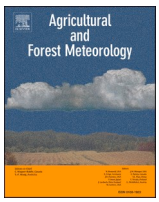
Agricultural and Forest Meteorology 323 (2022) 109080

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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



Short communication

Comparative analysis of seven machine learning algorithms and five empirical models to estimate soil thermal conductivity

Tianyue Zhao [a], Shuchao Liu [b], Jia Xu [a], Hailong He [c], Dong Wang [d], Robert Horton [e], Gang Liu [a,*]

[a] Department of Land Use Engineering, College of Land Science and Technology, China Agricultural University, Beijing 100193, China
[b] Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[c] College of Natural Resources and Environment, Northwest A&F University, Yangling, Shaanxi 712100, China
[d] School of Computer Science and Technology, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China
[e] Department of Agronomy, Iowa State University, Ames, IA 50011, USA

ARTICLE INFO

Keywords:
Soil thermal conductivity
Thermal conductivity model
Empirical model
Machine learning
Random forest
Feature importance ranking.

ABSTRACT

Soil thermal conductivity (λ) is an important thermal property that is crucial for surface energy balance and water balance studies. 1602 measured soil thermal conductivity values representing 189 soils were used to evaluate five empirical models (i.e., de Vries (1963) model (de Vries 1963), Campbell (1985) model (Campbell1985), Johansen (1975) model (Johansen 1975), Côté and Konrad (2005) model (Côté and Konrad 2005), and Lu et al. (2007) model (Lu 2007)) and seven machine learning (ML) algorithms (i.e., Decision Tree (DT), Random Forest (RF), Gradient Boosting Decision Tree (GBDT), Linear Regression (LR), K-Nearest Neighbors (KNN), Neural Network (NN), and Gaussian Process (GP)) to estimate λ. Our results demonstrated that the average root mean squared error (RMSE) values of ML were 66% and 82% of the empirical model values on validation and test sets respectively. The three best ML algorithms (GBDT, NN, RF) performed significantly better than the three best empirical models (Lu 2007, Côté and Konrad 2005, Johansen 1975): $0.183 < \text{RMSE} < 0.259$ (W m^{-1} K^{-1}) for ML algorithms and $0.293 < \text{RMSE} < 0.320$ (W m^{-1} K^{-1}) for empirical models. For ML, we recommend the GBDT, NN and RF algorithms. For empirical models, we recommend to use three normalized models (Lu 2007, Côté and Konrad 2005, Johansen 1975) over the physically-based model (DV1963) and the regression model (CG1985). The feature importance rankings performed by the RF and GBDT algorithms show that soil moisture content and soil bulk density are the most critical factors affecting λ. Soil moisture content and soil bulk density together account for more than 80% of the influence importance value of λ. RF gives more consistent feature importance ranking results than GBDT, therefore, we recommend the use of RF for selecting features.

1. Introduction

Soil thermal conductivity (λ) is an important thermal property (Côté and Konrad, 2005; Johansen, 1975; Lu et al., 2007). λ is necessary for quantitative descriptions of soil heat transfer, hydrothermal coupling, and other related heat and mass transfer processes (He et al., 2018, 2020a; Li and Shao, 2005; Peters-Lidard et al., 1998). Because rapid and accurate measurements of λ require specialized equipment and skill (He et al., 2020b), researchers have developed a number of empirical λ models (Côté and Konrad, 2005; Johansen, 1975; Lu et al., 2007). These models account for relationships between λ and other soil parameters such as texture, bulk density (ρ_b), water content (θ), and porosity (n) (He et al., 2020b). However, there is no widely recognized empirical model that is applicable to all soil types.

Machine learning (ML) is a rapidly growing data analysis method (Jordan and Mitchell, 2015). It has been applied in natural science studies (Butler et al., 2018). ML is a general term for a class of algorithms that find implicit patterns in large numbers of datasets for classification and prediction. This method has also been applied to agricultural soil analysis (Liakos et al., 2018), such as for soil classification (Padarian et al., 2020), soil moisture inversion (Senanayake et al., 2021), and estimation of soil bulk density ρ_b (Al-Shammary et al., 2018).

Abbreviations: DT, Decision Tree; GBDT, Gradient Boosting Decision Tree; GP, Gaussian Process; KNN, K-Nearest Neighbors; RF, Random Forest; LR, Linear Regression; ML, Machine learning; NN, Neural Network.

* Corresponding author.
E-mail address: liug@cau.edu.cn (G. Liu).

https://doi.org/10.1016/j.agrformet.2022.109080
Received 19 February 2022; Received in revised form 29 June 2022; Accepted 4 July 2022
Available online 13 July 2022

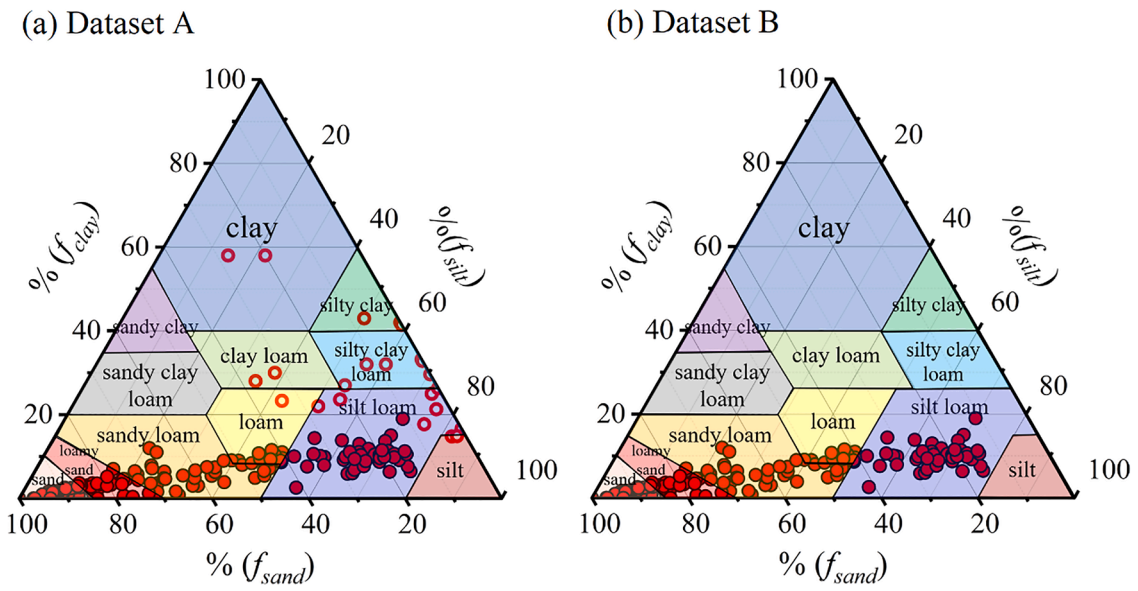


Fig. 1. Soil samples of two dataset classifications plotted based on the USDA soil texture classification triangle. (a) Dataset A: 189 soils with 1602 data points;(b) Dataset B: 164 soils with 1398 datapoints (By removing the 25 hollow red circle of (a)).

Recently, ML algorithms have been used to estimate λ (Liu et al., 2020; Rizvi et al., 2020; Yurttakal, 2021; Zhang et al., 2020b). However, ML methods have not been tested extensively and comprehensively. Generally, ML algorithms require large training datasets to avoid overfitting (Lever et al., 2016), however, the number of included data points has, thus far, been relatively small (40 < number of data points < 614, e.g. (Liu et al., 2020; Yurttakal, 2021; Zhang et al., 2020a; Zhang et al., 2020b. See Appendix for more information). The generalizability of the ML algorithm results needs to be further tested by increasing the number of data points. In addition, prior studies included only a single soil type or only included a few of the factors that affect λ. For example Rizvi et al. (2020) only estimated λ for sandy soils, while Cui et al. (2020) only considered dry density and θ, while the effects of porosity, quartz content and organic matter content were neglected. Moreover, most of the previous studies (Cui et al., 2020; Rizvi et al., 2020; Zhang et al., 2020b) did not rank the importance of the factors (ρ_b, n, soil moisture content (θ) and sand content (f_{sand}), silt content (f_{silt}), clay content (f_{clay}), organic matter content (f_{OM}), quartz content (f_q)) that affect λ. Furthermore, the vast majority of the earlier studies were based on Neural Network algorithms and their derivatives (Appendix and Liu et al., 2020; Zhang et al., 2020a; Zhang et al., 2020b), without considering other commonly used ML algorithms such as GBDT, RF, and GP. Therefore, there is a lack of comprehensive and extensive comparisons and evaluations of the performances of various ML algorithms.

In this study, we evaluated and compared the ability of seven mainstream ML algorithms (Decision Tree (DT), Random Forest (RF), Gradient Boosting Decision Tree (GBDT), Linear Regression (LR), K-Nearest Neighbors (KNN), Neural Network (NN), and Gaussian Process (GP)) to estimate λ. The ability of these ML algorithms was also compared with five popular empirical λ models. Finally, we used feature importance ranking to analyze the contributions of eight factors, including f_{sand}, f_{silt}, f_{clay}, f_{OM}, f_q, ρ_b, n and θ to estimate λ values.

2. Materials and methods

2.1. Database

The measured λ values used in this study were extracted from the following articles: Wang et al. (2020) (e.g., soil numbers 10–26, 28–31, 52–53, 57–62 of Table 1), Ghuman and Lal (1985), Ochsner et al. (2001), Bachmann et al. (2001), Lu et al. (2011), and Zhao et al. (2018). We collected a total of 1602 (189 soils) measured λ (θ) values for a range of water content θ values. Of the 1602 groups of λ (θ) data points, most were from China (1254) (See Appendix Fig. A1 for a spatial map of the sampling locations of the dataset) and Canada (282), and others were from Germany (24), the United States (20), Japan (18), and Nigeria (4). According to the spatial distribution of these data points, there is no noticeable spatial clustering. Because of the limited data source (1602 data points from 189 soils), to avoid overfitting issues related to small datasets (Lever et al., 2016), we did not split the dataset into a training dataset and a test dataset based on soil type.

All λ (θ) values were measured by transient methods (Table A-2 of the Appendix), including the traditional single-probe (SPHP) method and the dual-probe (DPHP) method. Bristow et al. (1994), Liu and Si (2011) and Kim and Oh (2020) report that SPHP and DPHP methods provide similar values of soil thermal conductivity (difference in results between the two methods is less than 10%). In this research, we include SPHP and DPHP values in the dataset. For the soil particle size analysis (and soil texture classification), either the pipette method or the laser particle size analyzer was used (Table A-2 of the Appendix). The 189 soil samples included silt loam (71 soils, 547 data points), loamy sand (31 soils, 291 data points), sandy loam (31 soils, 285 data points), sand (33 soils, 284 data points), loam (17 soils, 143 data points), clay loam (2 soils, 23 data points), clay (2 soils, 13 data points), silty clay (1 soil, 7 data points), and Peat (1 soil, 9 data points). Specific soil texture information according to USDA classification is presented in Fig 1.

In the following analysis, we used two different datasets. Dataset A includes the whole dataset (189 soils and 1602 data points), and Dataset B (164 soils with 1398 data points) is a subset of Dataset A. Dataset B does not include data for the finer-textured soils only represented by a few soil samples (Fig. 1). We need Dataset B, which only includes soils with limited clay content (f_{clay} < 20%), to test whether the performance of ML is improved or not when the sparse soil samples for a specific soil texture (in this case, we removed those soils with high clay content f_{clay}) was abandoned. It is well known that the volumetric content of quartz is a key parameter for modeling soil thermal conductivity (Tarnawski et al., 2012). However, due to expensive experimental apparatus (such as X-ray diffraction and X-ray fluorescence), reliable and accurate data of f_q for field soils is uncommon. Tarnawski et al. (2009) found weak correlations between f_q with the sum of f_{sand} and other soil materials. Therefore, if the original data sources did not provide f_q values, we assumed $f_q = 0.5\, f_{sand}$. This assumption was also used by Hu et al. (2017)

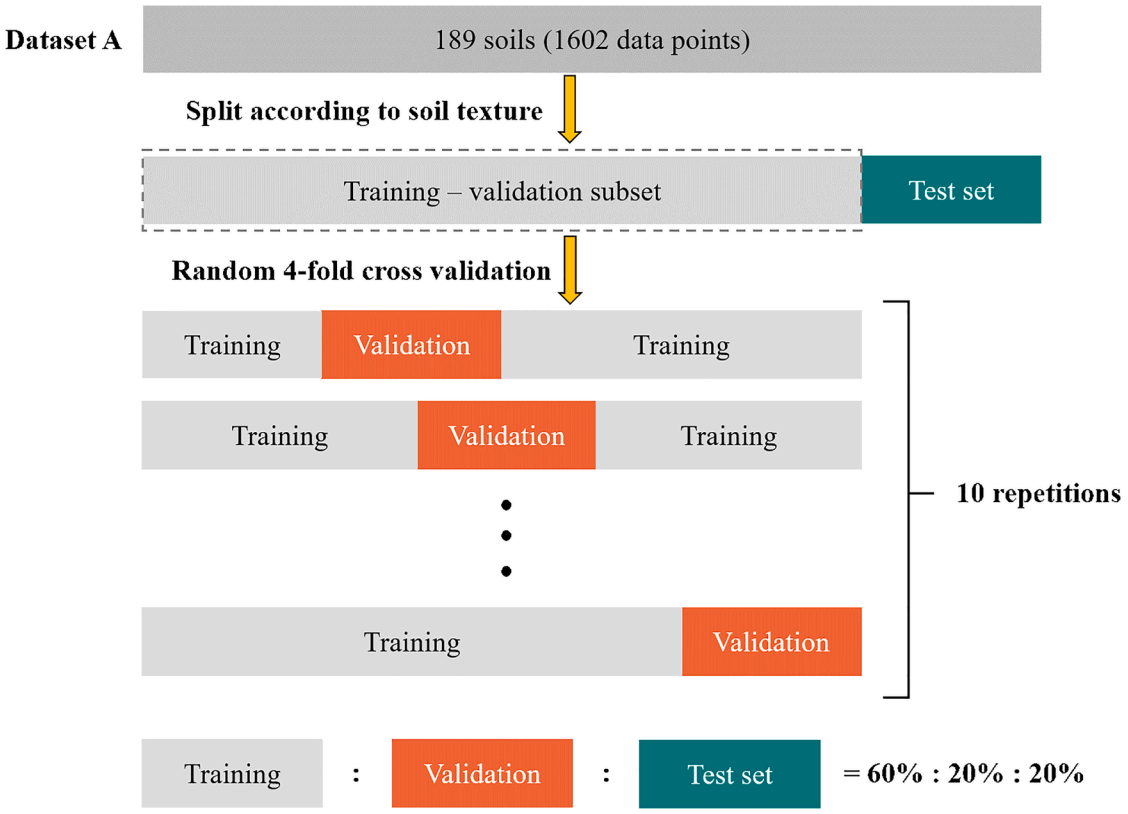


Fig. 2. Illustration of the split that we used in this study. According to soil texture classification, approximately 20% of the dataset A (1602 data points of 189 soils) are randomly selected as the test set, never to be used during training and validation. The training-validation subset is randomly divided into training and validation sets according to 75:25 split ratio (Here, we used random 4-fold cross-validation split with ten repetitions. For illustration purposes, we selected dataset A as an example).

and Zhao et al. (2018). Meanwhile, our ML predictions (results not presented here) indicate that RMSE difference between samples with $f_q = 0.5\, f_{sand}$ and $f_q = 1.0\, f_{sand}$, is less than 3%. The 1602 λ (θ) data points we obtained included the following parameter values: f_{sand}, f_{silt}, f_{clay}, θ, f_{OM}, ρ_b, n, f_q. When soil particle density ρ_s was not reported, but n and ρ_b were known, ρ_s was calculated from the relationship $n = 1\text{-}\rho_b/\rho_s$. When ρ_s was not known directly, we assumed $\rho_s = 2.65$ g cm^{-3}. We included both ρ_b and n in the analysis (ρ_b is strongly correlated with n and their correlation coefficient is 0.92, figure not shown here). For soil with a specified ρ_b value, n is determined by ρ_s, which is not a constant (Tolimir et al., 2020) but correlated to the soil mineral component (Birhanu et al., 2016).

To avoid spatial clustering (i.e., having the spatial location of the calibration data points close to the locations of test data points), we used a random data set split (Feinberg et al., 2018; Smirnov et al., 2020) based on the soil texture classification (Fig. 2). We randomly selected soils which contained approximately 20% of the data points in the total dataset (dataset A and dataset B) as the test set. In this way, the test dataset contained the entire λ (θ) data points of the selected soils. No partial λ (θ) data points of any of the 189 soils (or 164 soils for dataset B) existed in the test set. Therefore, all datapoints for the test set were unseen in the training-validation set, which insured that the test dataset was unbiased. As illustrated in Fig. 2, the training-validation/test split was accomplished by randomly selecting approximately 80% (based on the total datapoints of λ (θ)) of the soil data as the training-validation set, and the remaining 20% of the data (for dataset A: 1602 × 20% ≈ 320 data points) was selected as the test dataset. As a result, there were no data points in the test set that also belonged to the soils in the training-validation set. The random k-folds cross validation method (Meyer et al., 2019) with $k = 4$ was used to further split the training-validation set. A 75:25 split ratio (Wada et al., 2019) was used to randomly divide the data into the training set and the validation set (Feinberg et al., 2018; Smirnov et al., 2020). In this way, for both dataset A and dataset B, the size of the validation (80% × 25% = 20% of the full dataset) was approximately the same as the size of the test set (20% of the full dataset). Note that, unlike the traditional method, to avoid the drawback (higher error rate when k ≤ 8 for k-fold cross validation (Guo et al., 2019)) of the traditional 4-fold cross validation (Wada et al., 2019) (we performed traditional k-fold cross validation and the results showed that k = 4 resulted in a large variance for the predicted λ, which agreed with the findings of Bryce Meredig (2018). To overcome the shortcoming of a 4-fold cross validation, we used random splits (Feinberg et al., 2018; Smirnov et al., 2020) to split the training/validation subset (a total of 10 splits were performed for each given training/validation subset).

2.2. Five empirical models

There are at least 38 empirical thermal conductivity models of unfrozen soils (He et al., 2020b). Each empirical model is usually only applicable to specific soil types (e.g., sandy soils or peaty soils). None of them can be used to model all soil types (He et al., 2021). In this study, four widely evaluated and accepted empirical models (Campbell, 1985; Johansen, 1975; Côté and Konrad, 2005; Lu et al., 2007) and a mixing model (de Vries, 1963) are selected for comparison with selected ML algorithms.

de Vries (1963) is a physically based model which uses the weighted average of the thermal conductivity value of each soil component.

$$\lambda = \frac{\theta\lambda_w + k_a(n-\theta)\lambda_a + k_s(1-n)\left(\lambda_q^{f_q}\lambda_o^{1-f_q}\right)}{\theta + k_a(n-\theta) + k_s(1-n)} \quad (1)$$

In our calculation, we set air thermal conductivity $\lambda_a = 0.025$ W m^{-1} K^{-1} (Montgomery, 1947). λ_q (thermal conductivity of quartz) = 7.7 W m^{-1} K^{-1} (Clauser and Huenges, 1995), λ_o (thermal conductivity of other minerals) = 2.0 W m^{-1} K^{-1} ($f_q > 0.2$), $\lambda_o = 3.0$ W m^{-1} K^{-1} ($f_q < 0.2$), λ_w (thermal conductivity of water) = 0.57 W m^{-1} K^{-1} were set in this study. For dry soil, the de Vries (1963) model introduced a correction multiplier (1.25) into the Eq. (1). Farouki (1981) calculated the weighting factor k as

$$k_s = \frac{1}{3}\left[\frac{2}{1+0.125\left[\left(\lambda_q^{f_q}\lambda_o^{1-f_q}\right)/\lambda_w - 1\right]} + \frac{1}{1+0.75\left[\left(\lambda_q^{f_q}\lambda_o^{1-f_q}\right)/\lambda_w - 1\right]}\right] \tag{2}$$

$$k_a = \frac{1}{3}\left[\frac{2}{1+g_a\left[\left(\lambda_q^{f_q}\lambda_o^{1-f_q}\right)/\lambda_w - 1\right]} + \frac{1}{1+(1-2g_a)\left[\left(\lambda_q^{f_q}\lambda_o^{1-f_q}\right)/\lambda_w - 1\right]}\right] \tag{3}$$

where g_a is a shape factor

$$g_a = \begin{cases} 0.333 - (0.333 - 0.035)(n-\theta)/n & \text{for } 0.09 \le \theta \le n \\ 0.013 + 0.944\theta & \text{for } 0 \le \theta \le 0.09 \end{cases} \tag{4}$$

Campbell (1985) proposed the following empirical regression equation for λ

$$\lambda = A + B\theta - (A-D)\exp\left[-(C\theta)^E\right] \tag{5}$$

where $A = 0.65\text{-}0.78\rho_b + 0.60\rho_b{}^2$, $B = 1.06\rho_b\theta$, $C = 1 + 2.6\, f_{clay}{}^{-0.5}$, $D = 0.03 + 0.10\rho_b{}^2$, and $E = 4$.

Johansen (1975) presented a λ (θ) model for unsaturated soils which used a relationship between the Kersten coefficient (K_e) and λ.

$$\lambda = K_e\left(\lambda_{sat} - \lambda_{dry}\right) + \lambda_{dry} \tag{6}$$

$$K_e = \begin{cases} 0.7\lg S_r + 1 & S_r > 0.05 \quad \text{coarse} \\ \lg S_r + 1 & S_r > 0.1 \quad \text{fine} \end{cases} \tag{7}$$

Here, for unfrozen soil, fine soil and coarse soil is defined as $f_{clay} >$ 5% and $f_{clay} \le$ 5%, respectively (He et al., 2021). K_e is related to soil saturation S_r ($S_r = \theta/n$), where λ_{sat} and λ_{dry} are the thermal conductivity values of saturated and dry soil, respectively.

$$\lambda_{sat} = \lambda_w^n\left(\lambda_q^{f_q}\lambda_o^{1-f_q}\right)^{1-n} \tag{8}$$

$$\lambda_{dry} = \frac{0.135\rho_b + 64.7}{2700 - 0.947\rho_b} \tag{9}$$

Côté and Konrad (2005) improved the Johansen (1975) model by introducing a parameter k that reflects soil texture, and the relationship between K_e and S_r is

$$K_e = k \cdot S_r / [1 + (k-1)S_r] \tag{10}$$

For gravels (and coarse sands), medium sand (and fine sands), silty soils and peaty soils, k is 4.60, 3.55, 1.90 and 0.60, respectively. Based on the values of f_{sand} and f_{clay}, He et al. (2021) gave detailed standards to distinguish various soils according to their texture (fine, medium and coarse; See Table 3 of He et al. (2021) for more details). In this study we followed these standards. Because the complete mineral compositions of the majority of our soil are unknown, λ_{sat} is calculated according to the Johansen (1975) model (Eq. (8)), but λ_{dry} is computed by

$$\lambda_{dry} = \chi 10^{-\eta n} \tag{11}$$

where χ and η are parameters related to soil texture as described by Côté and Konrad (2005). For natural soils, χ and η is 0.75 and 1.20, respectively.

In order to extend the λ (θ) model to small water content conditions, Lu et al. (2007) modified the K_e expression to

$$K_e = \exp\left\{\alpha\left[1 - S_r^{(\alpha-1.33)}\right]\right\} \tag{12}$$

where α is a soil texture parameter, which is 0.96 and 0.27 for coarse-textured and fine-textured soils respectively. For the Lu (2007) model, λ_{sat} is calculated using the Johansen (1975) model (Eq. (8)), and λ_{dry} is assumed to be linearly related to porosity (n) (Lu et al. (2007).

$$\lambda_{dry} = -0.56n + 0.51 \tag{13}$$

Johansen (1975), Côté and Konrad (2005) and Lu (2007) models were called normalized soil thermal conductivity models (He et al., 2020a). After evaluating the performance of 38 normalized empirical models on a dataset which contained 669 data points (71 soils), He et al. (2020b) recommended the use of Johansen (1975), Côté and Konrad (2005), Lu (2007) model.

2.3. Seven ML algorithms

We include the following ML algorithms: Decision Tree (DT), Gradient Boosting Decision Tree (GBDT), Neural Network (NN), Random Forest (RF), Gaussian Process (GP), K-Nearest Neighbors (KNN), and Linear Regression (LR). Detailed descriptions of these algorithms can be found in Aurélien (2017). DT is based on a tree-like structure for making decisions, and has the advantage of being very interpretable. RF is a type of DT in which each tree has a release to extract a portion of the training set. It can be used for classification and regression problems and can handle high-dimensional data. GBDT is a boosting algorithm based DT, which first gives an estimation of the target value, and then uses another "tree" to estimate the error between the estimated value and the true value. This method obtains the final estimated value by continuously reducing the error, but the training is time consuming. LR refers to the use of a linear function to estimate the relationship between variables and target values. This algorithm can give regression formulas, a feature that significantly distinguishes it from other machine learning algorithms. The principle of KNN is to calculate the distance of a given value from all other samples and rank them. The disadvantage of this algorithm is that the computation is time consuming when the amount of data is large. NN in this study refers to artificial neural network (ANN), which is a supervised learning algorithm. NN is mainly divided into input layer, hidden layer, and output layer. The principle is to find the weights of the connections between neurons and the threshold values of the neurons. Because the initial values of the neural network are given randomly, a large amount of data is needed to improve the estimation accuracy. Statistics and Bayesian theory help GP to be suitable for processing high-dimensional nonlinear small samples. All ML algorithms and calculations were performed with Wolfram Mathematica 12.1 (Wolfram Research, Inc., Champaign, IL). The "L2Regularization" (Smirnov et al., 2020) of "LogisticRegression", as well as "FeatureFraction" (Quezada, 2019), "LeafSize" (Song and Lu, 2015) and "DistributionSmoothing" (Quezada, 2019) were used to prevent overfitting of data.

2.4. Model performance assessment

In this study, root mean square error (RMSE), Nash–Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) and average deviations (AD), were selected to evaluate the estimation accuracy of both ML algorithms and empirical thermal conductivity models.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{m}(Y_i - \widehat{Y}_i)^2}{m}} \tag{14}$$

$$NSE = 1 - \frac{\sum_{i=1}^{m}(Y_i - \widehat{Y}_i)^2}{\sum_{i=1}^{m}(Y_i - \overline{Y})^2} \tag{15}$$

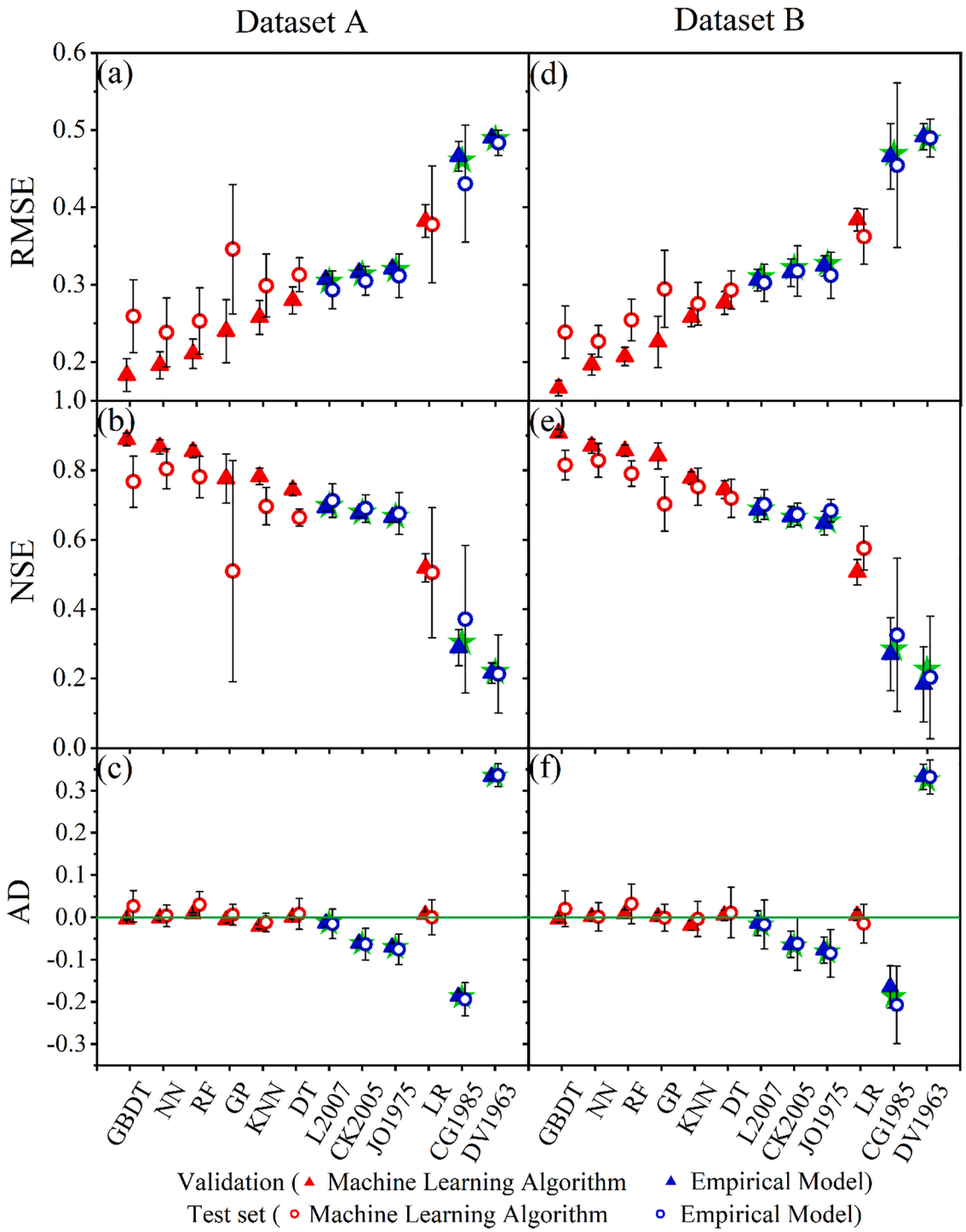


Fig. 3. Performance of seven ML algorithms with five empirical models. (a), (b), (c): 189 different soils (1602 data points, see Fig 1(a) for classification of soils according to texture); (d), (e), (f): 164 different soils (1398 data points, see Fig 1(b) for classification of soils according to texture). (Note: The green pentagram represents the result of using the full dataset or all datapoints (dataset A: 1602 data points, dataset B: 1398 data points)).

$$AD = \frac{\sum_{i=1}^{m}(\widehat{Y}_i - Y_i)}{m} \tag{16}$$

where Y_i and $\widehat{Y}_i$ represent measured and predicted values of λ, respectively. $\overline{Y}$ is the mean value of Y_i and m represents the total number of data points. In our analysis, we found that mean absolute error (MAE) was strongly correlated with RMSE (Pearsons' correlation coefficient is $r \approx 0.94$). Therefore, we decide not to use MAE. In this study, analyses of variance and correlation analyses were performed using SPSS software (version 21.0, SPSS Inc., Chicago, IL). The comparisons of different ML algorithms, as well as empirical models were performed by using one-way ANOVA, followed by a least significant difference test ($P < 0.05$).

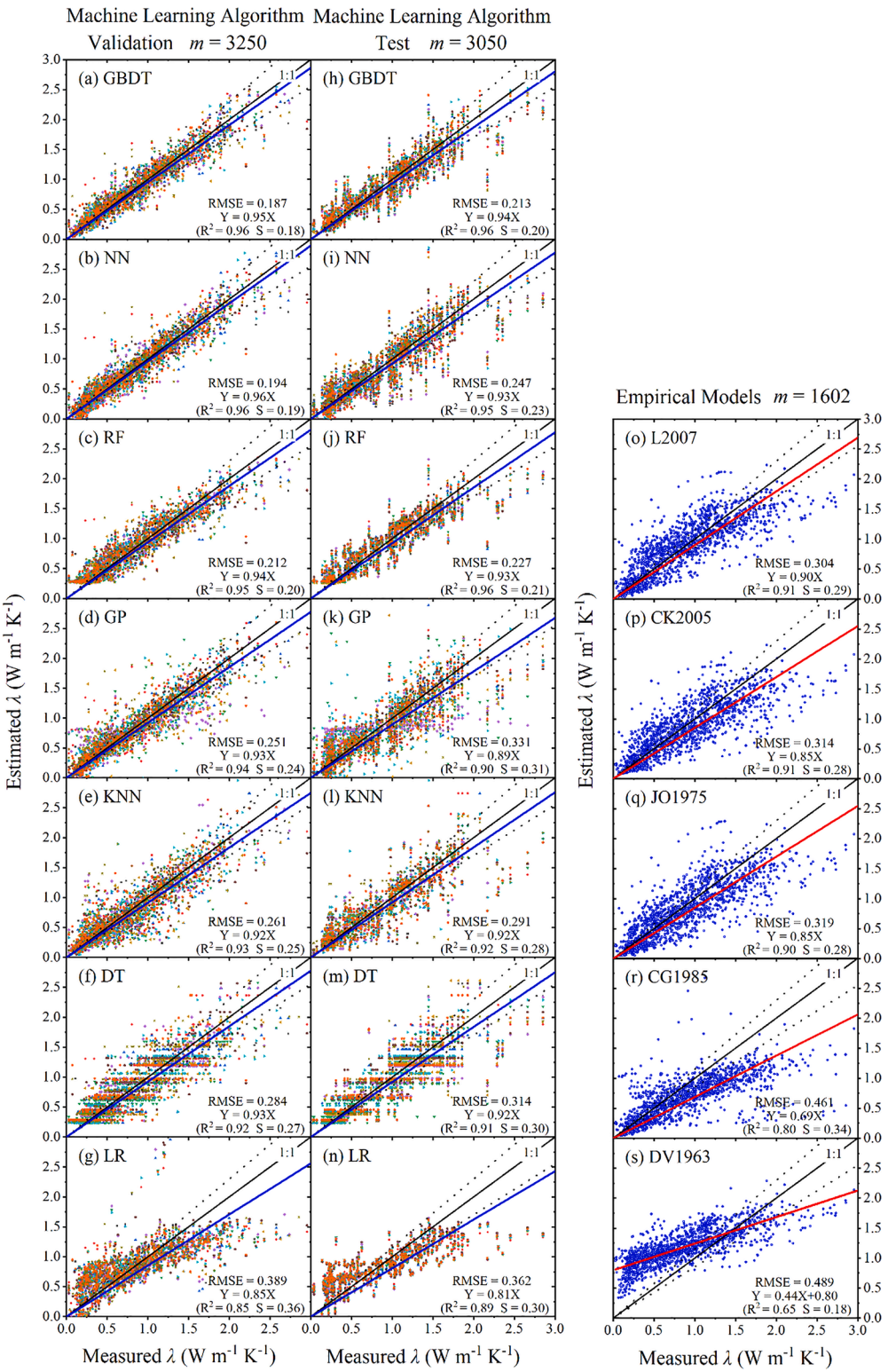


Fig. 4. Performance of the ML algorithms and the empirical models in dataset A. Lines (red and blue) are the best fitted lines with linear regression. R^2 and standard error (S) are the coefficient of determination and standard error of each linear regression. Abbreviation: Gradient Boosting Decision Tree (GBDT), Neural Network (NN), Random Forest (RF), Gaussian Process (GP), Nearest Neighbors (KNN), Decision Tree (DT), Linear Regression (LR), Lu et al. (2007) model (L2007), Côté and Konrad (2005) model (CK2005), Johansen (1975) model (JO1975), Campbell (1985) model (CG1985), and de Vires (1963) model (DV1963). Fig. 4a–g and Fig. 4h–g are the results of 10 repetitions for random sampling of the validation set and the test set, respectively. For both Fig. 4a–g and Fig. 4h–n, there are roughly 3200 data points (1602 × 20% × 10 ≈ 3200). The different color points indicate the results of repetitions of the ML algorithms in Fig. 4a–n. For the five empirical models, the full dataset (all 1602 data points) were included in the plot and RMSEs calculation. Here *m* represents the total number of data points. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

3. Results and discussion

3.1. Empirical models versus machine learning algorithms

Fig. 3a, b, d, and e indicate that among all of the ML algorithms, GBDT, NN, and RF perform the best. For dataset A (Fig. 3a–c), on the validation dataset, GBDT > NN > RF, (0.183 W m^{-1} K^{-1} < RMSE < 0.210 W m^{-1} K^{-1}, 0.854 < NSE < 0.889, -0.005 W m^{-1} K^{-1} < AD < 0.008 W m^{-1} K^{-1}). While, for the test dataset, NN > RF > GBDT, (0.238 W m^{-1} K^{-1} < RMSE < 0.259 W m^{-1} K^{-1}, 0.767< NSE < 0.804, 0.003 W m^{-1} K^{-1} < AD < 0.030 W m^{-1} K^{-1}). The significance test results showed that RF, which ranked third, was significantly different from GBDT, NN in the validation dataset, while there was no significant difference among these three algorithms in the test set. We ranked the seven ML algorithms based on their performances for dataset A (both the validation and test set): GBDT or NN > RF > KNN > (DT/GP/LR). Overall,

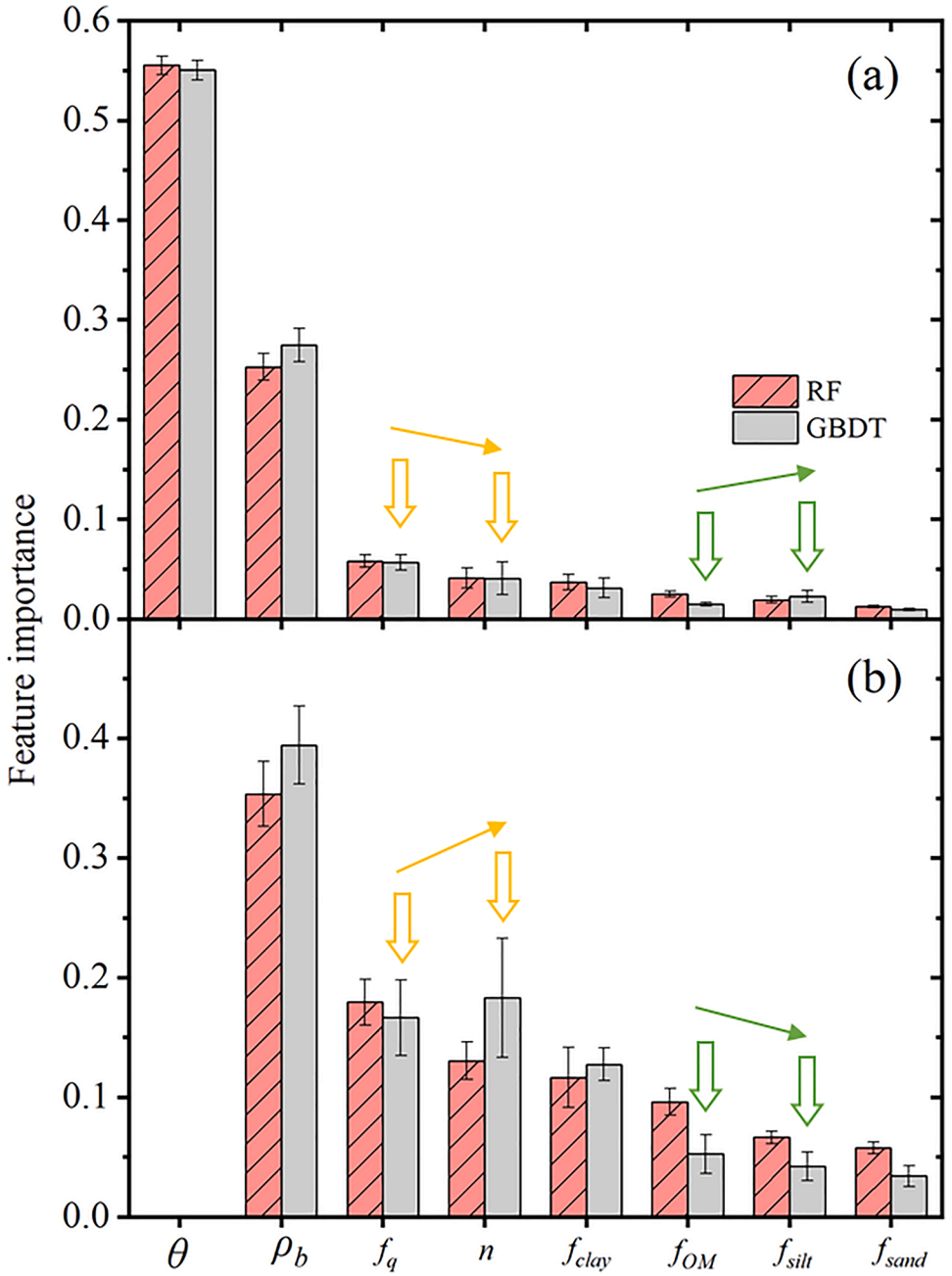


Fig. 5. Feature importance for sand content (f_{sand}), silt content (f_{silt}), clay content (f_{clay}), organic matter content (f_{OM}), quartz content (f_q), soil bulk density (ρ_b), soil porosity (n), and soil water content (θ). Eight factors are sorted in (a) and seven factors (θ was excluded) are sorted in (b). Error bars of each column represent standard deviation.

Fig. 3a, b, d, and e show that our ML ranking is similar to other reports. For example, NN outperformed RF (Ahmad et al., 2017), KNN (Alireza et al., 2010) and DT (Li et al., 2019); RF outperformed KNN (Ren et al., 2020), GP (Hultquist et al., 2014; Ly et al., 2021); GBDT outperformed DT (Lv et al., 2021). For saturated hydraulic conductivity predictions, Fig. 3 of Araya and Ghezzehei (2019) illustrated that the overall performance is BRT (similar to GBDT) > RF > KNN, which agrees with our results of Fig. 4. We expect a similar ranking of the performance of these seven ML algorithms when they were used in other branch of soil physics. However, to our knowledge, few publications exist to test this statement.

We selected three indicators (RMSE, NSE and AD) to analyze the performance of each model (Fig. 3), and the ranking given by these three indicators was quite consistent. RMSE and NSE values in Fig. 3 are highly correlated with the Pearsons' correlation coefficient ($r > 0.98$). AD did not perform as well as RMSE and NSE, e.g., it is difficult to see the ranking of GBDT, NN, RF, KNN, DT, GP in Fig. 3c and f. However, Fig. 3 clearly demonstrated that ML algorithms performed significantly better than the empirical models. As to the empirical models, Campbell (1985) and de Vries (1963) were significantly different from the other empirical models: the AD of the other models ranged from -0.085 W m^{-1} K^{-1} to 0.032 W m^{-1} K^{-1} except for the Campbell (1985) and de Vries (1963) (AD of Campbell (1985) < -0.165 W m^{-1} K^{-1} and the AD of de Vries (1963) > 0.336 W m^{-1} K^{-1}).

Similar to the conclusion of Li et al. (2017) and Jahan et al. (2021), Fig. 3a, b, d, and e also show that DT and LR perform the worst. The linear assumption of LR was not suitable to capture the nonlinear λ (θ) relationship, and thus, performed poorly. GBDT, which was derived

Table A1
Summary of the dataset and methods used in this research and in other publications.

	Data points	Soil texture	ML method	Feature importance ranking
this research	1602	sand, loam, clay, peat	RF, LR, DT, GBDT, KNN, NN, GP	Yes
Liu et al. (2020)	40	clay	BPNN, GA-BPNN	No
Zhang et al. (2020a)	257	sand, loam, clay	ANN	No
Zhang et al. (2020b)	614	sand, clay	ANN	No
Rizvi et al. (2020)	80	sand	GMDH	No
Cui et al. (2020)	609	sand, clay, boulder, breccia, weathered rock	SVR	No
Yurttakal (2021)	257	sand, loam, clay	XGBoost, LightGBM	Yes

Abbreviations: ML, Machine learning; DT, Decision Tree; RF, Random Forest; GBDT, Gradient Boosting Decision Tree; XGBoost, Extreme Gradient Boosting Decision Tree; LightGBM, Light Gradient Boosting Decision Tree; LR, Linear Regression; K-Nearest Neighbors; GP, Gaussian Process; KNN, NN, Neural Network; ANN, Artificial Neural Network; BPNN, Back Propagation Neural Network; GA-BPNN, Back Propagation Neural Network optimized by genetic algorithm; SVR, Support Vector Regression; GMDH, Group Method of Data Handling.

directly from the classical DT, significantly outperformed DT (Lv et al., 2021). We also found that, unlike other ML methods, the results of the GP method had a large variance (Fig. 3a and b), especially for dataset A. The reason might be that the GP method assumed that the data points had a Gaussian normal distribution (Hultquist et al., 2014), and thus, for cases when the dataset was more discrete or only had a few data points, the accuracy of the GP prediction decreased.

Fig. 3 also indicates that ML algorithms generally outperform the empirical models for both the validation and test cases. For dataset A (Fig. 3a–c), comparing the three best performing ML algorithms (GBDT, NN, RF) with the three empirical models Lu (2007), Côté and Konrad (2005), Johansen (1975), we find that for the validation set, the average RMSE of these three ML algorithms was 62.4% of the average RMSE of the three empirical models. For the test set, the prediction accuracy of the three ML algorithms decreased slightly, and the average RMSE value was 82.5% of the RMSE values for the three empirical models. In the validation set, when the best performing empirical model ((Lu, 2007), RMSE = 0.306 W m^{-1} K^{-1} in the validation, RMSE = 0.293 W m^{-1} K^{-1} in the test set) was selected for comparison, the RMSEs of GBDT, NN and RF were reduced by 31.4 to 40.4%, and NSE was improved by 23.2 to 28.3%; in the test set, the RMSE of GBDT, NN and RF decreased by 11.6 to 18.8%, and NSE improved by 7.7 to 12.9%. ML algorithms always performed better on the validation set compared to the test set. One possible explanation was that the soils of the test set (and thus the λ (θ) curve) were absolutely independent from the soils of the training set, and the soils of these two sets had zero common shared elements (intersection was zero). On the contrary, the soils of the validation set could overlap with the soils in the training set. In other words, the validation set contained a partial portion of the specific λ (θ) curve of a given soil that had already been used in the training set. Therefore, to ensure an unbiased prediction for the ML training, a test that was completely independent from the training set (λ (θ) data sets belonging to different textured soils) should be used. Although our results demonstrated that ML algorithms outperform the empirical models, this is not surprising, because these empirical models used few soils (8 soils for Lu (2007) as an example) for calibrating various parameters. On the contrary, our ML algorithms used at least 90 soils (for Dataset B: 60% of 164 soils is roughly 98 soils) for training.

The results for dataset A (Fig. 3a, b, c) and dataset B (Fig. 3d, e, f) showed that the ML algorithms based on dataset B (removing the data of soils that were sparsely distributed in the soil texture map: 25 soils and 204 data points of dataset A) had higher accuracy and reduced fluctuation (especially for GP). This agrees with our intuition, by limiting the types of soils of dataset B only to soils (for a specific texture) with sufficient samples, we can improve the prediction accuracy of ML algorithms. Switching from dataset A to dataset B, we noticed that, for both RMSE and NSE, the number of ML algorithms that outperformed the empirical models increased from three (GBDT, NN, RF) to six (GBDT, NN, RF, GP, KNN, DT). Comparing Fig. 3a and b and Fig. 3d and e, we found that the variance of GP on dataset B decreased significantly (40.5% for the RMSE indicators (from 0.084 to 0.050) and 75.5% for the NSE indicators (from 0.318 to 0.078)), which indicated that the GP algorithm was more sensitive to the spatial aggregation, or that the GP prediction accuracy was improved by removing the 25 soil samples (204 data points) that were more discrete in the spatial distribution map of soil texture (Fig. 1a). We concluded that although dataset A contained more soil textures than dataset B, the distribution of λ (θ) data points for each texture classification was not uniform, and sparsely distributed soils could reduce the ML prediction accuracy. In the future, we can systematically extend the soil thermal conductivity dataset of each specific texture, to avoid sparse datasets (in the soil texture classification map, e.g., Fig. 1a in the red circle part of the data volume) caused errors, and to further improve the ML prediction accuracy.

Fig. 4 shows the prediction of the ML algorithms for dataset A versus

Table A2
References for the soil texture classification and the soil thermal property measurements of 189 soils (1602 data points) that we used.

Literature source	Data points	No. of soil	Country	Soil texture	Method of soil particle-size analysis	Method of λ measurement
Bachmann et al. (2001)	24	1	Germany	Loamy sand	Null	DPHP
Chen et al. (2012)	484	42	China	Silt loam, Sandy loam, Sand, Loamy sand	Laser Particle Size Analyzer	SPHP
Ghuman and Lal (1985)	4	1	Nigeria	Clay	Null	SPHP
Kasubuchi et al. (2007)	18	2	Japan	Loam, Clay	Null	DPHP
Lu et al. (2007)	118	10	China, US	Loam, Silt loam, Sandy loam, Clay loam, Sand	Pipette method	Thermo-TDR
Lu et al. (2011)	6	1	China	Silty clay	Pipette method	Thermo-TDR
Ochsner et al. (2001)	49	4	China	Silt loam, Sandy loam, Clay loam	Pipette method	Thermo-TDR
Tarnawski et al. (2015)	234	39	Canada	Loam, Silt loam, Sandy loam, Sand, Loamy sand	Laser Particle Size Analyzer	Thermal-conductivity probe
Zhao et al. (2018)	617	85	China	Loam, Silty loam, Sandy loam, Sand, Loamy sand	Laser Particle Size Analyzer	DPHP
Zhao et al. (2019)	48	4	Canada	Sand, Loamy sand	Null	DPHP

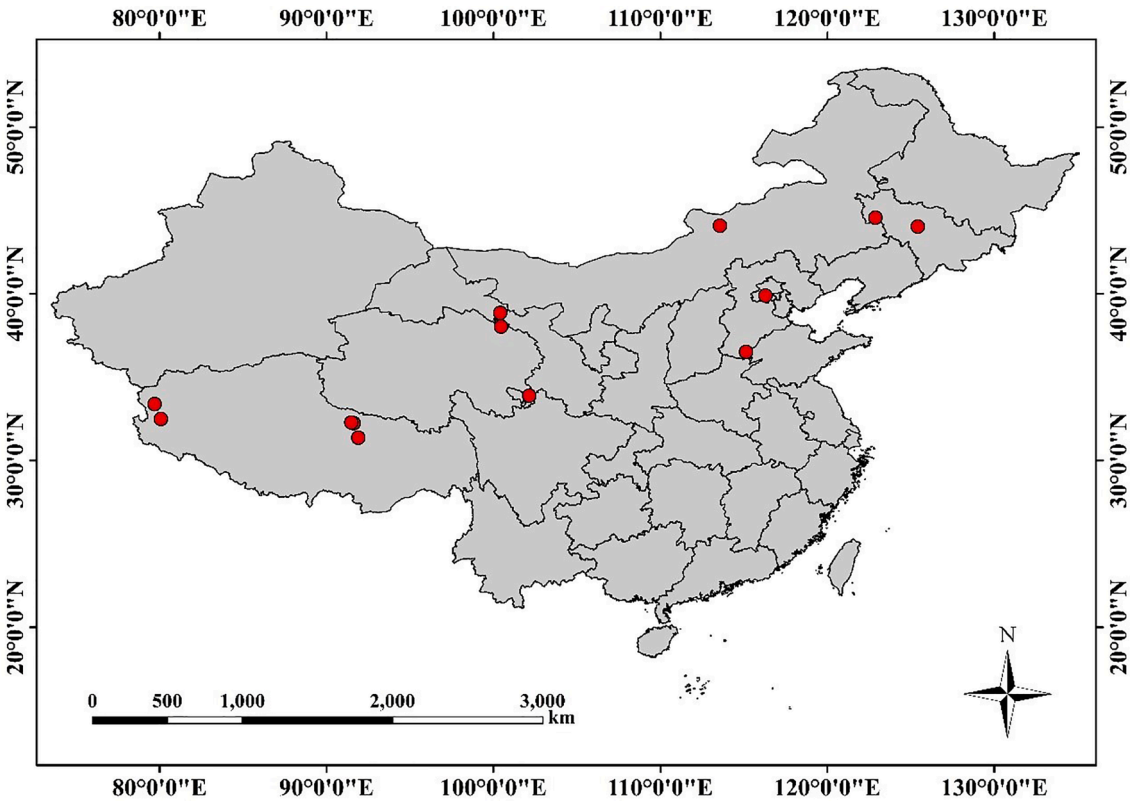


Fig. A1. Map of the Chinese sampling locations (140 soils, 1254 data points). In this study, we collected a total of 1602 (189 soils) values of λ (θ). Of the 1602 data points, the majority were from China (1254) and Canada (282).

the empirical models. Consistent with the results in Fig. 3, ML algorithms generally outperform the empirical models. The three best performing ML algorithms (GBDT, NN, RF) correspond to scatter plots with the least data dispersion (Fig. 4a, b, c, h, i, j), however, the predictions of the empirical models (1602 data points) are loosely distributed around the 1:1 line. The standard errors of ML algorithms are significantly smaller than the empirical models (Fig. 4o–s)). The standard errors of the three best performing ML algorithms are between 0.18 and 0.23, while the standard errors of the best three empirical models (Lu (2007), Côté and Konrad (2005), Johansen (1975)) are distributed between 0.28 and 0.29. The three empirical models mentioned above have at least 21.7% higher standard errors compared to the three best performing ML algorithms. Similar to the results displayed in Fig. 3, the scatter plot in Fig. 4 vividly demonstrates that the accuracy of the validation set is higher than that of the test set. However, an independent test set (and cross validation), other than the validation set should be used to further evaluate ML algorithms to obtain unbiased results.

Fig. 3 shows that the ranking of the three best performaning empirical models is: Lu (2007), Côté and Konrad (2005) and Johansen (1975). For dataset A (Fig. 3a–c), the mean values of RMSE, NSE and AD for the three models on the validation set are 0.314 W m^{-1} K^{-1}, 0.677, and -0.049 W m^{-1} K^{-1}, respectively. On the test set, the mean values of RMSE, NSE, and AD are 0.303 W m^{-1} K^{-1}, 0.692, -0.052 W m^{-1} K^{-1}, respectively. We note that for our dataset, the RMSE of Lu (2007) is roughly twice the value reported by Lu et al. (2007). Although this is one of the best empirical models, it failed to achieve the accuracy reported by Lu et al. (2007), when the soil sample set is large and contains a variety of soil textures. This result is consistent with He et al. (2020a), who evaluated 24 empirical models on 16 soils (439 data points), and found that none of the empirical thermal conductivity models applied well to the entire range of soils. In addition, our results (Figs. 3 and 4) further show that even the best performing empirical models (Lu et al. (2007), Côté and Konrad (2005), Johansen (1975) do not perform as well as the best ML algorithms (GBDT, NN, RF). Specifically, for the validation set, the average RMSE of the three ML algorithms is 60.2% of the RMSE of the three empirical models. For the test set, the average RMSE of the three ML algorithms is 77.1% of the RMSE of the three empirical models. Except for LR, all of the ML algorithms outperform the empirical models. However, most recent studies on soil thermal conductivity with ML algorithms do not consider cross validation (Cui et al., 2020; Rizvi et al., 2020; Zhang et al., 2020a).

The prediction accuracy of all 5 empirical models were ranked as Lu (2007) > Côté and Konrad (2005) > Johansen (1975) > Campbell (1985) > de Vries (1963). Among them, Campbell (1985) and de Vries (1963) performed the worst (with RMSE > 0.43 W m^{-1} K^{-1} and NSE < 0.37). This agrees with the results of both He et al. (2020a) and Liu et al. (2021). The prediction accuracy of the normalized empirical models (Lu (2007), Côté and Konrad (2005) and Johansen (1975)) was higher than those of the semi-physical model (de Vries (1963)) and the regression model (Campbell (1985)). The significance test based on RMSE, R^2 and NSE revealed that (except for the NSE of Johansen (1975) on the validation set which was significantly different from Lu (2007)) there was no significant difference among the three models (Lu (2007), Côté and Konrad (2005) and Johansen (1975)). Tong et al. (2016); Yan et al. (2019); Zhang et al. (2018); Zhao et al. (2019) also found that although Lu (2007) and Côté and Konrad (2005) redefined the K_e~S_r and λ_{dry} relationship based on the Johansen (1975) model, the performance of these three models were similar. However, we have to emphasize that Lu (2007) was proposed to improve the performance of Côté and Konrad (2005) at low water content, especially on fine-textured soils. If we compare these three models at low S_r, we might find superiority of Lu (2007) over both Côté and Konrad (2005) and Johansen (1975). The poor predictions of the Campbell (1985) model were related to the fact that the model needed to be calibrated for different soils to obtain the model parameters. The Campbell (1985) model was prone to underestimation (Wang et al. (2012) when parameters were not calibrated. This

was consistent with our results (Fig. 4r: Campbell (1985) model underestimated thermal conductivity by 30%). Another drawback was that the Campbell (1985) model was not suitable for soils with zero clay content (Campbell, 1985) . The de Vries (1963) model was the worst performer, compared to all of the ML algorithms as well as the other empirical models. The poor performance of the de Vries (1963) model might originate from the underlying model assumptions. To calculate the temperature gradient ratios, soil solid particles were assumed to be ellipsoidal (de Vries, 1963 pp. 214), whereas the actual soil particles were not always ellipsoidal (Tang and Li, 2018). Based on the above results and analysis, for empirical model, we do not recommend the use of the Campbell (1985) and de Vries (1963) models. This statement needs to be further validated in the future for datasets which include a wide range of soil textures and data points.

3.2. Feature importance

It is important to quantify the influence of major parameters or factors (θ, f_{sand}, f_{silt}, f_{clay}, f_{OM}, f_q, ρ_b, n) on the soil thermal conductivity. In ML, feature importance ranking can measure contributions of each input feature (factor) to the performance of a ML model. This method has become a powerful tool for facilitating understanding of a learning system and discovery of key factors in a specific domain. By ranking the importance scores, we used feature importance (Casalicchio et al., 2018) to analyze the importance of different features and identified the most important potential feature of the model. In this study, we used two feature importance ranking algorithms, the RF algorithm (Albert and Linville, 2020) and the GBDT algorithm (Weng et al., 2019). We ranked the characteristic importance of eight parameters (θ, f_{sand}, f_{silt}, f_{clay}, f_{OM}, f_q, ρ_b, n). As shown in Fig. 5, the importance rankings given by the two algorithms are basically the same. Both RF and GBDT show that θ and ρ_b are the two most important features, and their combined importance value exeeds 80%. In particular, the importance value of θ exceeds the sum of the importance values of the other factors, and ρ_b is the second most important feature as expected. Significance tests show that the feature importance of θ is significantly higher than that of ρ_b for both RF and GBDT. Meanwhile, both θ and ρ_b are significantly higher than the other factors importance. f_q is the third most important feature, which agrees with the conclusion of He et al. (2020a), therefore, it is necessary to consider the effect of quartz content on λ prediction.

The feature importance analysis is valuable for empirical model development. By selecting the most important features and omitting those with low feature importance scores, more robust and concise empirical models can emerge with prediciton accuracy comparable to the ML algorithms.

Fig. 5(a) and (b) differed in the number of features involved in the ranking. Fig. 5(a) included 8 features, while θ was removed in Fig. 5(b) in order to test whether the ranking of the remaining 7 features was sentitive or not to the presence of θ. Although the results of the GBDT and RF algorithms are similar (Fig. 5), the consistency of the results of the RF algorithm is better. The order of importance values given by the RF method in Fig. 5(a) is completely consistent with that of Fig. 5(b) ($\rho_b > f_q > n > f_{OM} > f_{clay} > f_{silt} > f_{sand}$). The following mechamism can explain why the bulk density and quartz content are more important than other factors. Higher bulk density leads to the decrease in porosity and more solid contacts. As thermal conductivity of soil solid is higher than that of water and air, thus more solid contact also leads to an increase in λ. Thermal conductivity of quartz is much higher than other minerals thus dominant in determining λ (Campbell, 1985). The big differences on the importance of ρ_b and n could be related to the violation of the commonly used constant soil particle density (2.65 g cm^{-3}) assumption. According to the research of Flint and Flint (2002) and Keller and Håkansson (2010), there exist significant soil particle density variations (Tolimir et al., 2020) especially for soil samples across the continental scale, just like the samples we used. There are two inconsistencies (or contradictions) in GBDT (yellow arrows in Fig. 5a) indicating that $f_q > n$, which contradicts the results in Fig. 5(b) ($f_q < n$); green arrows of Fig. 5(a) indicate that $f_{OM} < f_{silt,}$ which contradicts with the results in Fig. 5(b) ($f_{OM} > f_{silt}$).

Our results presented above demonstrate that feature importance ranking using the RF algorithm is more reliable than using the GBDT algorithm. Therefore, we recommend using RF for future feature importance ranking. Our GBDT results differ from the GBDT results of Yurttakal (2021) (which contain 257 λ values), who report the top three parameters as f_{clay}, f_q and S_r. The differing results may be related to different sizes of datasets analyzed, for example our 1602 values versus their 257 values. Another possible explaination is that their analysis did not consider f_{silt} and f_{OM}, and they used S_r instead of θ.

Although f_{sand}, f_{silt}, f_{clay}, f_q, f_{OM} are the components of soil solids, the sum of the feature importance of these five components is not close to that of soil bulk density. For the RF algorithm, we find that, the sum of the feature importance score of these five variables (f_{sand}, f_{silt}, f_{clay}, f_q, f_{OM}) is 0.152, which is smaller than the feature importance score of soil bulk density (0.253). For the GBDT algorithm, the trend is similar to that of RF. The sum of the feature importance score of these five variables is 0.135, which is also smaller than the feature importance score of soil bulk density (0.275).

Although we find that the ML methods significantly improve the accuracy of λ estimations, thus, provide a potential way to improve soil heat flux and energy balance estimations. Other soil thermal conductivity related research, such as thermal inertia (Colombo et al., 2019) and soil moisture monitoring will also benefit from improved λ prediction. We also realize that the ML estimation process is a "black box" and cannot provide a physically meaningful interpretation of the estimation results. The feature importance ranking does quantify the key factors affecting λ, which can lead to fine tuning or further development of empirical soil thermal conductivity models. However, in this research, there are only a limited number of training data points for clay and peat soils, which cause the performances of the ML algorithms to worsen for these soil types. To further test and evaluate the effectiveness of ML algorithms to estimate thermal conductivity and to validate the findings of this study, additional soil types (especially for clay and peat soils) can be evaluated and the number of data points can be expanded further.

4. Conclusions

By evaluating seven ML algorithms and five empirical models of soil thermal conductivity (Campbell, 1985; Côté and Konrad, 2005; de Vries, 1963; Johansen, 1975; Lu et al., 2007), we found that three of the ML algorithms (GBDT, NN, RF) provided much better and consistent λ estimations than the empirical models. GBDT, NN and RF reduced RMSE by at least 11.6% and improved NSE by 7.7% on the test set of dataset A compared with the empirical models. All three reduced RMSE by at least 16.0% and improved NSE by 12.7% on the test set of the smaller dataset B compared with the empirical models. Our study demonstrated that λ values of a wide variety of soil types could be effectively estimated by ML algorithms. Empirical models based on the concept of normalization (Lu et al., 2007; Côté and Konrad, 2005, Johansen, 1975) outperformed physical models (de Vries (1963)) and regression models (Campbell (1985)). Feature importance ranking of RF was more consistent than the ranking of GBDT. The RF-based feature importance ranking indicated that the sum of the importance values for θ and ρ_b was 81%.

Declaration of Competing Interest

The authors declare that they have no known conflict of interest or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Natural Science Foundation of China

(Grant No. 42077008, Grant No. 41771257, and 42177291), U.S. National Science Foundation (2037504) and USDA-NIFA Multi-State Project 4188.

Appendix A

Table A1-A2, Fig A1.

References

Ahmad, M.W., Mourshed, M., Rezgui, Y., 2017. Trees vs Neurons: comparison between random forest and ANN for high-resolution prediction of building energy consumption. Energy Build. 147, 77–89. https://doi.org/10.1016/j.enbuild.2017.04.038.

Al-Shammary, A.A.G., Kouzani, A.Z., Kaynak, A., Khoo, S.Y., Norton, M., Gates, W., 2018. Soil bulk density estimation methods: a review. Pedosphere 28, 581–596. https://doi.org/10.1016/s1002-0160(18)60034-7.

Albert, S., Linville, L., 2020. Benchmarking current and emerging approaches to infrasound signal classification. Seismol. Res. Lett. 91. https://doi.org/10.1785/0220190116.

Alireza, E., Hadi, N., Teimouri, M., Ahmadi, M., 2010. Comparison of neural network and K-nearest neighbor methods in daily flow forecasting. J. Appl. Sci. 10, 1006–1010. https://doi.org/10.3923/jas.2010.1006.1010.

Araya, S.N., Ghezzehei, T.A., 2019. Using machine learning for prediction of saturated hydraulic conductivity and its sensitivity to soil structural perturbations. Water Resour. Res. 55, 5715–5737. https://doi.org/10.1029/2018wr024357.

Aurélien, G., 2017. Hands-On Machine Learning with Scikit-Learn and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems, 1st ed. O'Reilly Media, Newton, MA, USA.

Bachmann, J., Horton, R., Ren, T., van der Ploeg, R.R., 2001. Comparison of the thermal properties of four wettable and four water-repellent soils. Soil Sci. Soc. Am. J. 65, 1675–1679. https://doi.org/10.2136/sssaj2001.1675.

Birhanu, I., Muktar, M., Kibebew, K., 2016. Impact of deforestation and subsequent cultivation on soil fertility in Komto, Western Ethiopia. J. Soil Sci. Environ. Manag. 7, 212–221. https://doi.org/10.5897/jssem2016.0578.

Bristow, K.L., White, R.D., Kluitenberg, G.J., 1994. Comparison of single and dual probes for measuring soil thermal properties with transient heating. Aust. J. Soil Res. 32, 447–464. https://doi.org/10.1071/sr9940447.

Bryce Meredig, E.A., Church, C., Hutchinson, M., Ling, J., Paradiso, S., Blaiszik, B., Foster, I., Gibbons, B., Hattrick-Simpers, J., Mehtaf, A., Ward, L., 2018. Can machine learning identify the next high-temperature superconductor? Examining extrapolation performance for materials discovery. Mol. Syst. Des. Eng. 3, 819–825. https://doi.org/10.1039/c8me00012c.

Butler, K.T., Davies, D.W., Cartwright, H., Isayev, O., Walsh, A., 2018. Machine learning for molecular and materials science. Nature 559, 547–555. https://doi.org/10.1038/s41586-018-0337-2.

Campbell, G.S., 1985. Soil Physics with BASIC: Transport Models for Soil-plant Systems. Elsevier, New York.

Casalicchio, G., Molnar, C., Bischl, B. 2018. Visualizing the feature importance for black box models. 10.1007/978-3-030-10925-7_40.

Chen, Y., Yang, K., Tang, W., Qin, J., Zhao, L., 2012. Parameterizing soil organic carbon's impacts on soil porosity and thermal parameters for Eastern Tibet grasslands. Sci. China Earth Sci. 55, 1001–1011. https://doi.org/10.1007/s11430-012-4433-0.

Clauser, C., Huenges, E., 1995. Thermal conductivity of rocks and minerals. Rock Physics & Phase Relations, pp. 105–126. https://doi.org/10.1029/RF003p0105.

Colombo, R., Garzonio, R., Di Mauro, B., Dumont, M., Tuzet, F., Cogliati, S., Pozzi, G., Maltese, A., Cremonese, E. 2019. Introducing thermal inertia for monitoring snowmelt processes with remote sensing. 46,4308-4319.10.1029/2019GL082193.

Côté, J., Konrad, J.M., 2005. A generalized thermal conductivity model for soils and construction materials. Can. Geotech. J. 42, 443–458. https://doi.org/10.1139/t04-106.

Cui, F.Q., Zhang, W., Liu, Z.Y., Wang, W., Chen, J.B., Jin, L., Peng, H., 2020. Assessment for thermal conductivity of frozen soil based on nonlinear regression and support vector regression methods. Adv. Civil Eng. 2020, 1–12. https://doi.org/10.1155/2020/8898126.

de Vries, D.A., 1963. Thermal properties of soil. In: van Dijk, W.R. (Ed.), Physics of Plant Environment. North Holland Publishing, Amsterdam.

Farouki, O.T., 1981. The thermal properties of soils in cold regions. Cold Reg. Sci. Technol. 67–75. https://doi.org/10.1016/0165-232X(81)90041-0.

Feinberg, E., Sur, D., Husic, B., Mai, D., Li, Y., Yang, J., Ramsundar, B., Pande, V. 2018. Spatial graph convolutions for drug discovery. https://arxiv.org/abs/1803.04465.

Flint, A.L., Flint, L.E., 2002. 2.2 particle density. Methods Soil Anal. 229–240. https://doi.org/10.2136/sssabookser5.4.c10.

Ghuman, B.S., Lal, R., 1985. Thermal conductivity, thermal diffusivity, and thermal capacity of some Nigerian soils. Soil Sci. 139, 74–80. https://doi.org/10.1097/00010694-198501000-00011.

Guo, L., Liu, J., Lu, R., 2019. Subsampling bias and the best-discrepancy systematic cross validation. Sci. China Math. 64, 197–210. https://doi.org/10.1007/s11425-018-9561-0.

He, H.L., Dyck, M.F., Horton, R., Ren, T.S., Bristow, K.L., Lv, J.L., Si, B.C., 2018. Development and application of the heat pulse method for soil physical measurements. Rev. Geophys. 56, 567–620. https://doi.org/10.1029/2017rg000584.

He, H.L., He, D., Jin, J.M., Smits, K.M., Dyck, M.F., Wu, Q.B., Si, B.C., Lv, J.L., 2020a. Room for improvement: a review and evaluation of 24 soil thermal conductivity parameterization schemes commonly used in land-surface, hydrological, and soil-vegetation-atmosphere transfer models. Earth Sci. Rev. 211, 103419 https://doi.org/10.1016/j.earscirev.2020.103419.

He, H.L., Noborio, K., Johansen, O., Dyck, M.F., Lv, J.L., 2020b. Normalized concept for modelling effective soil thermal conductivity from dryness to saturation. Eur. J. Soil Sci. 71, 27–43. https://doi.org/10.1111/ejss.12820.

He, H.L., Flerchinger, G.N., Kojima, Y., Dyck, M.F., Lv, J.L., 2021. A review and evaluation of 39 thermal conductivity models for frozen soils. Geoderma 382. https://doi.org/10.1016/j.geoderma.2020.114694.

Hu, G.J., Zhao, L., Wu, X.D., Li, R., Wu, T.H., Xie, C.W., Pang, Q.Q., Zou, D.F., 2017. Comparison of the thermal conductivity parameterizations for a freeze-thaw algorithm with a multi-layered soil in permafrost regions. Catena 156, 244–251. https://doi.org/10.1016/j.catena.2017.04.011.

Hultquist, C., Chen, G., Zhao, K.G., 2014. A comparison of Gaussian process regression, random forests and support vector regression for burn severity assessment in diseased forests. Remote Sens. Lett. 5, 723–732. https://doi.org/10.1080/2150704x.2014.963733.

Jahan, S., Islam, M.D.S., Islam, L., Rashme, T.Y., Prova, A.A., Paul, B.K., Islam, M.D.M., Mosharof, M.K., 2021. Automated invasive cervical cancer disease detection at early stage through suitable machine learning model. SN Appl. Sci. 3, 806. https://doi.org/10.1007/s42452-021-04786-z.

Johansen, O., 1975. Varmeledningsevne av jordarter (Thermal conductivity of soils). University of Trondheim, Trondheim, Norway., US Army Corps of Engineers, Cold Regions Research and Engineering Laboratory, Hanover, N.H. CRREL Draft English Translation, p. 637.

Jordan, M.I., Mitchell, T.M., 2015. Machine learning: trends, perspectives, and prospects. Science 349, 255–260. https://doi.org/10.1126/science.aaa8415.

Kasubuchi, T., Momose, T., Tsuchiya, F., Tarnawski, V., 2007. Normalized thermal conductivity model for three Japanese soils. Trans. Jpn. Soc. Irrig. Drain. Rural Eng. (Jpn.) 251, 529–533.

Keller, T., Håkansson, I., 2010. Estimation of reference bulk density from soil particle size distribution and soil organic matter content. Geoderma 154, 398–406. https://doi.org/10.1016/j.geoderma.2009.11.013.

Kim, D., Oh, S., 2020. Measurement and comparison of thermal conductivity of porous materials using box, dual-needle, and single-needle probe methods-a case study. Int. Commun. Heat Mass Transf. 118, 9. https://doi.org/10.1016/j.icheatmasstransfer.2020.104815.

Lever, J., Krzywinski, M., Altman, N., 2016. Classification evaluation. Nat. Methods 13, 603–604. https://doi.org/10.1038/nmeth.3945.

Li, P., Qin, Z., Wang, X.H., Metzler, D., 2019. Combining decision trees and neural networks for learning-to-rank in personal search. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD). Assoc Computing Machinery, Anchorage, AK, pp. 2032–2040. https://doi.org/10.1145/3292500.3330676.

Li, Y., Shao, M., 2005. Latest advance of thermo-pulse method for measuring soil thermal properties (in Chinese with English abstract). Acta Pedol. Sin. 42, 134–139.

Li, Z., Wang, S.W., Chin, W.S., Achenie, L.E., Xin, H.L., 2017. High-throughput screening of bimetallic catalysts enabled by machine learning. J. Mater. Chem. A 5, 24131–24138. https://doi.org/10.1039/c7ta01812f.

Liakos, K.G., Busato, P., Moshou, D., Pearson, S., Bochtis, D., 2018. Machine learning in agriculture: a review. Sensors 18, 2674. https://doi.org/10.3390/s18082674.

Liu, C.Y., Hu, X.M., Yao, R., Han, Y.L., Wang, Y., He, W.T., Fan, H.B., Du, L.Z., 2020. Assessment of soil thermal conductivity based on BPNN optimized by genetic algorithm. Adv. Civil Eng. 2020, 1–10. https://doi.org/10.1155/2020/6631666.

Liu, G., Si, B.C., 2011. Single- and dual-probe heat pulse probe for determining thermal properties of dry soils. Soil Sci. Soc. Am. J. 75, 787–794. https://doi.org/10.2136/sssaj2010.0241.

Liu, L.M., He, H.L., Dyck, M., Lv, J.L., 2021. Modeling thermal conductivity of clays: A review and evaluation of 28 predictive models. Eng. Geol. 288, 17. https://doi.org/10.1016/j.enggeo.2021.106107.

Lu, S., Ren, T., Gong, Y., Horton, R., 2007. An improved model for predicting soil thermal conductivity from water content at room temperature. Soil Sci. Soc. Am. J. 71, 8–14. https://doi.org/10.2136/sssaj2006.0041.

Lu, S., Ren, T., Yu, Z., Horton, R., 2011. A method to estimate the water vapour enhancement factor in soil. Eur. J. Soil Sci. 62, 498–504. https://doi.org/10.1111/j.1365-2389.2011.01359.x.

Lv, Z., Lou, R., Feng, H., Chen, D., Lv, H. 2021. Novel machine learning for big data analytics in Intelligent support information management systems. 13,Article 7.10.1145/3469890.

Ly, H.B., Nguyen, T.A., Pham, B.T., 2021. Estimation of soil cohesion using machine learning method: a random forest approach. Adv. Civil Eng. 2021, 14. https://doi.org/10.1155/2021/8873993.

Meyer, H., Reudenbach, C., Wöllauer, S., Nauss, T., 2019. Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. Ecol. Modell. 411, 108815 https://doi.org/10.1016/j.ecolmodel.2019.108815.

Montgomery, R.B., 1947. Viscosity and thermal conductivity of air and diffusivity of water vapor in air. J. Meteorol. 4, 193–196. https://doi.org/10.1175/1520-0469(1947)004<0193:VATCOA>2.0.CO;2.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I – a discussion of principles. J. Hydrol. 10, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6.

Ochsner, T.E., Horton, R., Ren, T., 2001. A new perspective on soil thermal properties. Soil Sci. Soc. Am. J. 65, 1641–1647. https://doi.org/10.2136/sssaj2001.1641.

Padarian, J., Minasny, B., McBratney, A.B., 2020. Machine learning and soil sciences: a review aided by machine learning tools. Soil 6, 35–52. https://doi.org/10.5194/soil-6-35-2020.

Peters-Lidard, C.D., Blackburn, E., Liang, X., Wood, E.F., 1998. The effect of soil thermal conductivity parameterization on surface energy fluxes and temperatures. J. Atmospheric Sci. 55, 1209–1224. https://doi.org/10.1175/1520-0469(1998)055<1209:Teostc>2.0.Co;2.

Quezada, J.L.M., 2019. Forecasting Crashes, Credit Card Default, and Imputation Analysis on Missing Values by the use of Neural Networks. University of Texas at El Paso, Open Access Theses & Dissertations, p. 2006.

Ren, X., Mi, Z.Y., Georgopoulos, P.G., 2020. Comparison of Machine Learning and Land Use Regression for fine scale spatiotemporal estimation of ambient air pollution: Modeling ozone concentrations across the contiguous United States. Environ. Int. 142, 13. https://doi.org/10.1016/j.envint.2020.105827.

Rizvi, Z.H., Husain, S.M.B., Haider, H., Wuttke, F., 2020. Effective thermal conductivity of sands estimated by Group Method of Data Handling (GMDH). Mater. Today Proc. 26, 2103–2107. https://doi.org/10.1016/j.matpr.2020.02.454.

Senanayake, I.P., Yeo, I.Y., Walker, J.P., Willgoose, G.R., 2021. Estimating catchment scale soil moisture at a high spatial resolution: Integrating remote sensing and machine learning. Sci. Total Environ. 776, 145924 https://doi.org/10.1016/j.scitotenv.2021.145924.

Smirnov, A.G., Berrendorf, M., Shprits, Y.Y., Kronberg, E.A., Allison, H.J., Aseev, N.A., Zhelavskaya, I.S., Morley, S.K., Reeves, G.D., Carver, M.R., Effenberger, F. 2020. Medium energy electron flux in earth's outer radiation belt (MERLIN): a machine learning model. 18,e2020SW002532.10.1029/2020SW002532.

Song, Y.Y., Lu, Y., 2015. Decision tree methods: applications for classification and prediction. Shanghai Arch. Psychiatry 27, 130–135. https://doi.org/10.11919/j.issn.1002-0829.215044.

Tang, Y., Li, J., 2018. Test method and application for microstructures of undisturbed silty sand and sandy silt. Environ. Earth Sci. 77, 657. https://doi.org/10.1007/s12665-018-7847-y.

Tarnawski, V.R., Momose, T., Leong, W.H., 2009. Assessing the impact of quartz content on the prediction of soil thermal conductivity. Géotechnique 59, 331–338. https://doi.org/10.1680/geot.2009.59.4.331.

Tarnawski, V.R., McCombie, M.L., Leong, W.H., Wagner, B., Momose, T., Schönenberger, J., 2012. Canadian field soils II. Modeling of quartz occurrence. Int. J. Thermophys. 33, 843–863. https://doi.org/10.1007/s10765-012-1184-2.

Tarnawski, V.R., Momose, T., McCombie, M.L., Leong, W.H., 2015. Canadian field soils III. Thermal-conductivity data and modeling. Int. J. Thermophys. 36, 119–156. https://doi.org/10.1007/s10765-014-1793-z.

Tolimir, M., Kresovic, B., Zivotic, L., Dragovic, S., Dragovic, R., Sredojevic, Z., Gajic, B., 2020. The conversion of forestland into agricultural land without appropriate measures to conserve SOM leads to the degradation of physical and rheological soil properties. Sci. Rep. 10, 12. https://doi.org/10.1038/s41598-020-70464-6.

Tong, B., Gao, Z.Q., Horton, R., Li, Y.B., Wang, L.L., 2016. An empirical model for estimating soil thermal conductivity from soil water content and porosity. J. Hydrometeorol. 17, 601–613. https://doi.org/10.1175/jhm-d-15-0119.1.

Wada, A., Tsuruta, K., Irie, R., Kamagata, K., Maekawa, T., Fujita, S., Koshino, S., Kumamaru, K., Suzuki, M., Nakanishi, A., Hori, M., Aoki, S., 2019. Differentiating Alzheimer's disease from dementia with lewy bodies using a deep learning technique based on structural brain connectivity. Magn. Resonance Med. Sci. 18, 219–224. https://doi.org/10.2463/mrms.mp.2018-0091.

Wang, J.M., He, H.L., Dyck, M., Lv, J.L., 2020. A review and evaluation of predictive models for thermal conductivity of sands at full water content range. Energies 13, 15. https://doi.org/10.3390/en13051083.

Wang, S., Wang, Q., Fan, J., Wang, W., 2012. Soil thermal properties determination and prediction model comparison (in Chinses with English abstract). Trans. Chin. Soc. Agric. Eng. 28, 78–84.

Weng, Y., Fang, Y., Yan, H., Yang, Y., Hong, W., 2019. Bayesian non-parametric classification with tree-based feature transformation for NIPPV efficacy prediction in COPD patients. IEEE Access 7, 177774–177783. https://doi.org/10.1109/ACCESS.2019.2958047.

Yan, H., He, H., Dyck, M., Jin, H., Li, M., Si, B., Lv, J., 2019. A generalized model for estimating effective soil thermal conductivity based on the Kasubuchi algorithm. Geoderma 353, 227–242. https://doi.org/10.1016/j.geoderma.2019.06.031.

Yurttakal, A.H., 2021. Extreme gradient boosting regression model for soil thermal conductivity. Therm. Sci. 25, 1–7. https://doi.org/10.2298/tsci200612001y.

Zhang, M., Bi, J., Chen, W., Zhang, X., Lu, J., 2018. Evaluation of calculation models for the thermal conductivity of soils. Int. Commun. Heat Mass Transf. 94, 14–23. https://doi.org/10.1016/j.icheatmasstransfer.2018.02.005.

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

Zhang, T., Wang, C.J., Liu, S.Y., Zhang, N., Zhang, T.W., 2020b. Assessment of soil thermal conduction using artificial neural network models. Cold Reg. Sci. Technol. 169, 102907 https://doi.org/10.1016/j.coldregions.2019.102907.

Zhao, H., Zeng, Y.J., Lv, S.N., Su, Z.B., 2018. Analysis of soil hydraulic and thermal properties for land surface modeling over the Tibetan Plateau. Earth System Science Data 10, 1031–1061. https://doi.org/10.5194/essd-10-1031-2018.

Zhao, Y., Si, B., Zhang, Z., Li, M., He, H., Hill, R.L., 2019. A new thermal conductivity model for sandy and peat soils. Agric. For. Meteorol. 274, 95–105. https://doi.org/10.1016/j.agrformet.2019.04.004.