In-situ soil texture classification and physical clay content measurement based on multi-source information fusion

Chao Meng, Wei Yang*, Xinjian Ren, Dong Wang, Minzan Li

(*Key Lab of Smart Agriculture Systems, Ministry of Education, China Agricultural University, Beijing 100083, China*)

Abstract: Soil texture is one of the most important soil characteristics that affect soil properties. Rapid acquisition of soil texture information is of great significance for accurate farmland management. Traditional soil texture analysis methods are relatively complicated and cannot meet the requirements of temporal and spatial resolution. This research introduced a self-developed vehicle-mounted in-situ soil texture detection system, which can predict the type of soil texture and the particle composition of the texture, and obtain real-time data during the measurement process without preprocessing the soil samples. The detection system is mainly composed of a conductivity measuring device, a camera, an auxiliary mechanical structure, and a control system. The soil electrical conductivity (ECa) and the texture features extracted from the surface image were input into the embedded model to realize real-time texture analysis. In order to find the best model suitable for the detection system, measurements were carried out in three test fields in Northeast and North China to compare the performance of different models applied to the detection system. The results showed that for soil texture classification, ExtraTrees performed best, with Precision, Recall, and F1 all being 0.82. For particle content of soil texture prediction, the R^2 of ExtraTrees was 0.77, and RMSE and MAPE were 74.72 and 39.58. It was observed that ECa, Moment of inertia, and Entropy had larger weights in the drawn model influence weight map, and they are the main contributors to predicting soil texture. These results showed the potential of the vehicle-mounted in-situ soil texture detection system, which can provide a basis for fast, cost-effective, and efficient soil texture analysis.
Keywords: soil texture, soil sensor, electrical conductivity, soil surface image
DOI: 10.25165/j.ijabe.20231601.6918

1　Introduction

Soil texture is one of the important physical properties of soil. It represents the percentage combination of soil particles of different diameters in the weight of the soil. It is generally divided into three types: sand, loam, and clay. Soil texture affects many dynamic physical properties, such as electrical conductivity (ECa), organic carbon, cation exchange capacity, and so on[1]. Soils with different textures contain different minerals. For example, soils with more clay have more iron oxides and kaolinite[2]. Soil texture also affects the root tensile strength of some plants (for example, *Spartina patens*)[3]. Good-textured soil can well regulate plants' requirements for water, nutrients, air, and temperature during crop growth, thereby promoting high crop yields. For example, peanuts are important crops, and a coarser soil texture is needed to achieve better pod development and peanut growth[4].

The standard method for obtaining soil texture is the hydrometer method[5]. Although this method can provide very accurate soil texture results, it has obvious disadvantages: it cannot obtain a large data set at one time; it takes a lot of time to dry and grind samples. The soil samples need to be heated and treated with H_2O_2 and Na_3PO_4. The whole experiment process needs manual operation to ensure accuracy and safety; measurement accuracy depends on the experimental conditions and the operator's proficiency. Many new methods of soil texture acquisition have been studied by scholars around the world, but there are some problems. For example, although the laser particle size analyzer method based on the principle of particle diffraction[6] can generate high-resolution mechanical composition detail reports, it also has limitations because of the higher cost. In addition, there are also studies using the γ-ray method, scanning electron microscopy, vis-NIR technology, Diffuse Revelations Infrared Fourier Transformations Spectroscopy (DRIFT), and remote sensing technology to predict soil texture[7-9], but expensive instruments or early-stage required pretreatment limits their usefulness. In summary, the current research has time lag in obtaining soil texture, the spatial resolution is difficult to meet the demand and most of them require expensive equipment. Therefore, it is necessary to develop a low-cost, real-time, high-precision soil texture information acquisition method and detection system to meet the development needs of smart agriculture.

Images have been widely used in soil texture research in recent years. Swetha et al.[10] used customized darkroom and smartphone images to predict soil texture. After soil samples were dried, ground, and sieved in the laboratory, local features, color features, and texture features were extracted from the images. The results show that the prediction accuracy of clay particles ($R^2>0.97$) and sand particles ($R^2>0.96$) is high, while the prediction accuracy of powder particles ($R^2=0.62$-0.75) is medium. de Oliveira Morais et al.[11] used computer digital scanned images and used three multivariate correction methods Partial Least Square

(PLS), Successive Projections Algorithm- Multiple Linear Regression (SPA-MLR), and Least Squares Support Vector Machines Regression (LSSVMR) to predict the sand and clay content in the pre-treated soil samples, of which LSSVMR performed the best (R^2>0.9). de Oliveira Morais et al.[12] used digital image processing (image segmentation) and multivariate image analysis (MIA) of soil samples to predict soil texture. In the range of 25% to 75% of sand and clay, there was a strong correlation between the predicted and measured sand and clay contents (R^2>0.92). Sudarsan et al.[13] developed a set of image acquisition systems by using a small and cheap hand-held microscope, and estimated the soil texture and soil organic matter by using various calculation parameters of the obtained images. The predicted performance of sand was better (R^2=0.63). Although the above studies all need pretreatments such as air drying and sieving, it provides a basis for in-situ measurement of soil texture using images.

Soil ECa has been proven to have a strong correlation with texture for many years. Garcia-Tomillo et al.[14] found that clay, silt, and sand contents were significantly correlated with ECa (r=0.48, 0.24, and −0.36, p<0.05). Therefore, ECa is used as an auxiliary variable to interpolate texture maps by regression Kriging. Soil texture and ECa showed strong spatial dependence, and ECa and soil quality maps showed similar spatial distribution patterns. Andrade et al.[15] used portable X-ray Fluorescence spectrometry (pXRF) and Magnetic Susceptibility (MS) to analyze soil texture. Among the three particle size components, the prediction of clay and sand content was the most accurate in this study. The R^2 of the best model for sand, silt, and clay were 0.79, 0.44, and 0.71, respectively. The above results show that ECa can be used as an input to indirectly predict soil texture, and the results show that ECa has a good performance in predicting soil texture.

Therefore, in order to obtain the field soil texture information in real time and provide the basis for accurate management, this study introduces a self-developed vehicle-mounted in-situ soil texture detection system based on ECa and soil surface image information. According to the soil texture classification and soil texture particle composition, the best model is selected by using the field test data.

2 Materials and methods

2.1 Soil samples and experimental preparation

There are three experimental sites, Erdaohe Farm, Heilongjiang, China (Farmland A); Shangzhuang, Beijing, China (Farmland B) and Tongzhou, Beijing, China (Farmland C).

The geographical coordinates of Heilongjiang Province are 123.02°E-137.3°E and 43.7°N-53.9°N, with a total area of 473 000 km², and an average annual temperature of −4°C-5°C, annual precipitation 400-650 mm. It is mainly black soil with high fertility and fine texture.

The geographical coordinates of Beijing are 115.7°E-117.4°E, 39.4°N-41.6°N, covering an area of 16 410 km², with an annual average temperature of 11.5°C. The annual precipitation is about 540.7 mm. Most of the soil is loess and brown gray soil, which is fertile and moderate in texture.

The distribution of the experimental area, path, and sampling area is shown in Figure 1, and farmland information is shown in Table 1. All the experiments were carried out in the fallow period, in which farmland A is 1750 km² of a paddy field, farmland B is 2000 km² of a maize field, farmland C is 2500 km² of a maize field,

and the sampling plot is 3 m×5 m. The content of the experiment includes data collection of the detection system and soil sample collection. A total of 221 soil samples were collected in the experimental path, including 36 in farmland A, 95 in farmland B, and 90 in farmland C. Each soil sample is 1 kg, and the topsoil is collected at a depth of 5-10 cm. The standard values of ECa and soil texture were measured by quartering each soil sample equally. The standard values were matched with the measured values by GPS data.
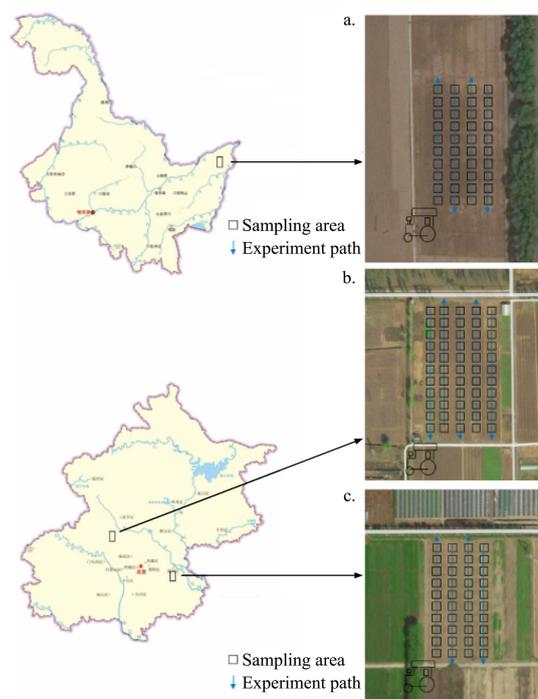
a. Erdaohe Farm, Heilongjiang, China   b. Shangzhuang, Beijing, China
c. Tongzhou, Beijing, China
Figure 1   Test site, path, and sampling point

Table 1   Experimental farmland information

Farmland	Location	Planting type	Area/km²	Samples
A	Heilongjiang	Rice	1750	36
B	Beijing	Maize	2000	95
C	Beijing	Maize	2500	90

The measurement method of ECa standard value is 5:1 static extraction method[16]. The soil samples were passed through a 1 mm sieve and dried in an oven for 24 h. Add 10 g soil sample and 50 mL deionized water into the flask. Put the cap on the bottle and put it on the reciprocating horizontal constant temperature oscillator to oscillate for 30 min. After oscillation, let it stand at room temperature (about 20°C) for 24 h. Take a proper amount of transparent solution, use ECa instrument (DDSJ-319L, Shanghai Yidian Scientific Instrument Co., Ltd., China) to measure, measure three times, take the average value, and record the data as the standard ECa value of soil samples.

The method of measuring soil texture is drying and sieving soil samples, using a laser particle size analyzer (NKT5200-h, Shandong Nijite Analytical Instrument Co., Ltd., China) for wet measurement. After obtaining the analysis report, soil samples were classified according to Kaczynski soil classification standard[17] in Table 2, and three soil texture types were measured: sandy loam (81 samples), light loam (99 samples), and medium loam (41 samples). The statistics of sample texture types of the three test sites are listed in Table 3.

Table 2   Kaczynski's standard for classification of soil texture

Soil Texture		Physical clay (<0.01mm) content			Physical sand (>0.01mm) content		
		Podzol	Grassland soil, red and yellow soil	Columnar alkaline soil, strong alkaline soil	Podzol	Grassland soil, red and yellow soil	Columnar alkaline soil, strong alkaline soil
Sand	Loose sand	0-5	0-5	0-5	100-95	100-95	100-90
	Tight sand	5-10	5-10	5-10	95-90	95-90	95-90
Loam	Sandy loam	10-20	10-20	10-15	90-80	90-80	90-85
	Light loam	20-30	20-30	15-20	80-70	80-70	85-80
	Medium loam	30-40	30-45	20-30	70-60	70-55	80-70
	Heavy loam	40-50	45-60	30-40	60-50	55-40	70-60
Clay	Light clay	50-65	60-75	40-50	50-30	40-25	60-50
	Medium clay	65-80	75-85	50-65	35-20	25-15	50-35
	Heavy clay	>80	>85	>65	<20	<15	<35

Table 3   Texture types of samples from three test sites

Test site	Sandy loam sample	Light loam sample	Medium loam sample	Total sample
Farmland A	18	2	16	36
Farmland B	4	66	25	95
Farmland C	59	31	0	90

2.2   Principle of soil texture measurement

Because it is difficult to measure soil texture directly, this study proposed a method to predict soil texture by using parameters highly related to soil texture.   The feasibility of indirect prediction of soil texture was analyzed in this study, in which image and ECa performed well in the prediction of soil texture.   ECa and soil surface images were selected to predict soil texture.   In the field experiment, through the self-designed ECa sensor and industrial camera, the required input parameters are obtained: ECa and the image features extracted from the soil surface image.   Finally, the soil texture information of the detection target is obtained by calling the prediction model.   The overall design scheme of the vehicle-mounted in-situ soil texture detection system is shown in Figure 2.
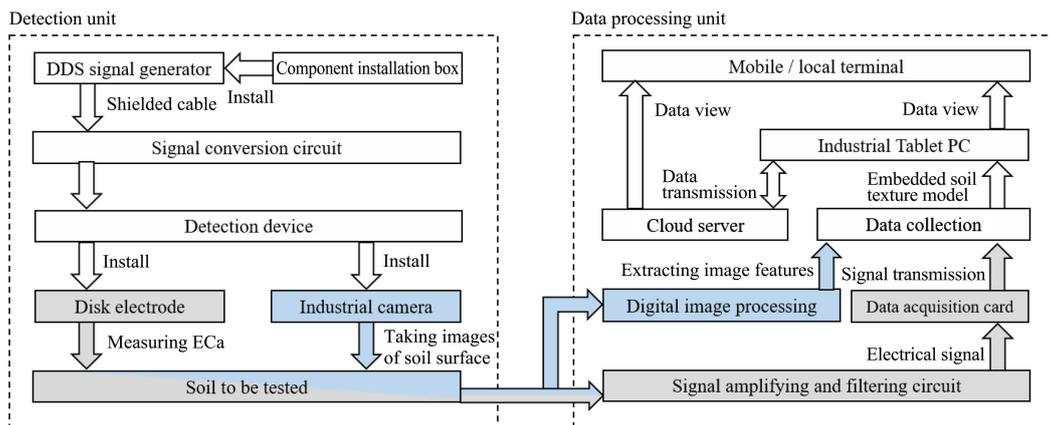
Note: Blue block diagram: ECa measurement process; Grey block diagram: soil surface image acquisition process; DDS: Direct Digital Synthesizer; PC: Personal Computer.

Figure 2   General design scheme of detection system

The most important part of the detection system is the data collection of two kinds of sensing devices: disc ECa electrode and industrial camera.   Four disk electrodes were used to measure ECa. Industrial cameras were used to obtain images of soil surfaces. Among them, industrial cameras and industrial tablets were directly connected through USB cables.   In order to achieve real-time and fast data measurement, a high-speed data acquisition card was selected to ensure the mobility and accuracy of data, and a GPS receiver was installed to record the location information.

2.3   ECa measurement principle and image analysis method

The method of measuring ECa is the four-colter Wenner array method[18], as shown in Figure 3 and Equation (1).   This method has been proven to be stable and accurate.

When $JM=KN=a-b/2$, $MN=b$,

$$\sigma = \frac{1}{\pi\left(\dfrac{a^2}{b} - \dfrac{b}{4}\right)} \frac{1}{\Delta V_{MN}} \tag{1}$$

where, σ is the soil ECa value calculated by the four-terminal method, μS/cm; a and b are the distance between probes, cm; I is the constant current source current, A; ΔV_{MN} is the M; N voltage between two probes, V.
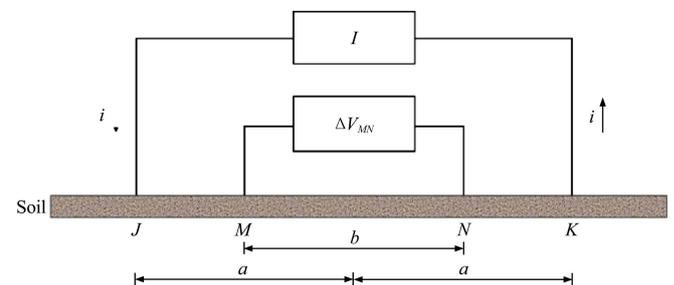
Note: i: Current; a and b: distance between electrodes.

Figure 3   Current-voltage four-electrode method

The method of extracting soil surface texture features is Gray-Level Co-occurrence Matrix (GLCM)[19,20].   GLCM is a classical second-order statistical algorithm.   In 1973, Haralick[21] proposed using GLCM to describe texture features.   This is because the texture is formed repeatedly and alternately by the gray distribution in the spatial position, so there must be a certain distance between the two pixels in the image space.   A certain gray-level relationship is called the spatial correlation of gray

levels in an image. The texture is described by studying the spatial correlation of gray levels. This is the basis of GLCM, which is composed of the joint probability density of the gray level. It can reflect the direction of the gray level of the image, the adjacent interval, and the change range of comprehensive information. It is the basis of analyzing the local pattern of an image and its arrangement rules. Based on the matrix, various statistics can be calculated: Energy, Entropy, Contrast, Uniformity, Correlation, Variance, Sum, etc. The common parameters and their calculation equations are as follows[22,23]:

1) Correlation

Correlation refers to the degree of correlation between related pixels and their adjacent pixels, which reflects the local gray correlation in the image.

$$Correlation = \sum_i \sum_j \frac{P(i,j)}{1+(i-j)^2} \qquad (2)$$

where, $P(\cdot)$ refers to the input image; (i, j) refers to the pixel coordinates, the value range of i and j is usually 0-255.

2) Energy

Energy measures the uniformity of image texture and represents the repeated information of pixel pairs.

$$Energy = \sum_i \sum_j P(i,j)^2 \qquad (3)$$

3) Entropy

Entropy in physics means the degree of regularity of an object. The more ordered, the smaller Entropy, and the more disordered it is, the larger Entropy. It represents the randomness and complexity of texture feature distribution.

$$Entropy = \sum_i \sum_j P(i,j) \log P(i,j) \qquad (4)$$

According to the research of predecessors[24,25] in our laboratory, energy, uniformity, Entropy, energy, and moment of inertia are highly correlated with soil roughness, soil bulk density, root mean square height and correlation length. By comparing the correlation between the root mean square height and the correlation length of soil surface roughness and the 12 texture feature parameters, the four texture features used in this study have the highest R^2. Therefore, this study used four image parameters: Energy, Entropy, Moment of inertia, and Correlation to predict soil texture.

2.4 Design of vehicle mounted in situ soil texture detection system

According to the principle of soil texture prediction based on ECa and soil surface image proposed in this study, the vehicle-borne soil texture detection system was developed. The structure diagram and the physical object are shown in Figures 4 and 5. The detection system mainly includes the detection unit, mechanical structure, and electronic control system.

The detection unit is the core of vehicle mounted in situ soil texture detection system, including the ECa measurement device and surface image acquisition device. The ECa measuring device consists of disc electrode, DDS signal generator, high-speed signal acquisition card, and signal processing circuit. As a constant current source, the signal generator provides the initial signal, which contacts the soil to be measured through four disc electrodes. After amplification, filtering, conversion, and other processing, the returned signal is collected by the signal acquisition card. The surface image acquisition device consists of an industrial camera (MV-SUA1201C-T, Shenzhen Mindvision Technology Co., Ltd., Shenzhen, China) and a tablet computer. After the industrial camera takes the soil surface image, the texture features are

extracted through image processing such as segmentation. Especially, due to the large measurement range of ECa in practice, the center line of the disk electrode is regarded as the measurement area, the camera is installed above the center line, and the area where the camera takes pictures and the center line overlap is regarded as the measurement point of the detection system.

In order to ensure the normal operation of the detection system in farmland, in addition to the core detection unit, it also includes auxiliary mechanical system and electronic control system. The auxiliary mechanical system provides platform support for the whole detection system, which is mainly composed of three-point suspension structure (connecting tractor), electronic components installation box, load support platform, etc. The electronic control system realizes the acquisition and processing of ECa measurement signals and the display of measurement results. The control part mainly consists of industrial plate, circuit processing module, GPS module, and so on. The ECa measurement and image characteristic values collected by industrial flat plate are input into the embedded soil texture prediction model to obtain soil texture information. The display part includes the local display on the industrial tablet and the display and data management on the mobile phone.
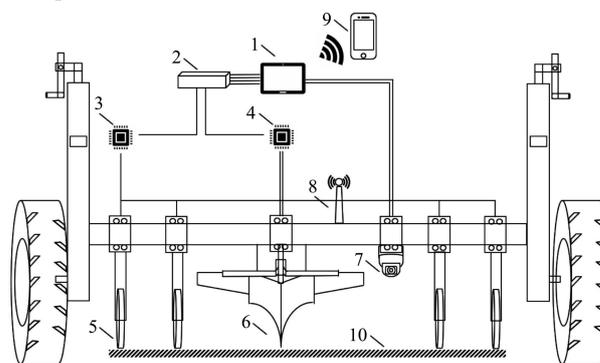
1. Industrial tablet computer　2. Data acquisition card　3. ECa processing circuit 4. Resistance processing circuit　5 Disc electrode　6. Deep loose plough 7. Industrial camera　8. GPS locator　9. Mobile phone　10. Centerline
Figure 4　Detection system structure diagram

Figure 5　Physical image of detection system

2.5 Model and evaluation method

In order to find the best model suitable for detection system embedding, six classic models were selected: AdaBoost, Light Gradient Boosting Machine (LightGBM), SVM, BP curve network, Random Forest (RF), Extremely Randomized trees (Extratrees).

AdaBoost[26] is an iterative algorithm. Its core idea is to train different prediction models (weak models) for the same training set, and then combine these weak models based on the error of the previous model to form a stronger model.

LigthGBM[27] uses the negative gradient of the loss function as the approximate residual value of the current decision tree to fit the new decision tree.

The basic idea of SVM[28] is to define the optimal linear

hyperplane and reduce the algorithm of the optimal linear hyperplane to a convex optimization problem.

BP neural network[29] is a kind of multilayer feedforward neural network trained according to the error back propagation algorithm, which has the ability of arbitrary complex pattern classification and excellent multi-dimensional function mapping.

RF[30] is a classifier with multiple decision trees, and its output categories are determined by the mode of the categories output by individual trees.

Each decision tree of ExtraTrees[31] uses all the same training samples, and the bifurcation value is obtained completely at random, so as to achieve bifurcation of the decision tree

The evaluation methods of the classification model are Precision, Recall, and the harmonic mean F1-score (F1). These three indicators are all methods to evaluate the prediction accuracy, and F1 is the weighted average of Precision and Recall. The regression model was evaluated by R^2, Root Mean Square Error (EMSE), and Mean Absolute Percentage Error (MAPE). The proportion of the model training set was 0.75, the number of training set samples of sandy loam, light loam, and medium loam were 60, 74, and 34, respectively, and the number of verification set samples was 21, 25, and 10, respectively.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

$$F1 = 2\frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

where, TP is the number of positive classes predicted as positive classes; TN is the number of negative classes predicted as negative classes; FP is the number of negative classes predicted as positive classes; FN is the number of positive classes predicted as negative classes.

3   Results and discussion

3.1   Descriptive statistics and soil textural variations

The descriptive statistics of physical clay (<0.01mm) content of Sandy loam, Light loam, and Medium loam show high variability, as listed in Table 4.   The physical clay (<0.01 mm) content of Sandy loam is between 10.09 μm and 19.89 μm, with an average value of 15.10 μm and an SD of 2.22.   The physical clay (<0.01 mm) content of Light loam also showed a wide range (13.57-29.99 μm), with an average value of 24.84 μm and an SD of 4.01, indicating a greater degree of dispersion.   The physical clay (<0.01 mm) content of Medium loam is between 30.05 and 42.08, with an average value of 34.84 and an SD of 3.5.   Light loam and Medium loam showed moderate skewness (−0.674 and 0.417).   These three kinds of soil separations all show a relatively flat non-normal distribution with negative kurtosis, which proves that the use of non-parametric regression such as RF is reasonable.

Table 4   Descriptive statistics of physical clay content of three germplasm types

Parameters	Mean/μm	Min/μm	Max/μm	SD	Kurtosis	Skewness	Total
Sandy loam	15.103	10.090	19.890	2.217	−0.844	0.075	81
Light loam	24.844	13.580	29.990	4.008	−0.397	−0.674	99
Medium loam	34.837	30.050	42.080	3.503	−0.777	0.417	41

Table 5 summarizes the descriptive statistics of input parameters used to predict soil texture information.   In order to display the mean values more directly, Figure 6 shows the mean values of a texture parameter of three texture types in four different directions.   It can be seen that with the increase of physical clay (<0.01 mm) content, the mean values of energy and correlation show an upward trend, while the mean values of Entropy and moment of inertia decrease with the increase of physical clay content.   The important thing is that with the change of angle, the four texture parameters basically maintain a similar trend.   The SD of Energy, Entropy, and Correlation were between 0.01 and 0.31, while the dispersion of moment of inertia was obvious (SD was 0.28-3.06).   The energy, moment of inertia, and correlation of the three germplasm sites all had positive skewness values, and the dispersion degree on the right side was large, showing a similar trend.   Entropy has positive bias and negative bias, but the difference is not significant.   Generally speaking, although there are differences in the degree of dispersion among individuals in the group, the difference in the mean value is very obvious, and the similar trend provides support for the classification.

Table 5   Descriptive statistics of input parameters

Parameters	Direction	Sandy loam			Light loam			Medium loam		
		Mean	SD	Skewness	Mean	SD	Skewness	Mean	SD	Skewness
Energy	0°	0.024	0.016	2.091	0.039	0.017	1.324	0.042	0.013	0.909
	45°	0.020	0.014	2.320	0.035	0.030	6.537	0.035	0.011	0.724
	90°	0.025	0.016	2.374	0.040	0.018	1.467	0.043	0.013	0.678
	135°	0.020	0.013	2.330	0.032	0.015	1.937	0.034	0.011	0.752
Entropy	0°	4.300	0.442	−1.133	3.685	0.468	0.754	3.557	0.292	−0.460
	45°	4.496	0.440	−1.453	3.878	0.469	0.590	3.735	0.310	−0.265
	90°	4.280	0.422	−1.511	3.663	0.458	0.676	3.534	0.296	−0.178
	135°	4.486	0.425	−1.478	3.882	0.462	0.615	3.756	0.309	−0.358
Moment of inertia	0°	3.500	2.108	0.148	1.277	1.597	2.210	0.646	0.284	0.572
	45°	5.435	3.056	0.080	1.941	2.303	2.065	0.973	0.487	0.874
	90°	3.180	1.715	0.115	1.152	1.316	2.087	0.625	0.305	0.721
	135°	5.212	2.903	0.123	1.949	2.343	2.150	1.013	0.478	0.482
Correlation	0°	0.066	0.027	4.684	0.113	0.068	3.999	0.115	0.040	1.889
	45°	0.062	0.028	4.425	0.113	0.096	6.617	0.113	0.040	1.861
	90°	0.067	0.027	4.851	0.108	0.047	0.846	0.115	0.040	1.873
	135°	0.062	0.028	4.518	0.105	0.047	0.778	0.112	0.040	1.844
ECa		316.369	100.570	0.522	229.838	70.173	1.578	247.080	60.825	0.208

Note: ECa: Soil electrical conductivity.

× Sandy loam   × Light loam   × Middle loam

a. Mean value of energy of three texture types

b. Mean value of entropy of three texture types

c. Mean value of moment of inertia of three texture types

d. Mean value of correlation of three texture types
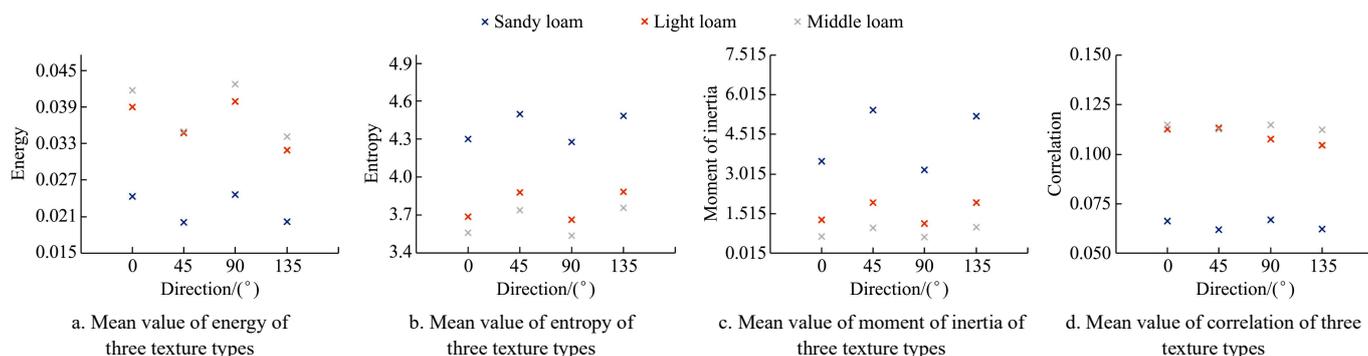
Figure 6   Mean value of a texture parameter of three texture types

3.2   Classification model

Six models such as Adaboost and LightGBR were used in the system.   The standard texture types measured in the laboratory are compared with the predicted texture classification of the detection system model.   Table 6 lists the comparison of the classification performance of different models.   Precision, Recall, and F1 were used as the evaluation indexes of the classification model.   Figure 7 shows the confusion matrix of the prediction performance of the six models.   Among them, ExtraTrees performed best, with Precision, Recall, and F1 all being 0.82.   Secondly, Adaboost's Precision, Recall, and F1 are 0.79, 0.70, and 0.69, respectively. SVM, BP Nerve Network, and Random Forest also performed well. Unfortunately, the Precision of the LightGBR model is poor, only 0.52.   From the perspective of the confusion matrix, SVM and ExtraTrees are the best predictors of Sandy loam.   The former incorrectly judges 2 samples as Light loam, and the latter incorrectly judges 1 sample as Light loam and 1 sample as Medium

loam.   The best predictor of Light loam is BP Nerve Network, which misjudged 3 samples out of 25 samples.   Adaboost is the best predictor of medium load.   In traditional laboratory texture measurement, the higher the content of physical clay (<0.01 mm), the more uncertainty it shows.   However, due to the small number of samples, it is impossible to determine whether AdaBoost model still performs well in a large range of samples.

Table 6   Comparison of texture classification of different models

Parameters	Precision	Recall	F1
Adaboost	0.79	0.70	0.69
LightGBR	0.52	0.55	0.53
SVM	0.74	0.73	0.70
BP Nerve Network	0.70	0.57	0.60
Random Forest	0.68	0.69	0.68
ExtraTrees	**0.82**	**0.82**	**0.82**

Note: SVM: Support Vector Machine; BP: Back propagation.

a. Performance of Adaboost model in prediction set

b. Performance of LightGBR model in prediction set

c. Performance of SVM model in prediction set

d. Performance of BP Nerve Network model in prediction set

e. Performance of RF model in prediction set

f. Performance of ExtraTrees model in prediction set
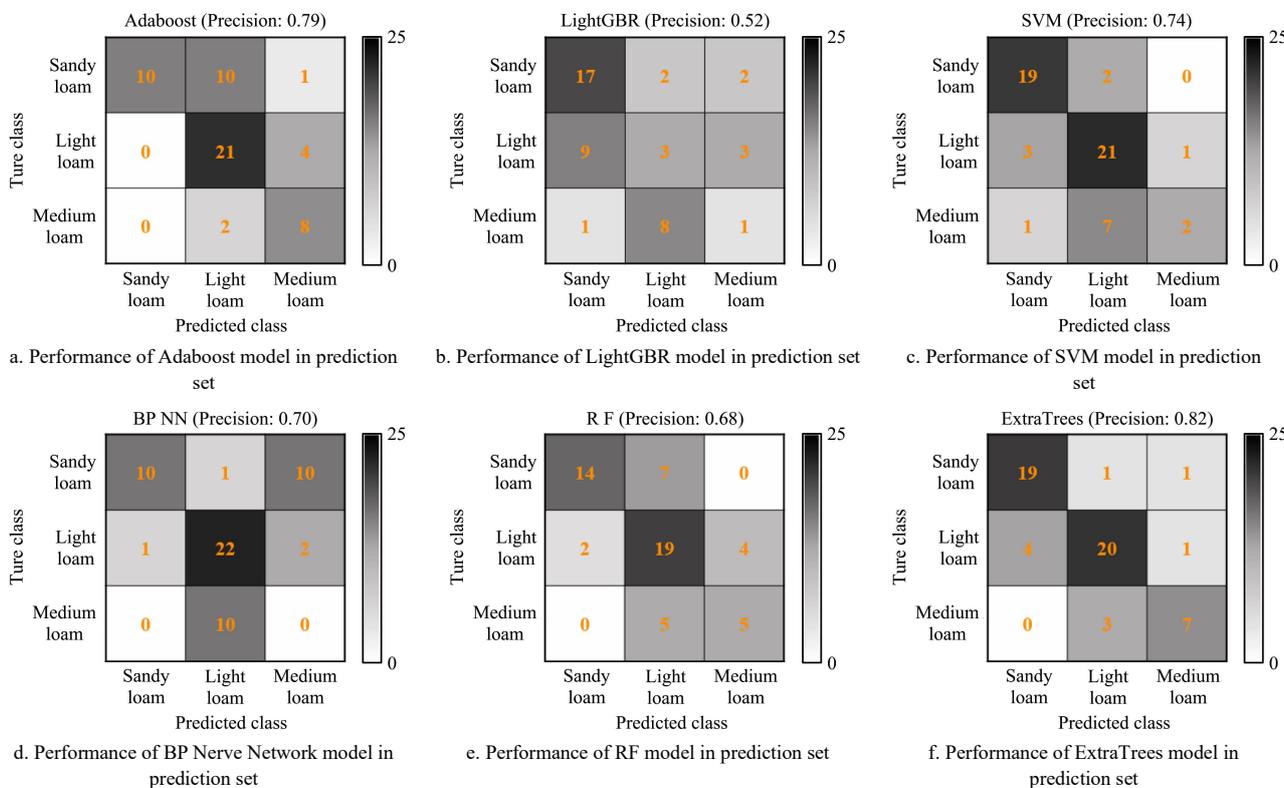
Figure 7   Confusion matrix of texture detection results of model in test set

A wise choice of input variables is essential for portable and user-friendly devices because adding too many predictors often leads to over-simplification of the predictive model.   In this study, a total of 17 parameters were used as the input of the model.   For the best-performing ExtraTrees model, Figure 8 shows the weight

map of the influence model of each input.   The first 4 weights are ranked: ECa, Moment of inertia 0°, Moment of inertia 135°, and Entropy 90°, and the following feature importance is observed in descending order of ECa, Moment of inertia, Entropy, Correlation, and Energy.   The close relationship between ECa and soil texture

has been confirmed. The performances of Moment of inertia and Entropy in Table 5 and Figure 6 were also confirmed. With the change of direction, the mean values of the three texture types maintain the same changing trend and spacing. Moment of inertia reflects the strength of the texture, and the depth of grooves is large. When the content of physical clay (<0.01 mm) is small, the particles have a certain distance. Entropy reflects the complexity of the texture. Literature [32] introduced that the skin texture becomes more and more sparse with age, and the entropy value decreases. This also verifies that as the content of physical clay (<0.01 mm) increases, the entropy value decreases.
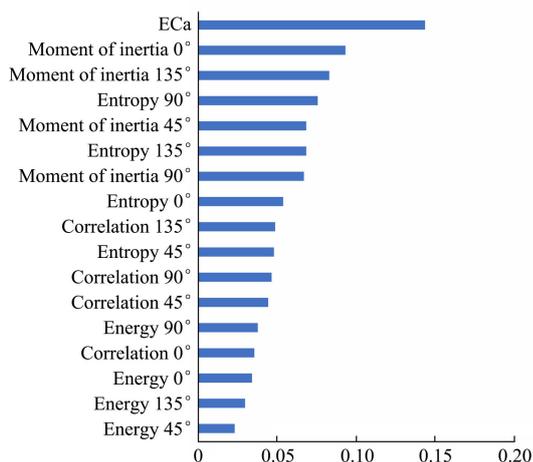
Figure 8   Classification model influence weight

3.3 Physical clay content prediction

In order to find the best model for predicting physical clay (<0.01 mm) content by the detection system, the performance of six models including Adaboost and LightGBR were compared. The model evaluation results are listed in Table 7. The maximum R^2 of ExtraTrees is 0.77, RMSE and MAPE were 74.72 and 39.58 respectively, followed by R^2 of Adaboost, LightGBR, and Random Forest were 0.74, 0.70, and 0.65. SVM has the lowest RMSE (72.06), and LightGBR has the lowest MAPE (35.00). Taken together, ExtraTrees performed the best, while SVM and BP Nerve Network did not perform very well. The reason why ExtraTrees performs well may be that each of its decision trees uses all the original training sample data, which helps reduce the bias of the model. Since the best-split attribute and threshold are randomly selected and set, it has a strong generalization ability. The poor performance of SVM may be due to the fact that SVM is used to solve support vectors by quadratic programming, which is difficult to implement for large-scale training samples The reason why BP neural network does not perform well may be: it is highly dependent on samples, because of the complexity of the soil, there may be contradictory samples and redundant samples in the sample set, so it is difficult for BP neural network to achieve the expected performance.

Table 7   Evaluation results of different models

Parameters	R^2	RMSE	MAPE
Adaboost	0.74	77.71	41.22
LightGBR	0.70	73.86	**35.00**
SVM	0.49	**72.06**	41.90
BP Nerve Network	0.58	77.25	39.22
Random Forest	0.65	85.80	43.75
ExtraTrees	**0.77**	74.72	39.58

A total of 17 parameters are used as the input of the model. For the best-performing ExtraTrees model, the weight diagram of

the influence model of each input is drawn, as shown in Figure 9. The top 4 weight rankings are Moment of inertia 135°, ECa, Moment of inertia 90°, and Entropy 90°. And the importance of the following features is observed in descending order of Moment of inertia, ECa, Entropy, Correlation, and Energy. This is similar to the result shown in Figure 8, indicating that if high accuracy is not required, the only moment of inertia, ECa, and entropy can be used as input.
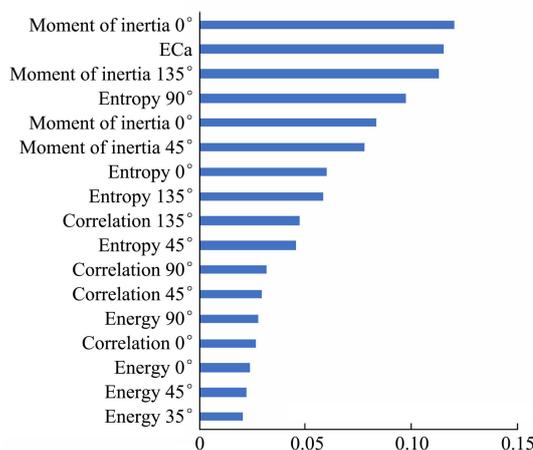
Figure 9   Physical clay content prediction model influence weight

3.4 Performance comparison and limitation

The soil texture prediction performance of this research was compared with the results of other scholars.

From the results of texture classification, the overall accuracy of the ExtraTrees model is 82.14% (46/56). The results of this study are better than the results of Demattê et al.[33] using satellite images to classify soil texture. Their accuracy rates for clay loam and sandy loam are 60.6% and 56.6%. It is also higher than the texture classification performed by Vibhute et al.[34] using hyperspectral remote sensing data (overall accuracy 71.18%). Higher than Taghizadeh-Mehrjardi et al.[35] and Dharumarajan et al.[36] using Digital Soil Mapping (DSM) to classify soil texture, the former has an overall accuracy rate of 67.5%, while the latter has an overall accuracy of 50%-65% in estimating six standard soil depth intervals. But it is usually lower than the results of texture prediction using ex-situ information (such as soil images that have been sieved and air-dried). For example, in the study of Barman et al.[37], the average accuracy is 91.37%.

From the results of texture particle composition, ExtraTrees model R^2 is 0.77, RSEM is 74.72. This study produced a better soil texture prediction effect than Aitkenhead et al.[38] (using Digital RGB photography and visible-range spectroscopy to predict soil texture, the R^2 of sand, silt, and clay are 0.27, 0.25, and 0.18, respectively). Qi et al.[39] used the microscope-based sensor to predict the soil texture. The R^2 of indoor measured and on-site wet image predictions for sand, silt, and clay are 0.78, 0.67, and 0.52, respectively. This research is superior to this method because of its time advantage and the ability to use in-situ information to measure large-scale data sets. The RMSE of the optimal model in this study is 74.72, which is inferior to the results of Sudarsan et al.[40]. Their study uses microscope images to study texture. The R^2 of the fine fractions (clay+silt) measured in the laboratory is 0.88, and the RMSE is 44.7, but slightly better than the results observed by Sudarsan et al.[41] using microscope computer vision technology (RMSE is 84.7). The results of this study are generally lower than those using spectroscopy techniques. For example, Benedet et al.[42] used pXRF to best predict the

content of sand, silt, and clay, with R^2 as high as 0.91, 0.81, and 0.83, respectively; Wang et al.[43] used Fourier transform near-infrared spectrometer and pXRF to study two data fusion methods to predict soil texture. The prediction accuracy R^2 of sand and clay content is 0.85 and 0.93, respectively. However, the advantage of this study is that the input of the detection system model is all in-situ information, and there is no need for pretreatment of soil samples such as grinding and sieving in the laboratory.

Soil moisture content is one of the main challenges of image-based near-end soil sensing. Because of the higher water-holding capacity of clay, image distortion may occur in soils with high clay content[44], and soil moisture content also affects ECa[45], it is necessary to conduct further research to measure the impact of moisture content on the performance of the detection system. At the same time, the lack of sufficient samples of other texture types limits the scope and performance of the test model to a certain extent. The next step should be to develop a detection system for more texture types to apply to all texture prediction models.

4 Conclusions

This study introduced a self-developed vehicle-mounted in-situ soil texture detection system and explores its optimal texture classification model and optimal soil texture particle composition prediction model. The detection system is mainly composed of soil electrical conductivity (ECa) sensors, cameras, auxiliary mechanical structures, and electronic systems. After the camera takes the image of the soil surface, the texture feature was extracted, the ECa was used as an input to the embedded model of the detection system, and finally, the predicted soil texture classification and particle composition are obtained. Data collection was carried out using the detection system in three test plots, and the best model research was carried out using the collected data. conclusion as below:

1) For soil texture classification, ExtraTrees performed best, with Precision, Recall, and F1 all being 0.82. In the drawn model influence weight map, the following feature importance is observed in descending order of ECa, Moment of inertia, Entropy, Correlation, and Energy.

2) For the prediction of soil particle content, the R^2 of ExtraTrees was 0.77, and RMSE and MAPE were 74.72 and 39.58, respectively. Approximately, the following feature importance is observed in the drawn model influence weight graph in descending order of Moment of inertia, ECa, Entropy, Correlation, and Energy. In general, ECa, Moment of inertia, and Entropy are the main contributors to predicting texture.

3) It is necessary to conduct more experiments to study the influence of soil moisture on the prediction performance of the model, and to study more texture types to expand the database for predicting soil texture models.

Acknowledgements

[References]

[1] Shahadat Hossain M, Mustafizur Rahman G K M, Saiful Alam M, Mizanur Rahman M, Solaiman A R M, Baset Mia, M A. Modelling of soil texture and its verification with related soil properties. Soil Research, 2018; 56(4): 421–428.

[2] Coblinski J A, Inda A V, Dematte J A M, Dotto A C, Gholizadeh A, Giasson E. Identification of minerals in subtropical soils with different textural classes by VIS-NIR-SWIR reflectance spectroscopy. CATENA, 2021; 203: 105334. doi: 10.1016/j.catena.2021.105334.

[3] Hollis L O, Turner R E. The tensile root strength of spartina patens varies with soil texture and atrazine concentration. Estuaries and Coasts, 2019; 42: 1430–1439.

[4] Hegde R, Bhaskar B P, Niranjana K V, Ramesh Kumar S C, Ramamurthy V, Srinivas S, et al. Land evaluation for groundnut (*Arachis hypogaea* L.) production in Pulivendula tehsil, Kadapa district, Andhra Pradesh, India. Legume Research, 2017; 42(3): 326–333.

[5] Li H R, Liu B, Wang R X, Liu W, Fang Y, Yang D L, et al. Particle-size distribution affected by testing method. Journal of Desert Research, 2018; 38(3): 619–627. (in Chinese)

[6] Wang W P, Liu J L, Zhang J B, Li X P. Evaluation and correction of measurement using diffraction method for soil particle size distribution. Transactions of the CSAE, 2014; 30(22):163–169. (in Chinese)

[7] Stefani Fae G, Montes Felipe, Bazilevskaya E, Masip Ano R, Kemanian A R. Making soil particle size analysis by laser diffraction compatible with standard soil texture determination methods. Soil Science Society of America Journal, 2019; 83(4): 1244–1252.

[8] Zhu Y, Zhang Z D, Liu C, Zhang X. Comparison of laser diffraction method and pipette method on soil particle size distribution determination - a case study of variously degraded kastanozem. Research of Soil and Water Conservation, 2018; 25(3): 62–67, 204. (in Chinese)

[9] Fisher P, Aumann C, Chia K, O'Halloran N, Chandra S. Adequacy of laser diffraction for soil particle size analysis. Plos One, 2017; 12(5): e0176510. doi: 10.1371/journal.pone.0176510.

[10] Swetha R K, Bende P, Singh K, Gorthi S, Biswas A, Li B, et al. Predicting soil texture from smartphone-captured digital images and an application. Geoderma, 2020; 376: 114562. doi: 10.1016/j.geoderma.2020.114562.

[11] de Oliveira Morais P A, de Souza D M, Madari B E, de Oliveira A E. A computer-assisted soil texture analysis using digitally scanned images. Computers and Electronics in Agriculture, 2020; 174: 105435. doi: 10.1016/j.compag.2020.105435.

[12] de Oliveira Morais P A, de Souza D M, de Melo Carvalho M T, Madari B E, de Oliveira A E. Predicting soil texture using image analysis. Microchemical Journal, 2019; 146: 455–463.

[13] Sudarsan B, Ji W J, Biswas A, Adamchuk V. Microscope-based computer vision to characterize soil texture and soil organic matter. Biosystems Engineering, 2016; 152: 41–50.

[14] García-Tomillo A, Mirás-Avalos J M, Dafonte-Dafonte J, Paz-González A. Mapping soil texture using geostatistical interpolation combined with electromagnetic induction measurements. Soil Science, 2017; 182(8): 278–284.

[15] Andrade R, Silva S H G, Faria W M, Poggere G C, Barbosa J Z, Guiherme L R G, et al. Proximal sensing applied to soil texture prediction and mapping in Brazil. Geoderma Regional, 2020; 23: e00321. doi: 10.1016/j.geodrs.2020.e00321.

[16] Kargas G, Londra P, Sgoubopoulou A. Comparison of soil EC values from methods based on 1:1 and 1:5 soil to water ratios and ECe from saturated paste extract based method. Water, 2020; 12(4): 1010. doi: 10.3390/w12041010.

[17] Wu K N, Zhao R. Soil texture classification and its application in China. Acta Pedologica Sinica, 2019; 56(1): 227–241.

[18] Pei X S, Meng C, Li M Z, Yang W, Zhou P. Measurement of soil electrical conductivity based on direct digital synthesizer (DDS) and digital oscilloscope. Int J Agric & Biol Eng, 2019; 12(6):162–168.

[19] Pare S, Bhandari A K, Kumar A, Singh G K. An optimal color image multilevel thresholding technique using grey-level co-occurrence matrix. Expert Systems with Applications, 2017; 87: 335–362.

[20] Awais M, Ghayvat H, Pandarathodiyil A K, Ghani W M N, Ramanathan A, Pandya S, et al. Healthcare professional in the loop (HPIL): Classification of standard and oral cancer-causing anomalous regions of oral cavity using textural analysis technique in autofluorescence imaging. Sensors, 2020; 20(20): 5780. doi: 10.3390/s20205780.

[21] Kulkarni A, Carrion-Martinez I, Dhindsa K, Alaref A A, Rozenberg R, van der Pol C B. Pancreas adenocarcinoma CT texture analysis: Comparison of 3D and 2D tumor segmentation techniques. Abdominal Radiology, 2021; 46(4): 1027–1033.

[22] Guan H X, Liu H J, Meng X T, Luo C, Bao Y L, Ma Y Y, et al. A quantitative monitoring method for determining maize lodging in different growth stages. Remote Sensing, 2020; 12(19): 3149. doi: 10.3390/rs12193149.

[23] Wang H P, Hui L. Classification recognition of impurities in seed cotton based on local binary pattern and gray level co-occurrence matrix. Transactions of the CSAE, 2015; 31(3): 236–241. (in Chinese)

[24] Liu Z. Image-based prediction of soil roughness and soil bulk density. Master dissertation. Beijing: China Agricultural University, 2020; pp.37–46. (in Chinese)

[25] Liu Z, Yang W, Li M Z, Zhou P, Yao X Q, Chen Y Q, et al. Soil roughness measuring system combined with image processing. IFAC-PapersOnLine, 2018; 51(17): 689–694.

[26] Dou P, Chen Y B. Remote sensing imagery classification using AdaBoost with a weight vector (WV AdaBoost). Remote Sensing Letters, 2017; 8(8): 733–742.

[27] Ge X, Sun J, Lu B, Chen Q S, Xun W, Jin Y T. Classification of oolong tea varieties based on hyperspectral imaging technology and BOSS-LightGBM model. Journal of Food Process Engineering, 2019; 42(8): e13289. doi: 10.1111/jfpe.13289.

[28] Huang M W, Chen C W, Lin W C, Ke S W, Tsai C F. SVM and SVM ensembles in breast cancer prediction. Plos One, 2017; 12(1): e0161501. doi: 10.1371/journal.pone.0161501.

[29] Wu M, Lyu B L. Prediction of viscosity of ternary tin-based lead-free solder melt using BP neural network. Soldering and Surface Mount Technology, 2020; 32(3): 173–180.

[30] Samat A, Liu S C, Persello C, Li E Z, Miao Z L, Abuduwaili J. Evaluation of ForestPA for VHR RS image classification using spectral and superpixel-guided morphological profiles. European Journal of Remote Sensing, 2019; 52(1):107–121.

[31] Mantas C J, Castellano J G, Moral-García S, Abellán J. A comparison of random forest based algorithms: random credal random forest versus oblique random forest. Soft Computing, 2019; 23(5): 10739–10754.

[32] Ou X, Pan W, Xiao P. In vivo skin capacitive imaging analysis by using grey level co-occurrence matrix (GLCM). International Journal of Pharmaceutics, 2014; 460(1-2): 28–32.

[33] Demattê J A M, Alves M R, da Silva Terra F, Bosquillia R W D, Fongaro C T, da Silva Barros P P. Is it possible to classify topsoil texture using a sensor located 800 km away from the surface? Revista Brasileira De Ciência Do Solo, 2016; 40: e0150335. doi: 10.1590/18069657rbcs20150335.

[34] Vibhute A D, Kale K V, Dhumal R K, Mehrotra S C. Soil type classification and mapping using hyperspectral remote sensing data. In: 2015 International Conference on Man and Machine Interfacing (MAMI), Bhubaneswar: IEEE, 2016; pp.1–4. doi: 10.1109/MAMI.2015.7456607.

[35] Taghizadeh-Mehrjardi R, Sarmadian F, Minasny B, Triantafilis T, Omid M. Digital mapping of soil classes using decision tree and auxiliary data in the ardakan region. Arid Land Research and Management, 2014; 28(2): 147–168.

[36] Dharumarajan S, Hegde R. Digital mapping of soil texture classes using Random Forest classification algorithm. Soil Use and Management, 2020; 38(1): 135–149.

[37] Barman U, Dev Choudhury R. Soil texture classification using multi class support vector machine. Information Processing in Agriculture, 2019; 7(2): 318–332.

[38] Aitkenhead M, Cameron C, Gaskin G, Choisy B, Coull M, Black H. Digital RGB photography and visible-range spectroscopy for soil composition analysis. Geoderma, 2018; 313: 265–275.

[39] Qi L, Adamchuk V, Huang H H, Leclerc M, Jiang Y, Biswas A. Proximal sensing of soil particle sizes using a microscope-based sensor and bag of visual words model. Geoderma, 2019; 351:144–152.

[40] Sudarsan B, Ji W J, Adamchuk V, Biswas A. Characterizing soil particle sizes using wavelet analysis of microscope images. Computers and Electronics in Agriculture, 2018; 148: 217–225.

[41] Sudarsan B, Ji W J, Biswas A, Adamchuk V. Microscope-based computer vision to characterize soil texture and soil organic matter. Biosystems Engineering, 2016; 152: 41–50.

[42] Benedet L, Faria W M, Silva S H G, Mancini M, Demattê J A M, Guilherme L R G, et al. Soil texture prediction using portable X-ray fluorescence spectrometry and visible near-infrared diffuse reflectance spectroscopy. Geoderma, 2020; 376: 114553. doi: 10.1016/j.geoderma.2020.114553.

[43] Wang S Q, Li W D, Li J, Liu X S. Prediction of soil texture using FT-NIR spectroscopy and PXRF spectrometry with data fusion. Soil Science, 2013; 178(11): 626–638.

[44] Kelley J, Higgins C W, Pahlow M, Noller J. Mapping soil texture by electromagnetic induction: a case for regional data coordination. Soil Science Society of America Journal, 2017; 81(4): 923–931.

[45] Bañón S, Álvarez S, Bañón D, da Ortuño M F, Sánchez-Blanco M J. Assessment of soil salinity indexes using electrical conductivity sensors. Scientia Horticulturae, 2021; 285: 110171. doi: 10.1016/j.scienta.2021.110171.