Heat and Mass Transfer (2022) 58:1695–1708
https://doi.org/10.1007/s00231-022-03209-y

ORIGINAL ARTICLE

Prediction of soil thermal conductivity based on Intelligent computing model

Caijin Wang[1] · Guojun Cai[1,2] · Xuening Liu[1] · Meng Wu[1]

Received: 7 December 2021 / Accepted: 24 March 2022 / Published online: 7 April 2022

Abstract

Thermal conductivity is a basic characteristic of the heat conduction properties of subsoil. Previous research shows that soil thermal conductivity has complex correlations with many soil physical parameters, such as dry density, water content, mineral composition and particle-size distribution. In this paper, several artificial intelligence calculation methods are used to study the soil heat conduction mechanism and establish predictive models of thermal conductivity: an artificial neural network (ANN), adaptive neural network-based fuzzy inference system (ANFIS) and support vector machine (SVM). Their modelling performance was evaluated by several metrics: correlation coefficient (R^2), root mean square error (*RMSE*), mean absolute error (*MAE*) and variance account for (*VAF*). Monte Carlo simulation was used to verify the robustness of the models, and the results of traditional empirical relationship models are used for comparison. The ANN, ANFIS and SVM models can accurately predict soil thermal conductivity, with $R^2 > 0.89$, $RMSE < 0.22$ (Wm^{-1} K^{-1}), $MAE < 0.14$ (Wm^{-1} K^{-1}) and $VAF > 88\%$. The ANN model had the best predictive accuracy, with $R^2 = 0.9535$, $RMSE = 0.1338$ (Wm^{-1} K^{-1}), $MAE = 0.0952$ (Wm^{-1} K^{-1}) and $VAF = 95.25\%$. The SVM model had similar accuracy, while that of the ANFIS model was lower. Monte Carlo simulations show that the SVM model provided the most robust predictions and that all three models were significantly better than the traditional empirical models. The SVM model is suggested as the best model for predicting soil thermal conductivity.

Nomenclature

ANN	Artificial neural network
ANFIS	Adaptive neural network-based fuzzy inference system
SVM	Support vector machine
R^2	Correlation coefficient
RMSE	Root mean square error
MAE	Mean absolute error
VAF	Variance account for
N_1, N_2, N_k	Input parameters
Y	Output parameter
x_{max}	Maximum values
x_{min}	Minimum values
x	Actual value
x_{norm}	Normalized value
λ_{dry}	Thermal conductivities of dry soil [Wm^{-1} K^{-1}]
λ_{sat}	Thermal conductivities of saturated soil [Wm^{-1} K^{-1}]
γ_d	Dry density [kgm^{-3}]
n	Porosity [%]
λ	Thermal conductivity [Wm^{-1} K^{-1}]
H_k	The predicted value
$w_{x,ji}$	Weighting factor
$b_{x,j}$	Bias factor at the jth hidden node
$w_{y,j}$	Weighting factor for the j^{th} hidden node
b_y	Bias factor in the output layer
M	Number of parameters evaluated by the same regression process
N	Sample number
y_0	Measured value
y_p	Predicted value
m	Number of Monte Carlo iterations

✉ Guojun Cai
focuscai@163.com

Caijin Wang
wangcaijin@seu.edu.cn

Xuening Liu
1455613868@qq.com

Meng Wu
542779989@qq.com

[1] Institute of Geotechnical Engineering, Southeast University, Nanjing, Jiangsu 211189, China

[2] School of Civil Engineering, Anhui Jianzhu University, Hefei 230601, China

S	Actual random variable considered
K_e	Normalized thermal conductivity of the soil
S_r	Soil saturation
λ_{water}	Thermal conductivities of water [Wm^{-1} K^{-1}]
λ_{solid}	Thermal conductivities of soil solid particles [Wm^{-1} K^{-1}]
κ	Empirical parameters
χ,η	Influencing parameters of thermal conductivity of dry soil
a, b	Parameters related to dry soil
α	Reflects the influence of soil type on the Kersten variable

1 Introduction

With the recent rapid rates of industrialization and urbanization, the rates of fossil fuel consumption and global warming are increasing. Accordingly, demands for renewable energy development and utilization are also increasing [1]. Soil has the potential to store energy but requires further development before it can be applied. Accurate measurement of the thermal conductivity of geotechnical materials is an important part of geotechnical engineering temperature field analysis and energy pile design. The thermal conductivity of soil is a complex parameter that is mainly dependent on the soil type, saturation, grain size and packing density. The type of soil is considered to be the most important influence on its thermal resistance [2–5]. Thermal conductivity is one of the most important thermophysical properties of soils, inasmuch it is directly connected with the heat transfer mechanism by conduction in materials [6].

In recent years, there has been much research on soil heat conduction and many predictive models have been proposed. Kersten [7] measured the thermal conductivity of 19 different types of soil using vertical heating from a single-column heat source. They analysed the influences of moisture content and dry density on thermal conductivity, proposed an empirical relationship model for predicting silt, clay and sandy soil. Coté and Konrad [8] measured the thermal conductivity of more than 650 soil samples and improved Johansen [9] normalized model, establish the κ-S_r relationshiped with κ as variable defined in par.6, and used κ to reflect the influence of the type of soil on its thermal conductivity. Based on the measured thermal conductivity of clay, silt, silty sand and fine sand, Erzin et al. [10] proposed predictive models for single samples and specific types of samples, respectively, using on an artificial neural network. Wang et al. [11] measured the thermal conductivity and electrical resistivity of 57 soil samples and found a linear relationship between them, although soil saturation was not considered. Based on published thermal conductivity data of 257 unsaturated soils, Zhang et al. [12] used dry density, porosity, saturation, quartz content, sand content and clay content as input to an artificial neural network predictive model. They found that the model can comprehensively consider the influences of these parameters and provide good predictions. Bi et al. [13] proposed a general model for calculating the thermal conductivity of frozen soil with consideration of soil composition and frost heave. The model is a function of unfrozen water content, frost heave effect, porosity and initial water content. In summary, predictions of soil thermal conductivity are commonly made using empirical relationship models, theoretical calculation models and artificial neural network models. Some authors assert that empirical models give much better accuracy than theoretical models [14]. Theoretical calculation models of soil thermal conductivity involve many parameters and the calculation process is complicated, leading to a difficult practical application. Artificial neural network models can consider the most factors and have the highest accuracy in predicting the thermal conductivity of soil.

The aim of this paper is to establish a predictive model of soil thermal conductivity. Firstly, the main factors influencing soil thermal conductivity are identified, then an artificial neural network model, fuzzy neural network model and support vector machine model are used to predict soil thermal conductivity. The results of the three models are compared with traditional empirical relationship models to verify their effectiveness. Robustness is an important indicator for evaluating intelligent computing models. Monte Carlo simulations are used to evaluate the robustness of prediction models. The research results are of great significance for discussing heat transfer mechanism of soil, and provide a new idea for accurately predicting soil thermal conductivity.

2 Factors influencing soil thermal conductivity

Thermal conductivity is an important soil thermophysical property. It is closely related to many factors, such as mineral composition, saturation, particle size distribution, dry density and porosity, among which mineral composition has the greatest influence [15]. In the present section the influence of soil geotechnical parameters on thermal conductivity is explored.

Moisture content is the most basic parameter describing the physical properties of soil and is also one of the main factors affecting soil heat transfer. Unsaturated soil is composed of a three-phase medium, and changes in water content mainly affect the proportions of liquid and gas in the soil. An increase in water content increases the proportion of water phase in the soil, while the proportion of gas phase decreases. Since the thermal conductivity of water is significantly greater than that of air, heat transfer through water films on particle surfaces and water bridges increases, so increasing the effective thermal conductivity

of the soil [7, 8]. As the water content increases, the thermal conductivity eventually reaches a maximum and doesn't increase any more, at a specific water content [16]. Soil with a greater dry density has closer interparticle contact, resulting in greater interparticle heat transfer and higher thermal conductivity.

The mineral composition of soil plays an important role in its thermal conductivity. The solid particles of soil are mainly composed of minerals such as quartz, feldspar and mica [17]. The thermal conductivity of quartz is about 7.69 Wm^{-1} K^{-1}, while that of other minerals is usually between 1.25–4 Wm^{-1} K^{-1}. The thermal conductivity of quartz is significantly higher than that of other minerals, so soils with a high quartz content have significantly higher thermal conductivity [18]. Therefore, the thermal conductivity of soil depends on its mineral composition, especially the quartz content.

The anisotropy of soils also influences their thermal conductivity. Macaulay et al. [19] tested the thermal conductivity of Australian siltstone and found that the bedding-direction angle of siltstone began to increase and its thermal conductivity also increased. The thermal conductivity of siltstone varies smoothly with the increase in bedding-direction angle at angles > 40°. This conclusion is consistent with Popov et al.'s [20] research on sedimentary rocks.

Many factors can affect the thermal conductivity of soils, such as particle-size distribution, particle size, particle shape and temperature [21–23]. The particle-size distribution changes the interparticle contact area, thus affecting thermal conductivity [24], as heat conducts more rapidly through particles than between them. At a given dry density, the smaller the particle size is, the more interparticle contact points and the lower the thermal conductivity of soil are [9]. Additionally, temperature affects the thermal motion of molecules. Mitchell and Soga [25] found that the thermal conductivity of all crystalline minerals decreases with increasing temperature. However, the thermal conductivity of soil water increases slightly as temperature increases. Furthermore, increases in temperature significantly increase the thermal conductivity of the air in soils, although some researchers have reached the opposite conclusion. Therefore, the influences on soil thermal conductivity are complex and require further study to more clearly understand them.

3 Artificial intelligence computing model

Artificial intelligence technology has developed rapidly and has recently been applied to the field of geotechnical engineering. It can accurately calculate engineering parameters in a time and cost-effective manner so that projects can be completed in less time and at lower cost. For this purpose, in this paper there different algorithms are described and tested in calculating the thermal conductivity of soils: an artificial neural network model (ANN), an adaptive neural network-based fuzzy inference system (ANFIS) and a support vector machine model (SVM).

3.1 Artificial neural network analysis

Artificial neural network models are based on an information processing system that mimics the structure and function of the human brain [26, 27]. An ANN is a complex network composed of many neurons that can process and learn a large amount of input data. Many neural network algorithms have been developed, such as the Levenberg–Marquardt algorithm, conjugate gradient algorithm and Bayesian regularization algorithm [28–30]. A neural network model consists of an input layer, one or more hidden layers and an output layer. The layers are composed of many neurons. The number of hidden layers depends on the complexity of the problem. For general geotechnical engineering, one hidden layer is enough effective [31]. The function of the input layer is to obtain input data, filter and pass them to the hidden layer. The main function of the hidden layer node function is to receive data from the input layer and learn from them to construct a training model. The output layer performs the final step of processing the data and deriving calculated values [32]. Artificial neural networks have been effectively applied to geotechnical engineering problems such as foundation pit support, tunnel monitoring, slope treatment and prediction of soil thermal conductivity [33–39]. Figure 1 shows a typical neural network structure, where N_1, N_2 and N_k are input parameters and Y is an output parameter. The advantage of ANN models is that there is no need to restrict the input data. The model only needs to learn the given input data and then calculate the result.

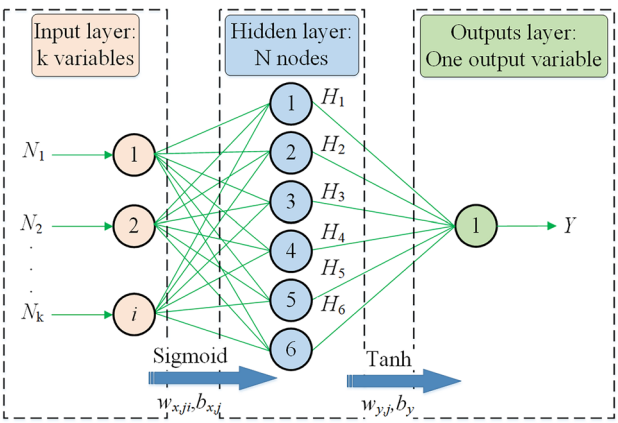


Fig. 1 The recurrent neural network architecture

This paper trains an established neural network model based on the Bayesian regularization algorithm, which uses ridge regression to convert nonlinear regression problems into statistical ones. The advantage of this method is that it has better performance than other backpropagation algorithms because an ANN based on a Bayesian regularization algorithm does not need to use cross-validation in the verification process, which can simplify model training and reduce calculation time.

3.2 Fuzzy inference system based on adaptive neural network

The adaptive neural network-based fuzzy inference system (ANFIS) is a new type of fuzzy inference system structure that organically combines fuzzy logic and neural networks. It uses a hybrid algorithm of backpropagation and least squares to adjust the first trial and conclusion parameters, and

Table 1 Physical properties 40 Canadian

Soil	Code	Texture	λ_{dry}	λ_{sat}	Quartz content(%)	γ_d (kgm^{-3})	n	GSD(%)		
								Clay	Silt	Sand
1	AB-1	Silt loam	0.2	1.2	55	2640	0.55	10	52	38
2	BC-1	Silty clay	0.2	1.21	21	2740	0.51	42	58	0
3	BC-2	Silty clay	0.2	1.3	19	2718	0.5	42	58	0
4	BC-3	Silt clay loam	0.2	1.12	27	2713	0.51	30	70	0
5	BC-4	Silty clay	0.2	1.14	17	2782	0.52	41	59	0
6	BC-5	Silt clay loam	0.2	1.35	17	2775	0.53	33	67	0
7	BC-6	Silty loam	0.2	1.46	37	2757	0.52	10	58	32
8	MN-1	Silt loam	0.3	1.97	38	2685	0.55	14	69	17
9	MN-2	Silt loam	0.2	1.27	20	2788	0.41	24	55	22
10	MN-3	Silt loam	0.2	1.05	21	2739	0.63	21	76	3
11	MN-4	Loamy sand	0.2	1.46	61	2706	0.47	3	15	81
12	NB-1	Silt loam	0.2	1.16	57	2590	0.54	15	82	3
13	NB-2	Silt loam	0.2	1.43	56	2540	0.56	17	83	0
14	NB-3	Silt loam	0.3	2.19	55	2569	0.62	10	66	24
15	NB-4	Silt loam	0.2	1.4	60	2588	0.54	10	64	26
16	NB-5	Silt clay loam	0.1	1.14	39	2707	0.54	33	67	0
17	NS-1	Silt loam	0.2	1.31	51	2708	0.55	10	57	32
18	NS-2	Sandy loam	0.3	1.46	61	2711	0.45	5	34	61
19	NS-3	Sandy loam	0.2	1.43	63	2680	0.4	5	37	57
20	NS-4	Sand	0.2	1.46	100	2662	0.36	0	0	100
21	NS-5	Loamy sand	0.3	1.6	72	2662	0.4	3	13	85
22	NS-6	Sandy loam	0.2	1.39	65	2684	0.51	6	38	56
23	NS-7	Silt loam	0.3	1.92	34	2781	0.57	12	67	22
24	ON-1	Silt loam	0.3	1.94	28	2704	0.43	8	56	37
25	ON-2	Silt loam	0.3	1.74	17	2758	0.51	18	75	7
26	ON-3	Loamy sand	0.2	1.76	41	2706	0.46	4	26	71
27	ON-4	Sand	0.3	2.17	38	2755	0.39	1	10	89
28	ON-5	Sandy loam	0.3	1.93	36	2754	0.38	7	37	56
29	ON-6	Loamy sand	0.2	1.73	38	2739	0.44	2	14	84
30	ON-7	Silt loam	0.2	1.84	25	2760	0.45	14	54	32
31	PE-1	Loam	0.2	1.52	66	2636	0.44	8	42	50
32	PE-2	Loam	0.2	1.57	58	2663	0.42	9	39	51
33	PE-3	Loamy sand	0.2	1.93	54	2656	0.41	3	14	83
34	QC-1	Sand	0.2	1.82	35	2727	0.43	2	5	93
35	QC-2	Loamy sand	0.3	1.98	42	2693	0.48	3	17	79
36	SK-1	Silt loam	0.2	1.6	48	2693	0.41	26	74	0
37	SK-2	Sandy loam	0.3	2.39	61	2703	0.45	6	27	67
38	SK-3	Silt loam	0.3	1.67	37	2702	0.53	15	83	2
39	SK-4	Loamy sand	0.2	1.59	67	2683	0.42	3	14	83
40	SK-5	Sandy loam	0.3	3.17	63	2677	0.45	5	28	68

automatically generates *If–Then* rules. A fuzzy reasoning system is very suitable for expressing fuzzy experience knowledge but lacks an effective learning mechanism. Although neural networks have a self-learning function, they cannot express the reasoning function of the human brain well. ANFIS organically combines a neural network and fuzzy reasoning. It not only exerts the advantages of both but also makes up for their shortcomings. ANFISs are hybrid models in which multiple nodes related to direction connections can estimate the fuzzy parameters of the model. This paper uses a subtractive clustering algorithm to train the ANFIS model, which can automatically estimate the number of clusters and their locations. The main process of the subtractive clustering algorithm is that it 1) selects the data most likely to be the first cluster centre and then 2) deletes all the data located around the first cluster centre that affects the radius definition. Finally, the process is repeated until all the data fall within the radius of the cluster centre [40].

3.3 Support vector machine model

The support vector machine (SVM) model is a generalized linear classifier that performs binary classification of data with a supervised learning method. Its theory is to establish a hyperplane to divide a dataset into different categories [41]. The SVM uses a hinge-loss function to calculate the empirical risk and adds a regularization term to the solution system to optimize the structural risk, and then determines the optimal hyperplane by optimizing the boundary. The support vector is determined as the training point closest to the optimal plane. SVM is widely used in geotechnical engineering, such as for landslide calculation, because it is the most accurate and precise modelling technique [42].

4 Thermal conductivity databases

The data used in this paper comes from 40 types of Canadian field soil [43–45]. Among them, 22 species are fine-grained soil (sand content < 40%) and 18 are coarse-grained soil (sand content > 40%). The measurements were carried out on moderately compacted samples at room temperature, and over a full range of degree of saturation ranging from dryness to full saturation, a total of 240 test data. The physical properties of the 40 Canadian soils are shown in Table 1. Their moisture content, dry density, clay content, powder content, sand content and quartz content were used as input parameters for the predictive model, with soil thermal conductivity as the output parameter. Table 2 lists the boundary values of the input and output parameters of the predictive model. In addition, the input and output parameters of all models are normalized to make them range between 0 and 1. The normalization equation is [10]:

$$x_{norm} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

where $x_{\max}$ and $x_{\min}$ are the maximum and minimum values of the variable x, respectively, x_{norm} is the normalized value.

5 Establishment and performance testing of the intelligent calculation model

5.1 Establishment of the intelligent calculation model

The ANN, ANFIS and SVM were used to establish an intelligent calculation model for predicting soil thermal conductivity. According to the analysis of the factors affecting the thermal conductivity of soil (Sect. 2), the input parameters for all calculation models were water content, dry density, clay content, powder content, sand content and quartz content, with the output parameter being soil thermal conductivity. In this paper, the Bayesian regularization algorithm was used to train the developed neural network calculation model. An ANN based on the Bayesian regularization algorithm is a mathematical calculation process that uses ridge regression to convert nonlinear regression into a statistical problem [30]. This method is better than other backpropagation methods because the

Table 2 The boundary values of the input and output parameters of the predictive model

Parameter	Name	Minimum value	Maximum value
Input parameters	Moisture content, w (%)	0	75
	Dry density, γ_d (kgm^{-3})	520	2790
	Clay content, c (%)	0	42
	Silt content, s_i (%)	0	83.4
	Sand content, s_a (%)	0	100
	Quartz content, q_c (%)	17	100
Output parameter	Thermal conductivity, λ (Wm^{-1} K^{-1})	0.1	3.2

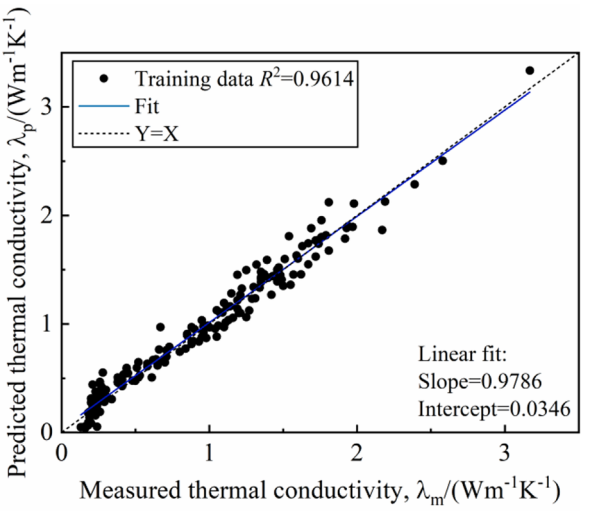


Fig. 2 Comparison of measurements of soil thermal conductivity with estimates of the training parts of the ANN (**a**); ANFIS (**b**) and SVM (**c**)

verification process does not need to use cross-validation technology, which can save calculation time. The data was divided into two parts: 70% for training and 30% for testing. The training part uses the data to update the weights of neurons, while model verification uses the test data. The number of hidden layer neurons in the ANN model increases from 1, and the optimal number of hidden layers and hidden layer neurons is obtained through training. We tried different activation functions during training, such as sigmoid, tanh and Relu, and achieved the best performance with the sigmoid activation function. The number of neural network training periods was 1000. We used the following standard procedures to model the artificial lift network [12, 46]:

(1) The input and output parameters are normalized so that they range between 0 and 1.
(2) By taking N_k inputs, the sigmoid activation function of each node in the hidden layer can be calculated as:

$$H_k = sigmoid\left(\sum_{i=1}^{N} w_{x,ji}N_k + b_{x,j}\right), \quad \text{where } sigmoid = \left(\frac{1}{1+\exp(-x)}\right) \tag{2}$$

where H_k, $w_{x,ji}$, and $b_{x,j}$ are the predicted value, weighting factor, and bias factor at the j^{th} hidden node, respectively.
(3) Calculation of the output layer nodes of the ANN is as follows:

$$Y_n = \sum_{j=1}^{M} w_{y,j}H_k + b_y \tag{3}$$

where $w_{y,j}$ is the weighting factor for the j^{th} hidden node and b_y is the bias factor in the output layer.
(4) Inverse normalization of output data:

$$Y = Y_n\left[x_{max} - x_{min}\right] + x_{min} \tag{4}$$

The ANFIS model uses a Gaussian membership function and subtractive clustering technology with an influence range of 0.7 for training. The SVM model is constructed by a cubic polynomial kernel function, the regularization constant c of the selection box constraint is 0.112, and the ε parameter of the SVM model is 0.0111. Figure 2 compares measurements of soil thermal conductivity with estimates of the intelligent calculation model (training set).

It can be seen from Fig. 2 that each intelligent calculation model accurately calculates the thermal conductivity of soil based on the training dataset. There were linear correlations between the predicted and measured results (Fig. 2). The best training effect was exhibited by the ANN model ($R^2 = 0.9614$), followed by the SVM model ($R^2 = 0.9551$), while the ANFIS model performed worst ($R^2 = 0.9212$). The slopes of the linear equations fitted to the ANN, ANFIS and SVM models are 0.9786, 0.8475 and 0.9551, respectively. Hence, the fits to the ANN and SVM cases are close to perfect ($y = x$). This shows that the deviations between the measured values and those predicted by the ANN and SVM are small.

5.2 Performance test of prediction model

The correlation coefficient (R^2), root mean square error (*RMSE*), average absolute error (*MAE*) and variance account for (*VAF*) were used to test the performance of the calculation models. An important indicator of correlation between the calculated and measured values is the correlation coefficient R^2. Three important indicators of the model, *RMSE*, *MAE* and *VAF*, were used to quantify the numerical differences between the calculated and measured values [47–49]. Their equations are:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y_0 - y_p)^2}{N - M}} \tag{5}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} \left| y_0 - y_p \right| \tag{6}$$

$$VAF = \left[1 - \frac{\mathrm{var}(y_0 - y_p)}{\mathrm{var}(y_0)} \right] \times 100 \tag{7}$$

where M is the number of parameters evaluated by the same regression process, N is the sample number, y_0 is the measured value, y_p is the predicted value, and var represents the variance. If $VAF = 100\%$ and $RMSE = 0$, the model is perfect. Table 3 shows the performance index values of each predictive model. Figure 3 compares the predicted and measured values of soil thermal conductivity λ.

It can be seen from Table 3 and Fig. 3 that each model has high accuracy in predicting the soil thermal conductivity λ. The ANN model performs best, second is the SVM model, while the ANFIS model has the lowest accuracy. It can be seen from Fig. 3b that the predicted values deviate greatly from the measured values and the data present a relatively large spread. The slopes of the linear equations fitted to the ANN, ANFIS and SVM models are 0.9452, 0.8197 and 0.9594, respectively, indicating that the fits of the ANN and SVM models to data are closest to the $y = x$ line.

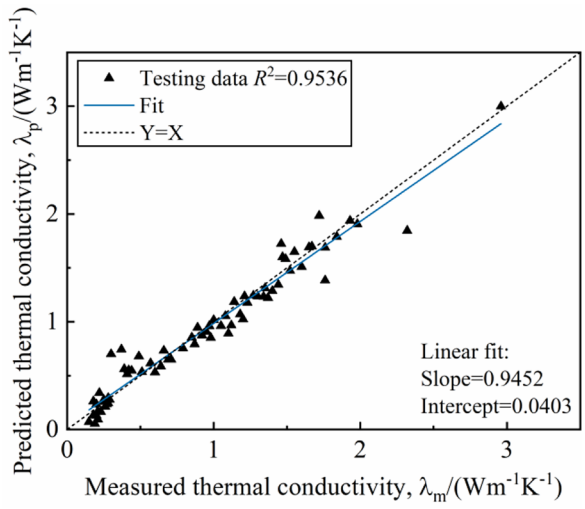
The function of a sample index comparison of predicted and measured values is shown in Fig. 4. It can be seen from Fig. 4 that the ANN and SVM calculated values are in good agreement with the measured values, while they deviate obviously from the predicted values of the ANFIS model. In order to analyse the error between the predicted and measured values, the error distributions of the predictive models are shown in Fig. 5. It can be seen from Fig. 5 that in the training and testing datasets, the error distribution of each model is mainly concentrated between − 0.3 ($\mathrm{Wm^{-1}\,K^{-1}}$) and 0.3 ($\mathrm{Wm^{-1}\,K^{-1}}$). The calculation errors of each model are very small; the error of the ANN model is the smallest; next is the SVM model and then the ANFIS model. However, it can be seen from Table 4 that the SVM model performs best in terms of *RMSE* (0.1333 $\mathrm{Wm^{-1}\,K^{-1}}$), while the *RMSE*s of the ANN and ANFIS models are 0.1338 ($\mathrm{Wm^{-1}\,K^{-1}}$) and 0.2105 ($\mathrm{Wm^{-1}\,K^{-1}}$), respectively when applied to the test dataset. The above analysis shows that the ANN model has the best performance. The results show that the weights and deviations of the ANN model based on Bayesian regularization improve its performance. The ANFIS and SVM models also predict the thermal conductivity of soil well.

5.3 Robustness analysis of forecasting model

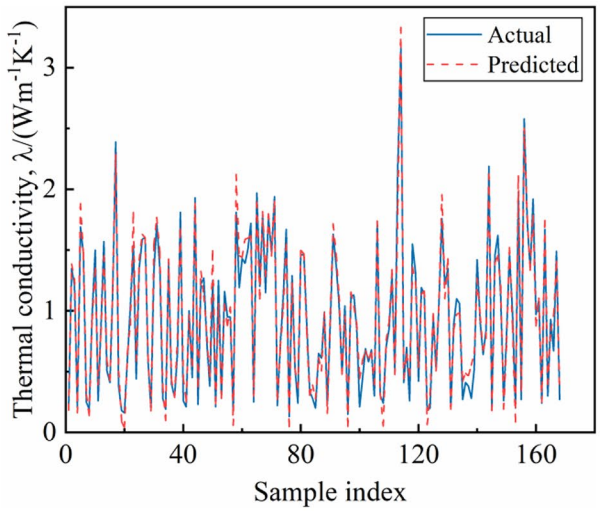
When using AI models for complex geotechnical engineering prediction, the Monte Carlo method is often used to analyse the robustness of the model by varying the model parameters [50]. Figure 6 shows the basic principle of the Monte Carlo method used to show the variability of parameters. It reproduces the output result multiple times by randomly varying the input parameters [51]. These random

Table 3 Predictive model performance indicators

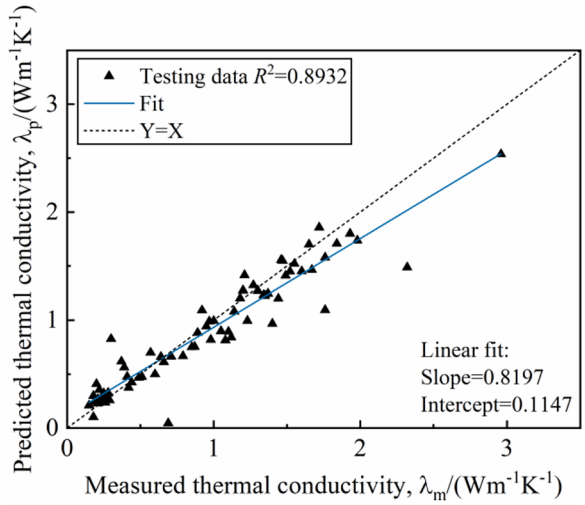
Model symbol	Data set	R^2	*RMSE*($\mathrm{Wm^{-1}\,K^{-1}}$)	*MAE*($\mathrm{Wm^{-1}\,K^{-1}}$)	*VAF*(%)
ANN	Training	0.9614	0.1032	0.0858	96.88
	Testing	0.9535	0.1338	0.0952	95.25
ANFIS	Training	0.9212	0.1773	0.1389	91.21
	Testing	0.8932	0.2105	0.1393	88.97
SVM	Training	0.9551	0.1263	0.1055	95.12
	Testing	0.9469	0.1333	0.0958	95.41



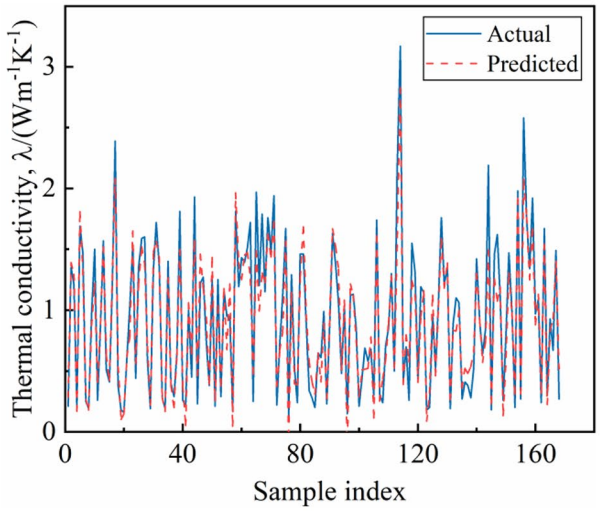
(a)



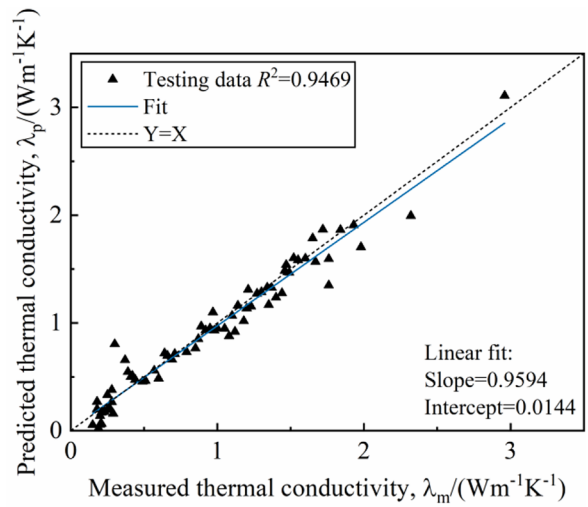
(a)



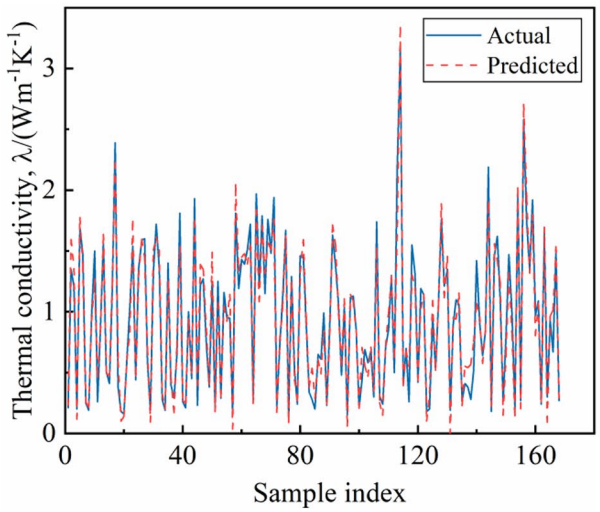
(b)



(b)



(c)



(c)

Fig. 3 Comparison of measured values and those calculated by the testing part of the ANN (**a**); ANFIS (**b**) and SVM (**c**)

◂**Fig. 4** Thermal conductivity of soil in function of sample index for the training part of ANN (**a**); ANFIS (**b**) and SVM (**c**); for the testing part of ANN (**d**); ANFIS (**e**) and SVM (**f**)

fluctuations propagate to the output solution, supplying different values of output quantities, and from them it is possible to quantitatively evaluate their probability density function.

The Monte Carlo method is widely used to evaluate the robustness of computational models because it can deal with irregular functions or complex models, especially ANNs. In this paper, the Monte Carlo method is used to propagate the input variability to the output. In order to study the robustness of the predictive models in calculating soil thermal conductivity λ, the results were statistically analysed. The optimal number of Monte Carlo simulation iteration was derived from the following equation [52]:

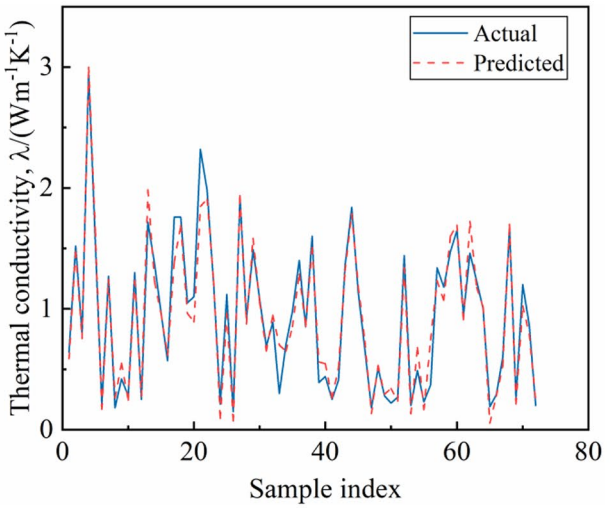
$$f = \frac{1}{m}\sum_{j=1}^{m} S_j \tag{8}$$

where m is the number of Monte Carlo iterations and S is the actual random variable considered. The optimal value of m represents the minimum number of iterations which is increased doesn't change any more the results. To evaluate the robustness of the models, 70% of the training data were randomly combined as a new input dataset, and 300 Monte Carlo simulations were generated. We then calculated the deviations between the calculated values of each model and the measured values. Each AI calculation model supplied 300 values of R^2, *RMSE* and *MAE*, so their average and variance and distribution can be calculated.

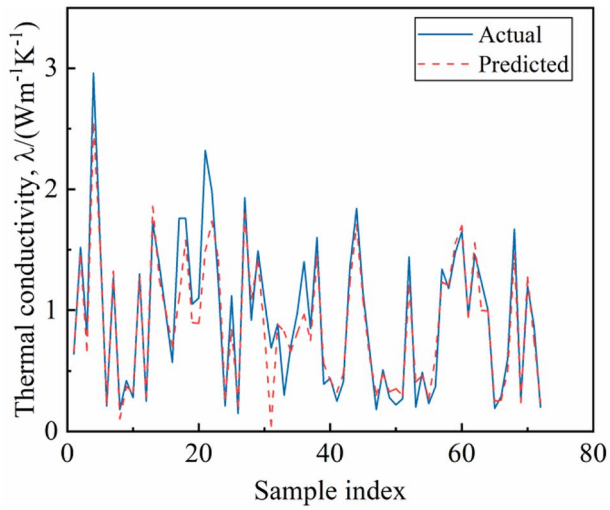
These probability distributions are shown in Fig. 7 while Table 4 show the percentiles of the distributions. Again it is confirmed that the SVM model presents the best robustness, followed by the ANN and ANFIS models.

6 Predictive model performance evaluation

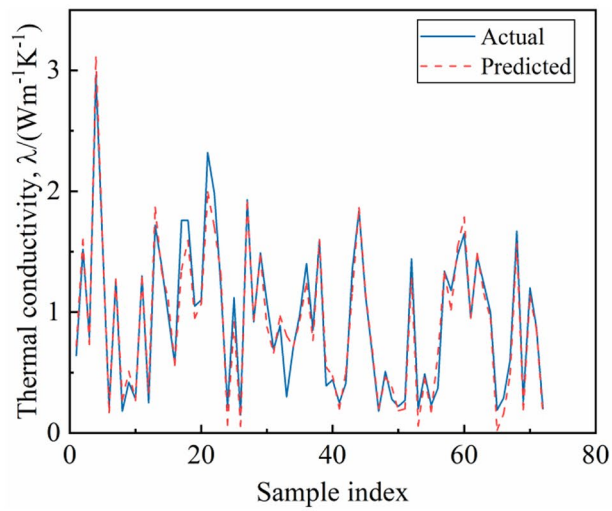
At present, the predictions of soil thermal conductivity is mainly based on theoretical and empirical models. Many parameters and requirements are involved in theoretical models of soil thermal conductivity, making them complicated. Empirical models calculate soil thermal conductivity, in a simpler way, so this method is commonly used in engineering design. In this paper the Johansen, Cote and Lu models are chosen and their results compared with those of the described AI models. The Johansen model [9] is an empirical relationship model based on the definition of normalized thermal conductivity. It can calculate the thermal conductivity of frozen and unfrozen soil as shown:



(d)



(e)



(f)

Fig. 4 (continued)

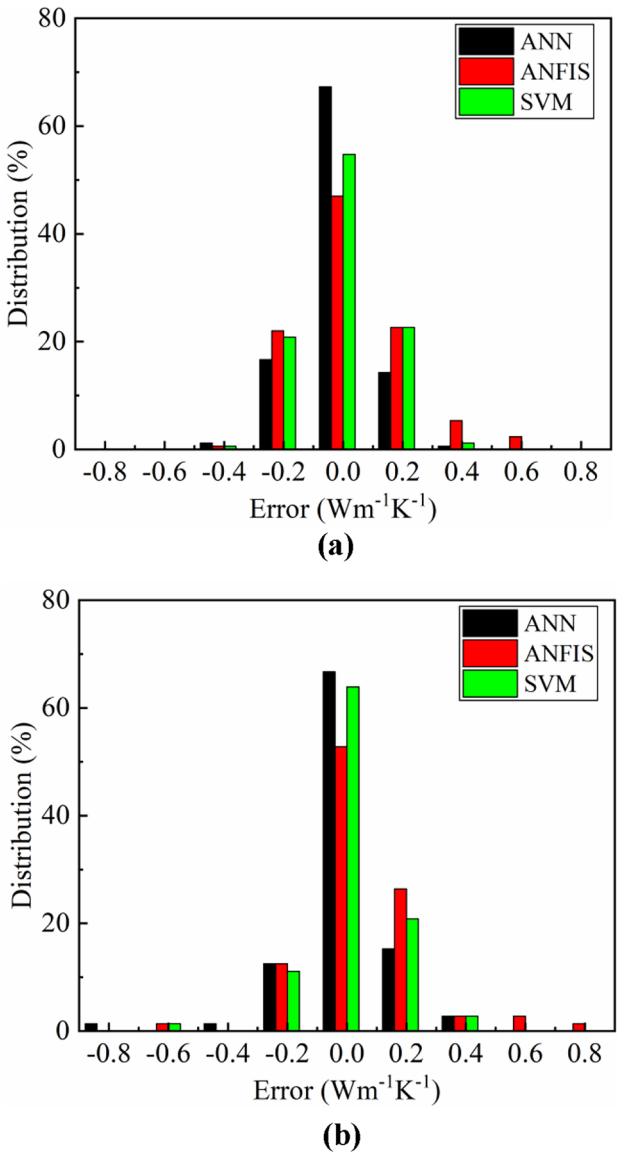


Fig. 5 Error distribution of measured − predicted thermal conductivity for (**a**) training data and (**b**) testing data (with a resolution of 0.2)

Table 4 Statistics of Monte Carlo simulation results

Criteria	Model	D25	D50	D75	Mean
R^2	ANN	0.8926	0.9512	0.9723	0.9132
	ANFIS	0.8731	0.8951	0.9056	0.8761
	SVM	0.9362	0.9459	0.9552	0.9449
RMSE	ANN	0.0838	0.1338	0.2518	0.1898
	ANFIS	0.1935	0.2105	0.2405	0.2285
	SVM	0.1093	0.1333	0.1623	0.1373
MAE	ANN	0.0432	0.0952	0.1832	0.1322
	ANFIS	0.1234	0.1394	0.1614	0.1464
	SVM	0.0839	0.0959	0.1109	0.0989

$$\lambda = (\lambda_{sat} - \lambda_{dry})K_e + \lambda_{dry} \tag{9}$$

where λ_{sat} and λ_{dry} are the thermal conductivities of saturated and dry soil, respectively, and K_e is the normalized thermal conductivity of the soil. When the soil saturation $S_r > 0.05$, K_e for coarse soils is given by:

$$K_e = 0.7 \log S_r + 1 \tag{10}$$

When $S_r > 0.1$, K_e again for coarse soils is:

$$K_e = \log S_r + 1 \tag{11}$$

Johansen improved the thermal conductivity model of De Vries [53] and proposed the following equation for dry soils:

$$\lambda_{dry} = \frac{0.137\ \rho_d + 64.7}{2650 - 0.947\ \rho_d} \tag{12}$$

where ρ_d is the dry density. In the calculation of the thermal conductivity of saturated soils, the following equation proposed by Sass [54] is widely used:

$$\lambda_{sat} = \lambda_{water}^{n} \lambda_{solid}^{1-n} \tag{13}$$

where λ_{water} and λ_{solid} are the thermal conductivities of water and soil solid particles, respectively, and n is porosity.

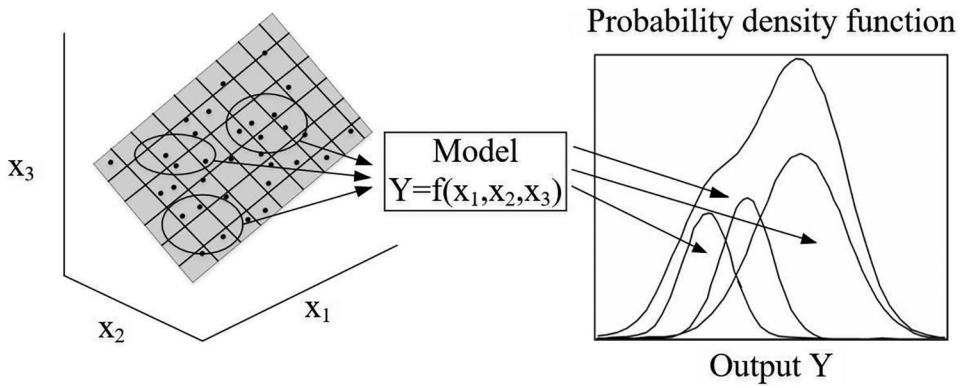
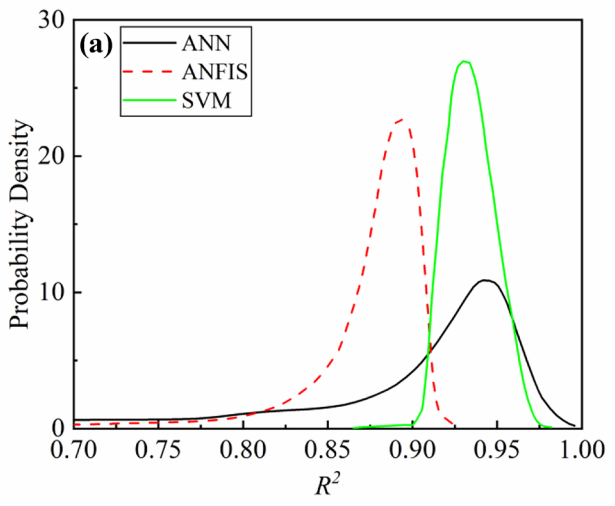
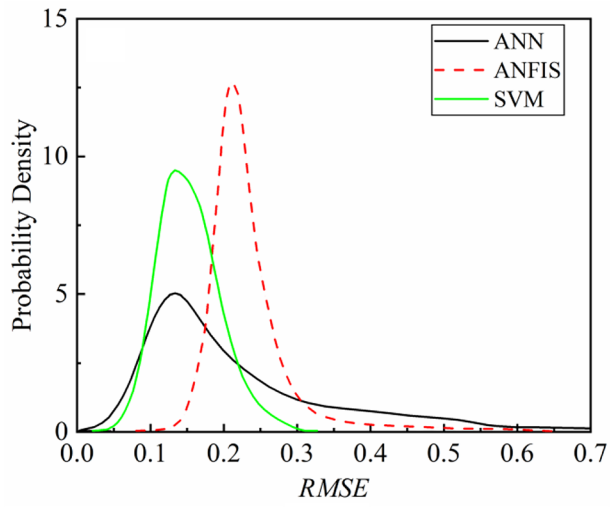


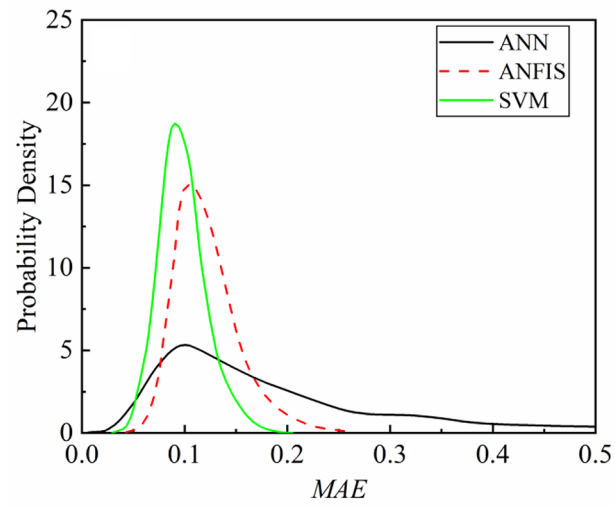
Fig. 6 Schematic diagram of Monte Carlo method basic principle



(a)



(b)



(c)

Fig. 7 Monte Carlo simulation test parameter probability distribution of (**a**) R^2; (**b**) *RMSE* and (**c**) *MAE*

The Cote model is based on the normalized thermal conductivity calculation model proposed by Cote and Konrad [8] according to Johansen [9], who established a generalized normalized prediction model:

$$\lambda = (\lambda_{\text{sat}} - \lambda_{\text{dry}})K_{\text{e}} + \lambda_{\text{dry}} \tag{14}$$

where K_{e} is the normalized thermal conductivity of the soil. Afterwards, the equations for K_{e} and λ_{dry} have been modified as follows:

$$K_{\text{e}} = \frac{\kappa S_r}{1 + (\kappa - 1)S_r} \tag{15}$$

$$\lambda_{\text{dry}} = \chi(10^{-\eta n}) \tag{16}$$

where κ is an empirical parameter used to express the influence of soil type and freezing state on K_{e}, and χ and η are parameters that describe the influence of the type and shape of granular soil on the thermal conductivity of dry soil.

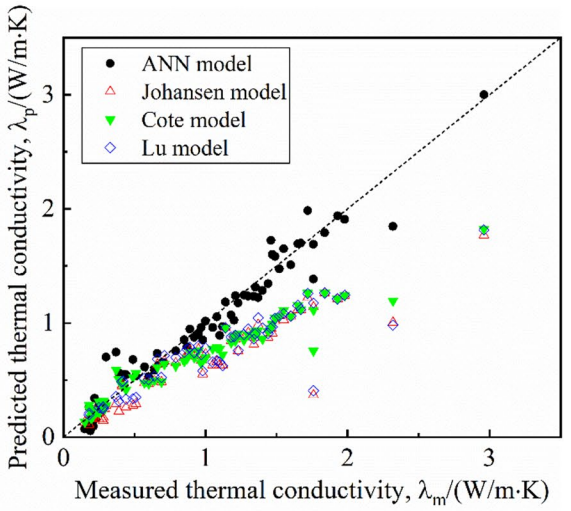
The model proposed by Lu et al. [55] was established on the base of a large number of thermal conductivity tests on 12 natural soils with different properties under various moisture contents. By fitting to the test data, a simpler thermal conductivity and porosity model of dry soils was obtained. Based on the normalized thermal conductivity model proposed by Johansen, the following empirical equation is proposed:

$$\lambda = \left[\lambda_{\text{water}}^{n}\lambda_{\text{solid}}^{1-n} - (b - an)\right]\exp[\alpha(1 - S_r^{\alpha - 1.33})] + (b - an) \tag{17}$$

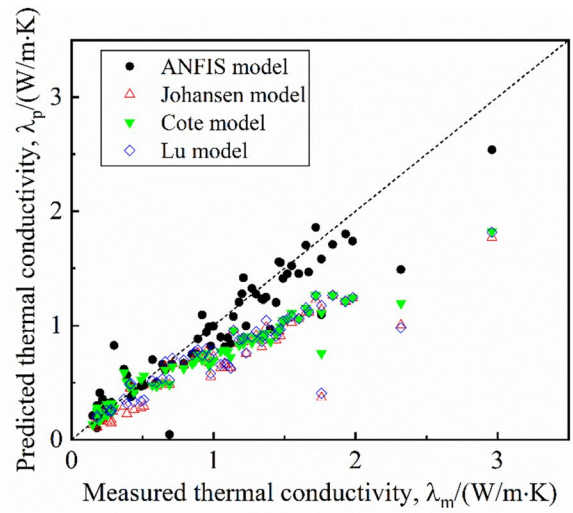
where a and b are parameters related to dry soil. It is recommended to take $a = 0.56$ and $b = 0.51$, respectively; α reflects the influence of soil type on the Kersten variable (0.96 and 0.27 for coarse-grained and fine-grained soils, respectively).

In the present paper the AI model results are compared with the λ values predicted by the Johansen, Cote, and Lu models, as shown in Fig. 8. The relative values are shown in Table 5. Figure 8 shows that the ANN, ANFIS and SVM models predict the thermal conductivity of the soils much better than the Johansen, Cote and Lu models. Even if with $\lambda < 1$ $\text{Wm}^{-1}\,\text{K}^{-1}$, the results of the Cote, Lu and Johansen models present small deviations from the experimental values, when $\lambda > 1$ $\text{Wm}^{-1}\,\text{K}^{-1}$, the predicted values are significantly lower than the measured ones, indicating that the prediction accuracy is low. Also Table 4 demonstrates that the performances of the empirical relationship models are significantly lower than those of the ANN, ANFIS and SVM models.

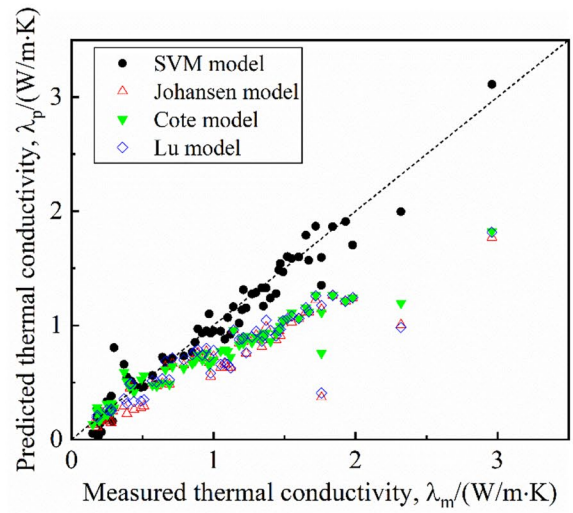
Again it is confirmed that the ANN model has the highest prediction accuracy, followed by SVM and ANFIS, while the empirical relationship models have lower accuracy. For the prediction of soil thermal conductivity, the ANN or SVM models are recommended.



(a)



(b)



(c)

◂**Fig. 8** Comparison of the forecasting models and the empirical relationship models (**a**) ANN, (**b**) ANFIS and (**c**) SVM

Table 5 Performance comparison of the intelligent computing and empirical relationship models

Model	R^2	$RMSE$($Wm^{-1}K^{-1}$)	MAE($Wm^{-1}K^{-1}$)	VAF(%)
ANN	0.9535	0.1338	0.0952	95.25
ANFIS	0.8932	0.2105	0.1393	88.97
SVM	0.9469	0.1333	0.0958	95.41
Johansen model	0.6464	0.4251	0.3631	75.05
Cote model	0.7744	0.3841	0.2805	78.34
Lu model	0.6432	0.4075	0.3335	71.49

7 Conclusions

This paper explored the prediction of soil thermal conductivity using ANN, ANFIS and SVM models. The following conclusions can be drawn:

(1) The factors influencing soil thermal conductivity were determined to be usesd as input parameters for predictive models. The most relevant are moisture content, dry density, clay content, powder content, sand content and quartz content.

(2) ANN, ANFIS and SVM models were tested to predict soil thermal conductivity, evaluating their R^2, $RMSE$, MAE and VAF. The results show that the three predictive models give results with high accuracy. Among them, the ANN model was most accurate, with $R^2 = 0.9535$, $RMSE = 0.1338$, $MAE = 0.0952$ and $VAF = 95.25\%$. The predictive accuracy of the SVM model was slightly lower, with $R^2 = 0.9469$, $RMSE = 0.1333$, $MAE = 0.0958$ and $VAF = 95.41\%$. The ANFIS model performed worst, with $R^2 = 0.8932$, $RMSE = 0.2105$, $MAE = 0.1393$ and $VAF = 88.97\%$. The robustness of the prediction models was analysed using Monte Carlo simulation and the SVM model was found to be the best performant.

(3) The prediction results of these three models were compared with those of traditional empirical relationship models and found to be significantly higher. The traditional models are only based on moisture content, dry density and soil type. So their estimates greatly deviate from measured values, making them less suited to be used in engineering design.

Funding Majority of the work presented in this paper was funded by the National Key R&D Program of China (Gran No. 2020YFC1807200), the National Natural Science Foundation of China (Grant No. 41877231, No. 42072299).

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Latifi N, Latifi S, Meehan CL, Abd Majid MZ, Tahir MM, Mohamad ET (2016) Improvement of problematic soils with biopolymer-an environmentally friendly soil stabilizer. J Mater Civ Eng 29(2):04016204. https://doi.org/10.1061/(ASCE)MT.1943-5533.0001706
2. Adam D, Markiewicz R (2009) Energy from earth-coupled structures, foundations, tunnels and sewers. Géotechnique 59(3):229–236. https://doi.org/10.1680/geot.2009.59.3.229
3. Dong Y, Mccartney JS, Lu N (2015) Critical review of thermal conductivity models for unsaturated soils. Geotech Geol Eng 33(2):207–221. https://doi.org/10.1007/s10706-015-9843-2
4. 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 94:14–23. https://doi.org/10.1016/j.icheatmasstransfer.2018.02.005
5. Xu X, Zhang W, Fan C, Li G (2019) Effects of temperature, dry density and water content on the thermal conductivity of genhe silty clay. Results Phys 16:102830. https://doi.org/10.1016/j.rinp.2019.102830
6. Lu N, Dong Y (2015) Closed-form equation for thermal conductivity of unsaturated soils at room temperature. J Geotech Geoenviron Eng 141(6):04015016. https://doi.org/10.1061/(ASCE)GT.1943-5606.0001295
7. Kersten MS (1949) Thermal Properties of Soils. University of Minnesota Engineering Experiment Station, Minneapolis (Bulletin No. 28)
8. Coté J, Konrad JM (2005) A generalized thermal conductivity model for soils and construction materials. Can Geotech J 42(2):443–458. https://doi.org/10.1139/t04-106
9. Johansen O (1975) Thermal Conductivity of Soils. University of Trondheim, Trondheim, Norway (Ph.D. thesis)
10. Erzin Y, Rao BH, Singh DN (2008) Artificial neural network models for predicting soil thermal resistivity. Int J Therm Sci 47(10):1347–1358. https://doi.org/10.1016/j.ijthermalsci.2007.11.001
11. Wang J, Zhang X, Du L (2017) A laboratory study of the correlation between the thermal conductivity and electrical resistivity of soil. J Appl Geophys 145:12–16. https://doi.org/10.1016/j.jappgeo.2017.07.009
12. Zhang N, Zou H, Zhang L, Puppala AJ, Cai G (2020) A unified soil thermal conductivity model based on artificial neural network. Int J Therm Sci 155:106414. https://doi.org/10.1016/j.ijthermalsci.2020.106414
13. Bi J, Zhang MY, Lai YM (2020) A generalized model for calculating the thermal conductivity of freezing soils based on soil components and frost heave. Int J Heat Mass Tran 150:119166. https://doi.org/10.1016/j.ijheatmasstransfer.2019.119166
14. Wang JM, He HL, Li M, Dyck M, Si B, Lv JL (2020) A review and evaluation of thermal conductivity models of saturated soils. Arch Agron Soil Sci. https://doi.org/10.1080/03650340.2020.1771315
15. Zhang T, Cai G, Liu S, Puppala AJ (2017) Investigation on thermal characteristics and prediction models of soils. Int J Heat Mass Tran 106:1074–1086. https://doi.org/10.1016/j.ijheatmasstransfer.2016.10.084
16. Salomone LA, Kovacs WD (1984) Thermal resistivity of soils. J Geotech Eng 110(3):375–389. https://doi.org/10.1061/(ASCE)0733-9410(1984)110:3(375)
17. Chen SX (2008) Thermal conductivity of sands. Heat Mass Transf 44(10):1241–1246. https://doi.org/10.1007/s00231-007-0357-1
18. Horai KI (1971) Thermal conductivity of rock-forming minerals. J Geophys Res Atmos 76(5):1278–1308. https://doi.org/10.1029/JB076i005p01278
19. Barry-Macaulay D, Bouazza A, Singh RM, Wang B, Ranjith PG (2013) Thermal conductivity of soils and rocks from the Melbourne (Australia) region. Eng Geol 164:131–138. https://doi.org/10.1016/j.enggeo.2013.06.014
20. Popov Y, Tertychnyi V, Romushkevich R, Korobkov D, Pohl J (2003) Interrelations between thermal conductivity and other physical properties of rocks: experimental data. Pure Appl Geophys 160(5–6):1137–1161. https://doi.org/10.1007/PL00012565
21. Anand J, Somerton WH, Gomaa E (1973) Predicting thermal conductivities of formations from other known properties. Soc Pet Eng J 13(5):267–272. https://doi.org/10.2118/4171-PA
22. Brigaud F, Chapman DS, Douaran SL (1990) Estimating thermal conductivity in sedimentary basins using lithologic data and geophysical well logs. AAPG Bull 74(9):1459–1477. https://doi.org/10.1306/0c9b2501-1710-11d7-8645000102c1865d
23. Midttomme K, Roaldset E (1998) The effect of grain size on thermal conductivity of quartz sands and silts. Petrol Geosci 4(2):165–172. https://doi.org/10.1144/petgeo.4.2.165
24. Haigh SK (2012) Thermal conductivity of sands. Geotechnique 62(7):617–625. https://doi.org/10.1680/geot.11.P.043
25. Mitchell JK, Soga K (2005) Fundamentals of Soil Behavior. John Wiley & Sons, New York
26. Fausett LV (1994) Fundamentals of neural networks: architectures, algorithms, and applications. Pearson, Prentice-Hall
27. Hagan MT, Demuth HB, Beale MH (2002) Neural network design. China Machine Press, Beijing
28. Levenberg K (1944) A method for the solution of certain problems in least squares. Q Appl Math 2(2):164–168
29. Møller MF (1993) A scaled conjugate gradient algorithm for fast supervised learning. Neural Netw 6:525–533. https://doi.org/10.1016/S0893-6080(05)80056-5
30. Burden F, Winkler D (2009) Bayesian regularization of neural networks. Methods Mol Biol
31. Orbanic P, Fajdiga M (2003) A neural network approach to describing the fretting fatigue in aluminium-steel couplings. Int J Fatigue 25(3):201–207. https://doi.org/10.1016/S0142-1123(02)00113-5
32. Nguyen TA, Ly HB, Jaafari A, Pham TB (2020) Estimation of friction capacity of driven piles in clay using artificial neural network. Vietnam J Earth Sci 42(3):265–275. https://doi.org/10.15625/0866-7187/42/3/15182
33. Boubou R, Emeriault F, Kastner R (2010) Artificial neural network application for the prediction of ground surface movements induced by shield tunnelling. Can Geotech J 47(11):1214–1233. https://doi.org/10.1139/T10-023
34. Fei W, Narsilio GA, Disfani MM (2021) Predicting effective thermal conductivity in sands using an artificial neural network with multiscale microstructural parameters. Int J Heat Mass Tran 170:120997. https://doi.org/10.1016/j.ijheatmasstransfer.2021.120997

35. Shrestha D, Wuttke F (2020) Predicting the effective thermal conductivity of geo-materials using artificial neural networks. E3S Web Conf 205:04001. https://doi.org/10.1051/e3sconf/202020504001
36. Rizvi ZH, Akhtar SJ, Sabeeh WT, Wuttke F (2020) Effective thermal conductivity of unsaturated soils based on deep learning algorithm. E3S Web Conf 205:04006
37. Liu C, Hu X, Yao R, Han Y, Wang Y, He W, Fan H, Du L (2020) Assessment of Soil Thermal Conductivity Based on BPNN Optimized by Genetic Algorithm. Adv Civ Eng 2020(3):1–10. https://doi.org/10.1155/2020/6631666
38. Rizvi Z, Zaidi H, Shoarian Sattari A, Wuttke F (2019) Effective thermal conductivity of unsaturated sand using artificial neural network (ANN) and lattice element method (LEM). Int J Therm Sci
39. Chayjan RA, Montazer GA, Hashjin TT, Hadi M, Ghobadian B (2007) Prediction of Pistachio Thermal Conductivity Using Artificial Neural Network Approach. Int J Agric Biol 9(6):816–820. http://www.fspublishers.org
40. Fattahi H (2016) Indirect estimation of deformation modulus of an in situ rock mass: an anfis model based on grid partitioning, fuzzy c-means clustering and subtractive clustering. Geosci J 20(5):681–690. https://doi.org/10.1007/s12303-015-0065-7
41. Vapnik V (1999) The Nature of Statistical Learning Theory, 2nd edn. Springer, New York
42. Bui DT, Tran AT, Klempe H, Pradhan B, Revhaug I (2016) Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides 13(2):361–378. https://doi.org/10.1007/s10346-015-0557-6
43. Tarnawski VR, McCombie ML, Leong WH, Wagner B, Momose T, Schönenberger J (2012) Canadian Field Soils II Modeling of Quartz Occurrence. Int J Thermophys 33(5):843–863. https://doi.org/10.1007/s10765-012-1184-2
44. Tarnawski VR, Momose T, Mccombie ML, Leong WH (2014) Canadian field soils iii thermal-conductivity data and modeling. Int J Thermophys 36(1):119–156. https://doi.org/10.1007/s10765-014-1793-z
45. Tarnawski VR, Leong WH (2016) Advanced Geometric Mean Model for Predicting Thermal Conductivity of Unsaturated Soils. Int J Thermophys 37(2):18. https://doi.org/10.1007/s10765-015-2024-y
46. Lin P, Ni P, Guo C, Mei G (2020) Mapping soil nail loads using Federal Highway Administration (FHWA) simplified models and artificial neural network technique. Can Geotech J 57(10):1453–1471. https://doi.org/10.1139/cgj-2019-0440
47. Gokceoglu C (2002) A fuzzy triangular chart to predict the uniaxial compressive strength of the ankara agglomerates from their petrographic composition. Eng Geol 66(1–2):39–51. https://doi.org/10.1016/S0013-7952(02)00023-6
48. Yukselen Y, Erzin Y (2008) Artificial neural networks approach for zeta potential of montmorillonite in the presence of different cations. Environ Geol 54(5):1059–1066. https://doi.org/10.1007/s00254-007-0872-x
49. Willmott CJ, Matsuura K (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res 30:79–82. https://doi.org/10.3354/cr030079
50. Pham BT, Le LM, Le TT, Bui KTT, Prakash I (2020) Development of advanced artificial intelligence models for daily rainfall prediction. Atmos Res 237:104845. https://doi.org/10.1016/j.atmosres.2020.104845
51. Christian Pds (2012) Stochastic Models of Uncertainties in Computational Mechanics. Amer Society of Civil Engineers, Reston, VA
52. Guilleminot J, Soize C (2013) Stochastic model and generator for random fields with symmetry properties: application to the mesoscopic modelling of elastic random media. Multiscale Model Sim 11:840–870. https://doi.org/10.1137/120898346
53. De Vries DA (1963) Thermal properties of soils. In: Van Wijk WR (ed) Physics of the Plant Environment. John Wiley & Sons, New York, pp 210–235
54. Sass JH, Lachenbruch AH, Munroe RJ (1971) Thermal conductivity of rocks from measurements on fragments and its application to heat-flow determinations. J Geophys Res 76(14):3391–3401. https://doi.org/10.1029/JB076i014p03391
55. 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(1):8–14. https://doi.org/10.2136/sssaj2006.0041