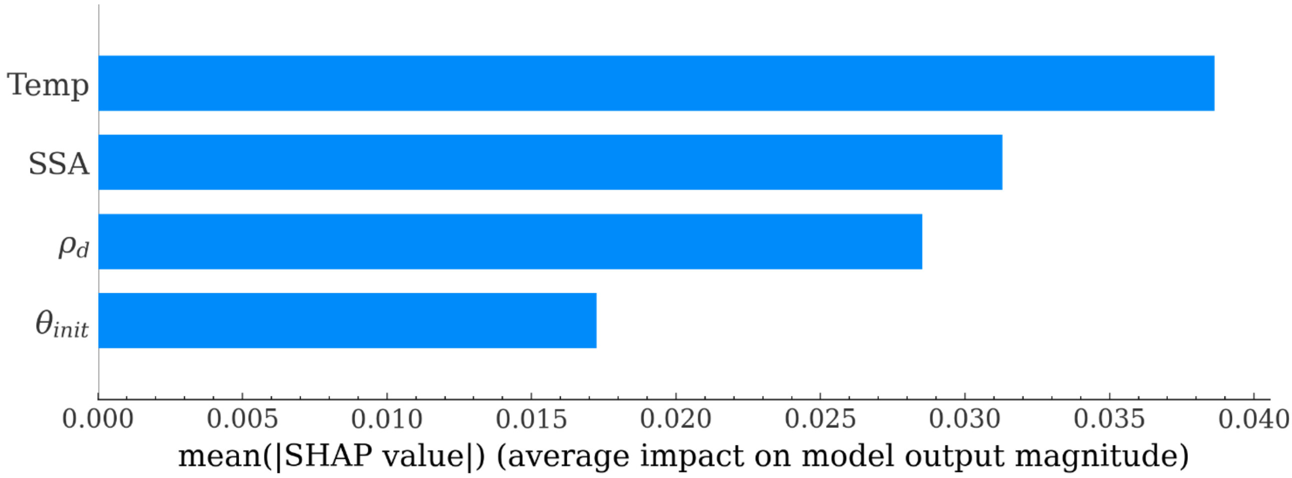


Fig. 8. Overall importance of the considered factors on the unfrozen water content

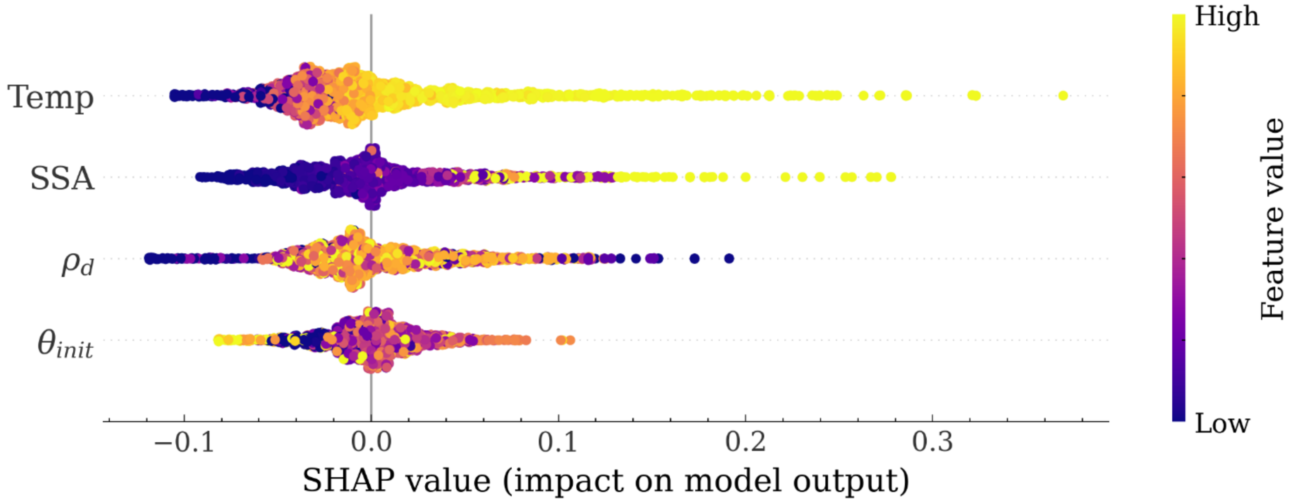


Fig. 9. Individual impact of the considered factors on the unfrozen water content

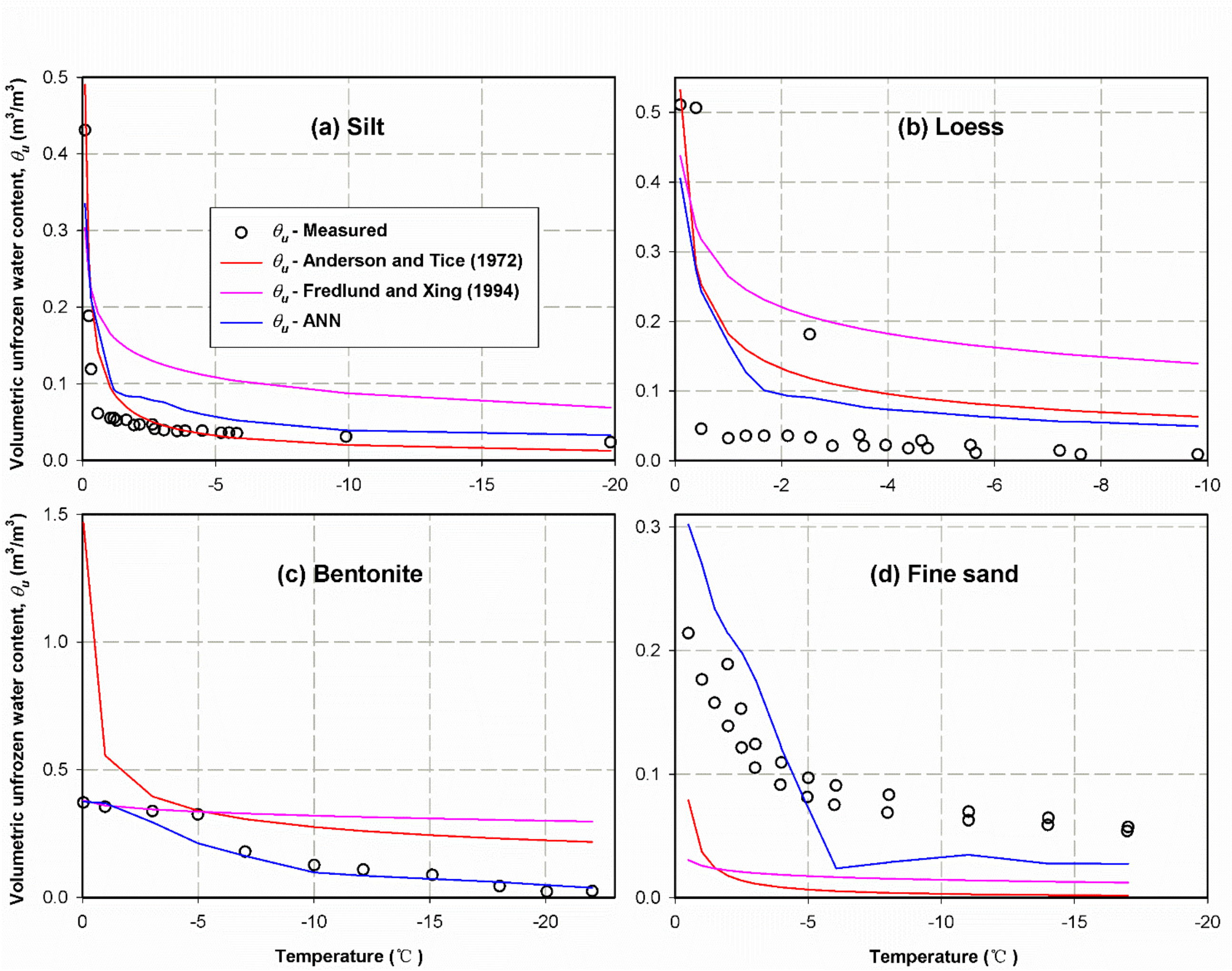


Fig. 10. Comparison between the ANN model and two traditional models

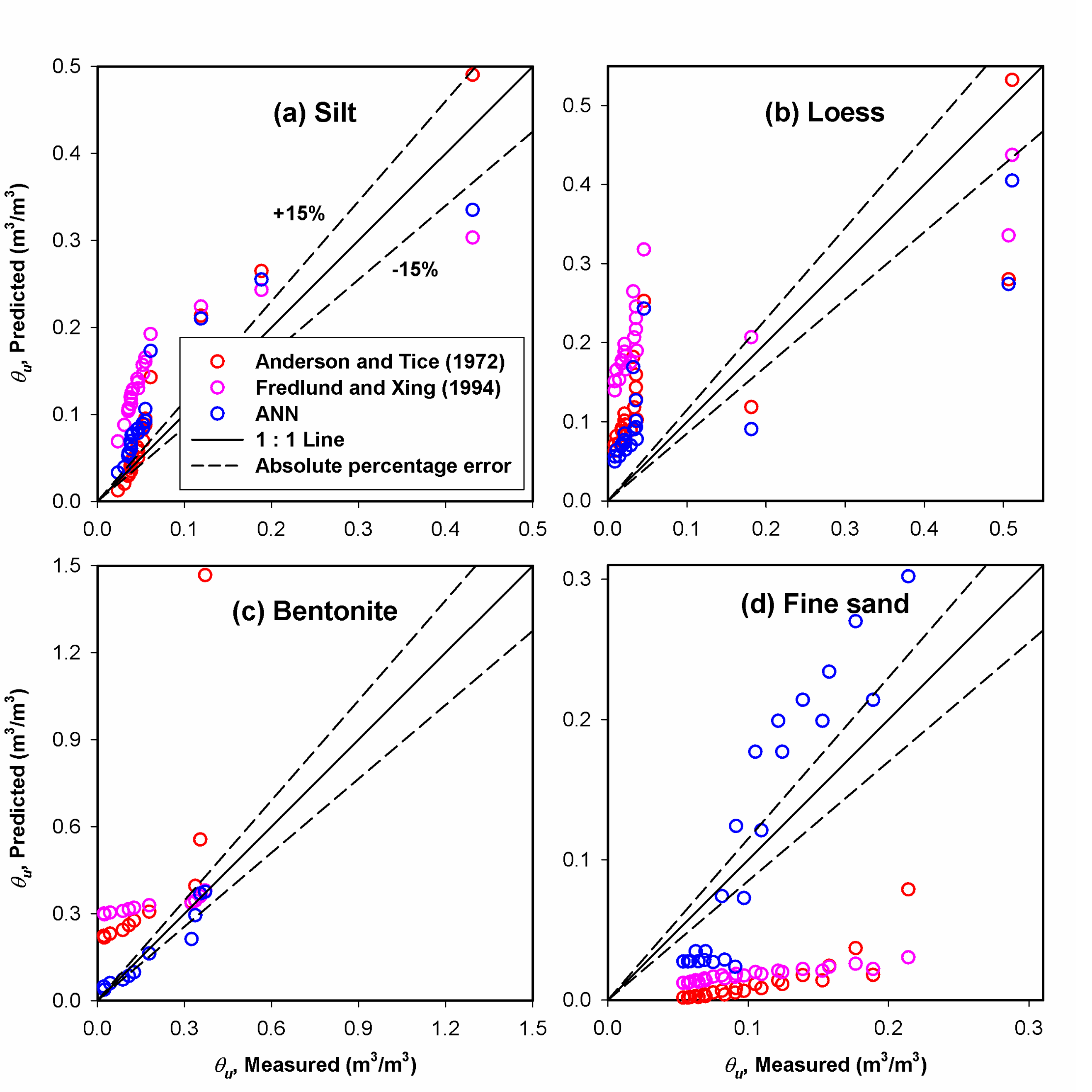


Fig. 11. Comparison of prediction accuracy of the ANN model and two traditional models

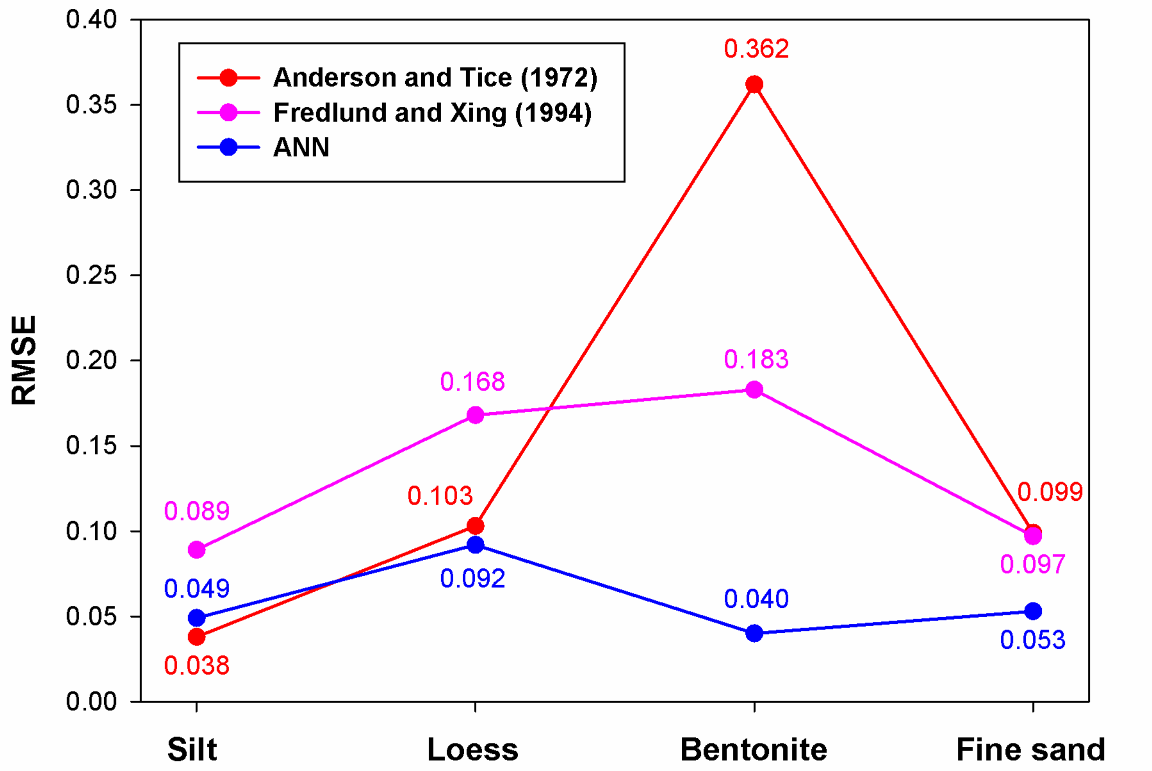


Fig. 12. The RMSE for the ANN model and two traditional models on four soils