

Article

Desert Soil Salinity Inversion Models Based on Field In Situ Spectroscopy in Southern Xinjiang, China

Yu Wang ¹, Modong Xie ², Bifeng Hu ^{3,4} , Qingsong Jiang ⁵, Zhou Shi ⁶ , Yinfeng He ⁷ and Jie Peng ^{1,*}

¹ College of Agriculture, Tarim University, Alar 843300, China

² College of Horticulture, Gansu Agricultural University, Lanzhou 730070, China

³ Department of Land Resource Management, School of Tourism and Urban Management, Jiangxi University of Finance and Economics, Nanchang 330013, China

⁴ Laboratory of Environment Remediation and Ecological Health, Ministry of Education, Zhejiang University, Hangzhou 310058, China

⁵ College of Information Engineering, Tarim University, Alar 843300, China

⁶ Institute of Applied Remote Sensing and Information Technology, College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China

⁷ Urumqi Natural Resources Comprehensive Survey Center, China Geological Survey, Urumqi 830026, China

* Correspondence: 11414049@zju.edu.cn or pjzky@163.com

Abstract: Soil salinization is prominent environmental issue in arid and semi-arid regions, such as Xinjiang in Northwest China. Salinization severely restricts economic and agricultural development and would lead to ecosystem degradation. Finding a method of rapidly and accurately determining soil salinity (SS) is one of the main challenges in salinity evaluation, saline soil development, and utilization. In situ visible and near infrared (Vis-NIR) spectroscopy has proven to be a promising technique for detecting soil properties since it can realize real-time, rapid detection of SS. However, it still remains challenging whether Vis-NIR in situ spectroscopy can invert SS with high accuracy due to the interference of environmental factors (e.g., light, water vapor, solar altitude angle, etc.) on the spectral in the field. To fill this knowledge gap, we collected Vis-NIR in situ spectral and lab-measured SS data from 135 surface soil samples in the Kongterik Pasture Nature Reserve (KPNR) in the desert oasis ecotone of southern Xinjiang, China. We used genetic algorithm (GA), particle swarm optimization (PSO), and simulated annealing (SA) algorithms to select the feature bands of SS. Subsequently, we combined extreme learning machines (ELM), back-propagation neural networks (BPNN), and convolutional neural networks (CNN) to build inversion models of SS. The results showed that different feature bands selection methods could improve the Vis-NIR in situ spectral prediction model accuracy. Either SS inversion models were built using full-band spectral data or feature-band spectral data. Compared with the full-band (401–2400 nm) spectral modeling, the validation set R^2 of ELM, BPNN, and CNN models built selected feature bands selected by PSO, GA, and SA, respectively, were improved by more than 0.06. The accuracy of predicting SS varied widely among modeling methods. The accuracy of CNN model was obviously higher than that of BPNN and ELM models. The optimal hybrid model for predicting SS constructed in this study is SA-CNN model ($R^2 = 0.79$, RMSE = 9.41 g kg⁻¹, RPD = 1.81, RPIQ = 2.37). This study showed that the spectral feature bands selection methods can reduce the influence of environmental factors on in situ spectroscopy and significantly enhance the inversion accuracy of SS. The present study provided that estimating SS using in situ Vis-NIR spectral is feasible.

Keywords: Vis-NIR in situ spectroscopy; soil salinity; feature bands selection method; deep learning; inversion model



Citation: Wang, Y.; Xie, M.; Hu, B.; Jiang, Q.; Shi, Z.; He, Y.; Peng, J. Desert Soil Salinity Inversion Models Based on Field In Situ Spectroscopy in Southern Xinjiang, China. *Remote Sens.* **2022**, *14*, 4962. <https://doi.org/10.3390/rs14194962>

Academic Editor: Lenio Soares Galvão

Received: 14 August 2022

Accepted: 4 October 2022

Published: 5 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Soil salinization poses a great threat to agricultural development and food security, especially in arid and semi-arid regions, including areas of China [1,2]. Presently, salinization affects more than 3% of the world's soil resources [3]. The global area of primary

salinized soils is about $9.55 \times 10^8 \text{ hm}^2$, and the area of secondary salinized soils is about $4.4 \times 10^7 \text{ hm}^2$ [4]. China has about $3.67 \times 10^7 \text{ hm}^2$ of salinized soils, which mainly occur in the Northeast Plain, North China Plain, and Northcentral and Northwest China [5]. Xinjiang is located in the northwest of China and accounts for more than 50% of total saline area in China. Therefore, dealing with the issue of soil salinization plays a critical role in maintaining sustainable agricultural development and ecosystem functions in Xinjiang and China. Currently, soil salinity (SS) is the principal method for determining soil salt levels [6]. A rapid, accurate, and effective SS method is needed to assess and resolve soil salination in China and elsewhere.

Traditional laboratory analysis approach for SS is time-consuming and difficult to meet the requirements of SS monitoring at large spatial scale. In recent years, Vis-NIR spectroscopy has provided an alternative means of rapidly obtaining information on soil properties. Numerous scholars have used Vis-NIR indoor spectroscopy combined with different modeling methods to construct soil property inversion models and obtained high accuracy [7–9]. Although Vis-NIR indoor spectroscopy for soil inversion has the advantage of high accuracy, it still cannot avoid the tedious work of soil sample collection, air-drying, grinding, and sieving and still suffers from a large workload and relatively long determination cycle. In contrast, the in situ spectroscopy method can omit the above soil sample collection and processing process, which is faster and simpler than the indoor Vis-NIR spectroscopy method [10]. This could significantly enhance the efficiency of soil salinization detection.

Moreover, SS can strongly vary in both vertical and horizontal space, and the spatial and temporal variation of SS is more pronounced especially in saline soils [11]. The high spatial and temporal variability of SS introduces difficulty to traditional measuring methods for obtaining SS in a time-sensitive and rapid manner, especially on a large spatial scale. In contrast, the Vis-NIR spectroscopy-based soil near-ground sensor technology can measure the in situ hyperspectral data of soil in a real time, low cost and quickly way.

However, when sensors are used to measure soil properties outdoors, we must overcome the interference of the soil moisture and soil structure. Meanwhile, soil Vis-NIR spectral are influenced by various environmental factors, such as cloud cover, atmospheric water vapor, and light [10]. Among them, water vapor has the most influence on soil spectra. Soil Vis-NIR spectral are strongly influenced by water vapor in the air. It can produce strong noise in the absorption band of the spectrum near 1400 and 1900 nm, which can seriously interfere with the spectral data [11]. Several studies have proposed that preprocessing methods can effectively reduce spectral noise and can eliminate the effects of some interfering factors [10]. It has also been shown that there is a high redundancy of information between different spectral bands, and the feature band selection method can eliminate irrelevant or redundant spectral features. Thus, the relations between soil properties and spectral features could be better explored and minimize the interference of irrelevant wavelengths as well as the influence of some environmental factors. However, whether the feature band selection methods can be applied to in situ hyperspectral inversion of soil salinization remains to be proven by research.

The models and algorithms used are another critical factor which would affect the accuracy of spectral inversion of soil properties. Among different algorithms, the artificial neural network models (ANN), support vector regression (SVR) models, and partial least squares regression (PLSR) models have proven to be able make accurate predictions for SS [8,12,13]. Zhang et al. developed a PLSR model and a principal component regression (PCR) model for SS and soil spectral reflectance using different spectral treatments. The model accuracy between the two was compared, and the results showed that the PLSR model outperformed the PCR model [14]. Mahajan et al. found that SVR, PLSR, and PCR models were more robust than MARS and RF models in soil property prediction models [8]. The use of the back-propagation neural network (BPNN) method to predict soil properties has been widely used in many recent studies. Meanwhile, the extreme learning machines (ELM) method is also gradually applied in soil property studies. ELM is

a superior modeling method compared to other algorithms as it can reduce the time for feature extraction and prediction and improve learning efficiency. Khosravina et al. used ELM and BPNN to better predict Pb and Zn in soil with R^2 of 0.87 and 0.81, respectively [15]. Although ELM techniques have been widely used in many fields, less research has been conducted on the application of SS monitoring [16]. To select the best salinity inversion model, many researchers have also compared multiple models [11,17,18], but no uniform results have been obtained. For conventional prediction models, the main difficulty in extracting useful information from Vis-NIR spectra depends on the amount of data and data dimensionality [19]. Conventional Vis-NIR spectral data require pre-processing to improve the model prediction accuracy, such as denoising, dimensionality reduction, spectral transformation and other processes. During these processes, there is a risk of introducing false information and eliminating valid information [19]. In contrast, the convolutional neural network (CNN) method is an end-to-end approach that avoids these processing procedures. It mines only the target information from the raw spectral data and is a superior approach for predicting soil properties [20]. Padarian et al. constructed inversion models for various soil properties using PLSR, Cubist, and CNN methods. Their results showed that CNN models outperformed PLSR and Cubist models [21].

In summary, this study mainly intended to explore the potential of field in situ spectroscopy for estimating SS. We took the KPNR in the desert oasis ecotone of South Xinjiang, China as the study area, and we introduced the feature band selection method into the study of in situ spectral estimation of SS. Three band selection algorithms, PSO, GA, and SA, were used to extract the characteristic spectral information of SS and combined with ELM, BPNN, and CNN methods, respectively, to build SS inversion models. The specific aims of this study are: (1) to reduce invalid information and noise in the in situ spectra using feature band selection methods; (2) to compare the accuracy of different modeling approaches for the estimation of SS; (3) to establish a combinatorial model for predicting SS with high accuracy using in situ spectroscopy.

2. Material and Methods

2.1. Study Area

The study area is located in the KPNR in Wensu County, Aksu Region, southern Xinjiang, China, and it is a typical desert oasis ecological zone that is a sparsely vegetated transition between barren desert and a fairly vegetated moist oasis [22]. The KPNR is at southern foot of the Tianshan Mountains and the northern edge of the Taklamakan Desert, spanning 61 km from east to west and 93 km from north to south. The elevation decreases from northwest to southeast. The area belongs to the middle and distal part of the alluvial fan. Different soil salinization conditions exist in the study area. The degree of soil salinization increases and then decreases from south to north and shows an increasing–decreasing–increasing trend from west to east [23]. The mean annual evaporation in the study area is 1331.56 mm; the mean annual precipitation is 47.21 mm, with precipitation mainly concentrated in May–October, which accounts for about 97% of the yearly precipitation. The average sunshine time is 3000 h, which is typical of a warm temperate continental arid climate. The study area has very abundant light and heat resources, but the soil has poor water retention capacity, low precipitation, strong evaporation, low vegetation cover, and natural vegetation dominated by salt-tolerant plants [24]. The geographical location of the study area is from 80°53′ to 81°10′E longitude and 40°52′ to 41°10′N latitude (Figure 1).

2.2. Soil Sample Collection and Analysis

Soil samples were collected from 13 October to 15 October 2021, with clear, cloudless weather and no rainfall during the sampling period. We collected samples mainly on both sides of the road (S215) that crosses the KPNR in the northwest–southwest direction, with a small number of samples in the southwest–northeast direction through the central area of the KPNR. Soil samples were collected at approximately 500 m intervals. A total of 135 surface (0–20 cm) soil samples were collected from the study area. Considering the

surface roughness of saline soils, there may be light leakage due to using contact probes, and the limited measurement range using contact probes makes it difficult to collect enough soil samples. Therefore, samples were selected in the bare soil area where the ground was relatively flat and measured using the non-contact fiber optic probe that comes with the instrument. The spectrometer probe was 1 m above ground level and measured vertically downward with a field of view of 25° . Soil samples were collected within a diameter of 44 cm and an area of 0.15 m^2 using the vertical downward direct point of the probe as the center of the circle. Location information of the sampling points is recorded using a handheld global positioning system (GPS). After measuring the soil spectrum, about 500 g of soil was collected at a depth of 0–20 cm within the spectral determination field of view, and the soil samples were placed in sealed bags, numbered, and brought back to the laboratory. They were then ground and passed through a 2 mm sieve after natural air-drying to determinate the SS (water-soluble salts) with a soil–water ratio of 1:5. The determination method uses the residue-drying mass method. The SS was used as the dependent variable and the in situ spectral reflectance as the predictor.

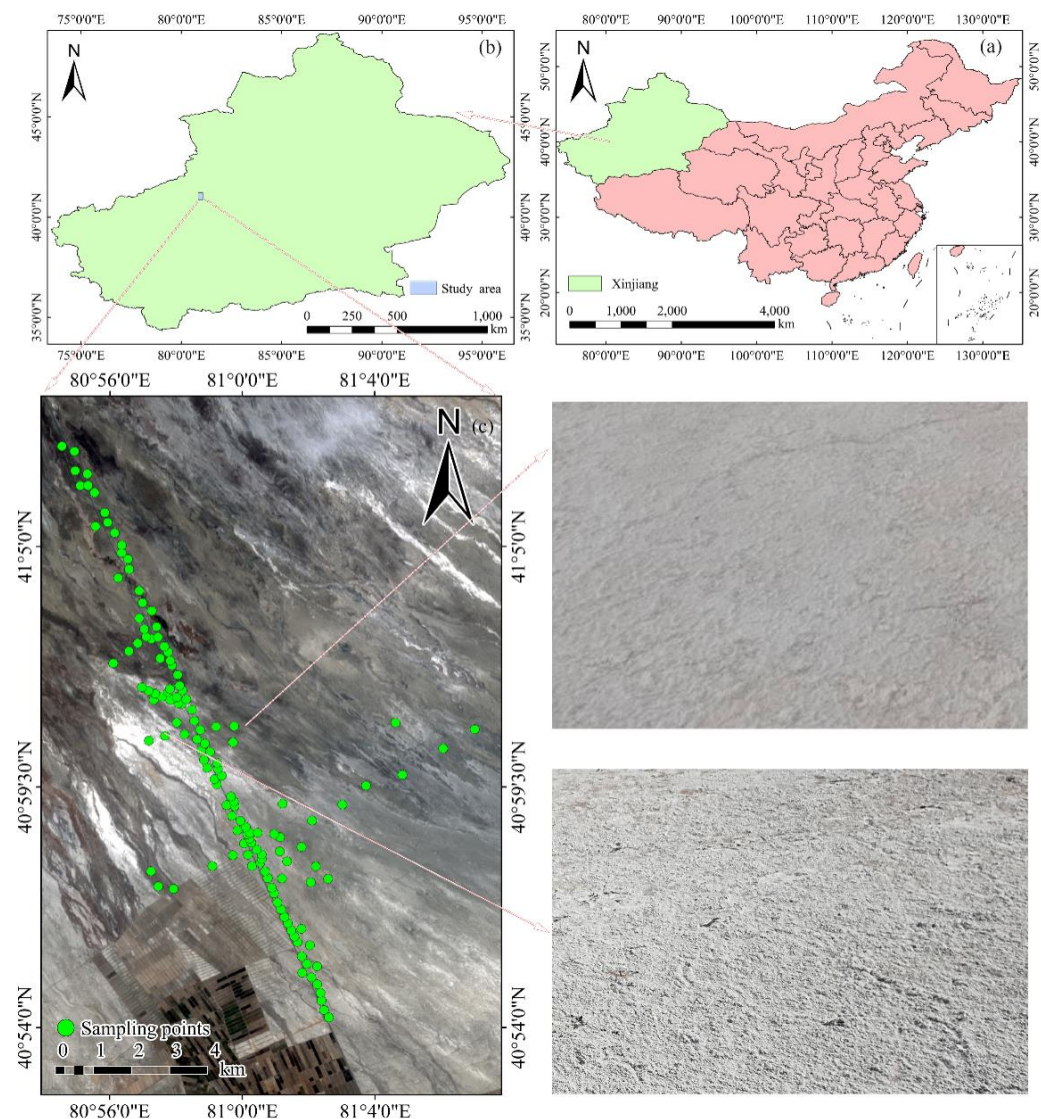


Figure 1. Study area and sampling points within Xinjiang Province. The location of Xinjiang, China (a), the location of the study area (b), and the distribution of sampling points (c).

2.3. In Situ Spectral Data Acquisition and Pre-Processing

In situ Vis-NIR spectroscopy of soil samples was measured using a FieldSpec4 geophysical spectrometer from ASD, USA [8]. The band width or sampling intervals of the Spectrometer were 1.4 nm and 2 nm in the 350–1000 nm and 1000–2500 nm wavelength ranges, respectively. The instrument is warmed for at least 30 min before the spectroscopic measurement to ensure that it operates in optimal condition. Vis-NIR in situ hyperspectral data acquisition was performed at 13:00–15:00. The sky was clear and cloudless during the time period of the spectral acquisition, and the light was intense [25]. Acquisition of spectral samples was performed using a pistol-type fiber optic handle accompanying the equipment, with the light source being sunlight. Dark current calibration of the instrument prior to spectral measurement, whiteboard calibration, and optimization. For spectral measurements, soil spectral profiles were collected at the center point of the probe vertically downward and in the surrounding direction, respectively. Every soil sample collected has ten spectral curves, and the average value of the spectral reflectance is taken as the spectral value of the soil sample and positioned using GPS. Vis-NIR in situ spectroscopies were removed from the edge bands of 350–400 nm and 2401–2500 nm with low signal-to-noise ratio and high noise, and the remaining bands were retained. Then, the Savitzky–Golay 9-point smoothing process was also performed. Figure 2 shows the in situ raw spectral profiles in the field.

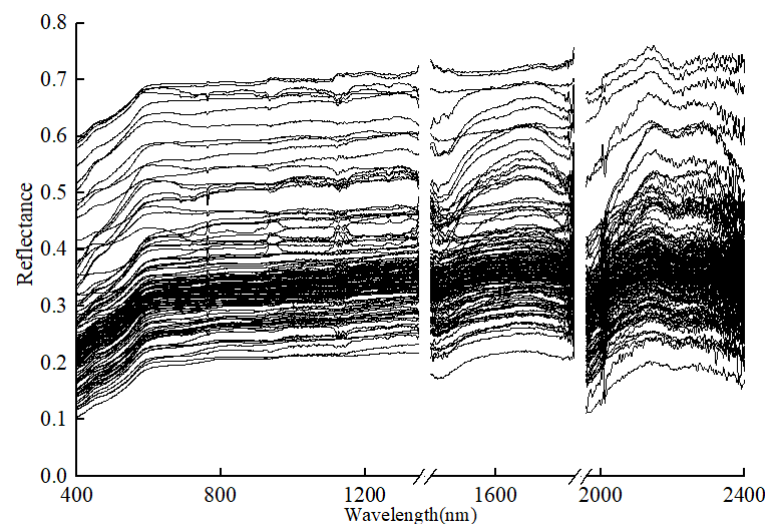


Figure 2. Vis-NIR in situ spectral profile.

2.4. Feature Bands Selection

The Vis-NIR spectral contain information of thousands of bands, but most of them show strong covariance. Feature band selection methods can reduce redundant information. It helps to reduce the complexity of prediction model building and try to make the model prediction more accurate. Therefore, we chose three methods—PSO, GA, and SA—for extracting the feature bands of SS separately and comparing their differences.

Particle swarm algorithm (PSO) is an intelligent optimization algorithm. Suppose that there are N particles forming a population in a D -dimensional space. Where the position and velocity information of particle I in the population is denoted as $X_i = (X_{i1}, X_{i2}, \dots, X_{iD})Z$, $V_i = (V_{i1}, V_{i2}, \dots, V_{iD})Z$, $i = 1, 2, \dots, N$. Calculate all the fitness-valued particles and find the optimal solution. Update the speed and position information based on Equations (1) and (2).

$$V_i^{z+1} = \omega V_i^z + c_1 r_1 (P_i^z - X_i^z) + c_2 r_2 (P_g^z - X_i^z) \quad (1)$$

$$X_i^{z+1} = X_i^z + V_i^{z+1} \quad (2)$$

V_i represents the speed of the particle, r_1 and r_2 are random values between 0 and 1, and X_i is the current position of the particle. ω is a non-negative inertia coefficient. c_1 and c_2 are learning coefficients used to adjust the individual extremes P_i and P_g , respectively [26]. In this study, PSO was used for bands selection. We set the number of particle populations, acceleration coefficients c_1 and c_2 , and the maximum number of iterations to 50, 2, and 100, respectively.

The simulated annealing algorithm (SA) is widely known as a meta-heuristic algorithm. It is generated from the physical annealing process of solids in metallurgy [27]. SA starts with a higher initial temperature. Then, the temperature is gradually reduced until the temperature reaches the equilibrium condition. At each temperature, multiple searches are performed, and each search generates a new solution via random perturbation. In each step, we will randomly select a feature subset based on the current optimal feature subset. If it works better, then we will adopt it and update the current optimal feature subset. If it does not, we will probably accept it depending on the current temperature. Accepting a subset of poorly performing features with a certain probability is crucial for SA algorithms because it helps prevent the algorithm from falling into a local optimum. With iterations, the SA algorithm converges on a good and stable final result. We set the initial temperature of SA to 1000, the cooling rate α to 0.95, the maximum number of iterations to 100, and the number of runs to 50.

Genetic algorithm (GA) is one of the most popular algorithms for feature selection [28]. This algorithm was first proposed by John Holland in the United States, is a computational model of biological evolutionary process that simulates the mechanism of natural selection and genetics of Darwinian biological evolution, and is a method of searching for the optimal solution by simulating the natural evolutionary process. The main operators include selection, crossover, and mutation. Starting from an arbitrary initial population, the population evolves to an increasingly better region of the search space by generating a group of individuals better suited to the environment through selection, crossover and mutation operations. In this way, the population evolves from generation to generation, finally finding the optimal solution to the problem. In this study, the number of iterations of the GA is set to 100, the variance probability to 0.05, and the crossover probability to 0.1.

2.5. Model Establishment and Accuracy Evaluation

The ELM is one of the feed-forward neural network models. Compared with traditional neural network algorithms, the only variable to set for ELM is the number of hidden layer neurons in the network. The input weights and hidden layer deviation values of the network are randomly generated. The learning process only calculates the output weights. Good generalization performance and fast learning are its advantages [29]. In this paper, the ELM model function we used was the Sigmoid. We determined the number of implicit neurons in the ELM model that are suitable for SS. Although ELM has good generalization ability, the model stability can be affected because of the redundant information variables of the input. Therefore, we used three feature band selection algorithms, thus obtaining an ELM model that stably predicts salinity information.

The BPNN is considered the most commonly used prediction method and consists of input layer, hidden layer, and output layer. Parameters must be determined based on the inputs and outputs when building it. For example, the weights between neurons in the input, hidden, and output layers; the learning rate; the activation function; the initialization threshold for the hidden layer; and the threshold for the output layer [30]. The model has the advantages of simple structure and high predictive power and can meet the need for estimating and inverting SS from hyperspectral data [31]. In this study, when we used BPNN models and CNN predictions, we used 30% of the modeling set randomly for internal validation to avoid overfitting. The best model parameters after several training sessions with full-band spectral data were used as parameters for all BPNN models in the study, where the number of implied layers was 10, the learning rate was 0.0001, and 1000 iterations were performed.

The CNN is an end-to-end supervised neural network, which is a representative deep learning algorithm. The earliest CNN was developed by LeCun in 1989. It was used to classify handwritten digital images, and then developed by other scholars to different domains. It is usually made up of one or more convolutional, pooling, and fully connected layers. The convolution operation extracts different features from the input layer. The hyperparameters of the convolution layer include filter size (F), stride(S), and padding (P). Filter size (F) is the convolution kernel. The stride (S) is the size of the distance of each slide of the convolution kernel. Zero padding (P) means adding an appropriate number of zeros to the edges of the input [32]. Pooling is also commonly called subsampling. When building a convolutional neural network, it is often used after the convolutional layer. Pooling is used to reduce the feature dimensionality of the convolutional layer output, effectively reducing the network parameters while preventing overfitting. The fully connected layer was reconstructed to extract features for mapping to the final output [33]. Malek et al. and Hong et al. explored a 1D-CNN network model for spectral regression [34,35]. We refer to the network architecture they built and propose our CNN model. In this study, we train the CNN models built on different datasets several times, adjust the values of the hyperparameters at each training, and finally use the parameters of the optimal training model they produce as the final hyperparameters.

In this paper, we constructed a 1D-CNN constructure (Table 1). Figure 3 provides the constructure of the proposed CNN. There are six trainable layers (three convolutional layers and three fully connected layers) in this network. First, the input layer is the spectral data, and it has 2000 bands ranging from 401 to 2400 nm. Here, we perform smoothing treatment on the spectral data to avoid the interference of spectral noise. Then, normalization is performed to finally obtain the spectral data that we use. The second layer is a convolutional layer with a kernel size of 10, 10 filters, a stride of 1, and a padding of the same value. Subsequent convolutional layers have 21 filters and 42 filters, and other parameters are unchanged. The maximum pooling layer has a kernel size of 1 and a stride of 2. The output of the convolutional layer is passed to the pooling layer, and the ReLU function is chosen as the activation function. Layers 5, 6, and 7 are fully connected layers with sizes of 20, 10, and 1, respectively. The activation function is chosen as the ReLU function, with a random dropout of 50% of the number of neurons (dropout (0.5)). The batch size of the network is 50, the maximum number of iterations is 2400, and the learning rate is set to 0.0001. Meanwhile, in this study, we obtain a new dataset after feature band selection for full-band spectral data. However, the input spectral dimensions are different for the feature band dataset, and in order to obtain optimal results, we adjust the parameters and hyperparameters of the network on the basis of the full-band CNN model to build a CNN model based on different feature band selection methods.

Table 1. Sequence and description of the layers used in the CNN structure.

Layer	Type	Kernel Size	Filters	Activation
1	Input	— —	— —	— —
2	Conv1	10×1	10	ReLU
3	Maxpooling1	1×1	— —	— —
4	Conv2	5×1	21	ReLU
5	Maxpooling2	1×1	— —	— —
6	Conv3	2×1	42	ReLU
7	Maxpooling3	1×1	— —	— —
8	Fully-connected1	— —	— —	ReLU
9	Fully-connected2	— —	— —	ReLU
10	Output	— —	— —	ReLU

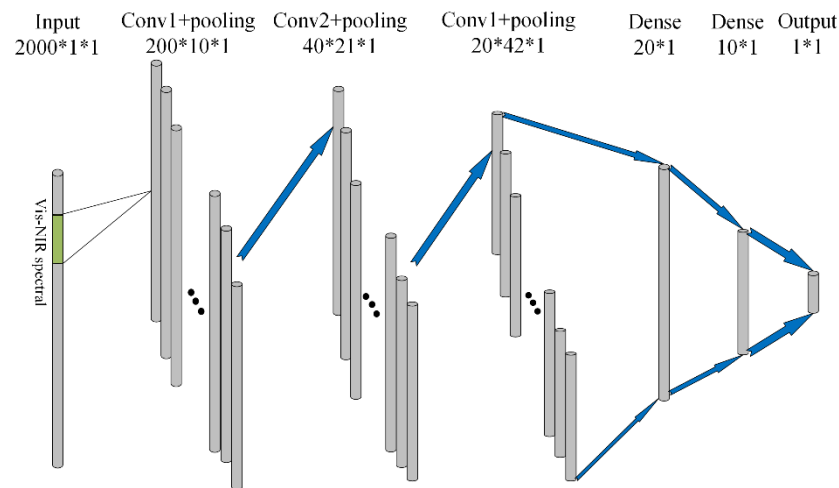


Figure 3. Structure of proposed CNN model for SS prediction.

The following metrics were used for model evaluation: coefficient of determination (R^2), root mean square error (RMSE), relative analysis error (RPD), and the ratio of performance to the interquartile range (RPIQ). R^2 is used to measure the prediction accuracy of the fitted model. RMSE is used to measure the difference between the predicted and observed values of the fitted model [36]. RPD explains the performance of the model. For RPD, when $RPD < 1.4$, the model is not valid; when $1.4 \leq RPD < 1.8$, the model has average predictive power; when $RPD \geq 1.8$, the model has good predictive power [37]. For RPIQ, when $RPIQ \geq 2.2$, the model has good predictive ability; $1.7 \leq RPIQ < 2.2$, the model has average predictive ability; $RPIQ < 1.7$ indicates low model reliability.

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (4)$$

$$RPD = \frac{SD_{yi}}{RMSE} \quad (5)$$

$$RPIQ = \frac{Q3 - Q1}{RMSE} \quad (6)$$

3. Results

3.1. Descriptive Statistics of SS

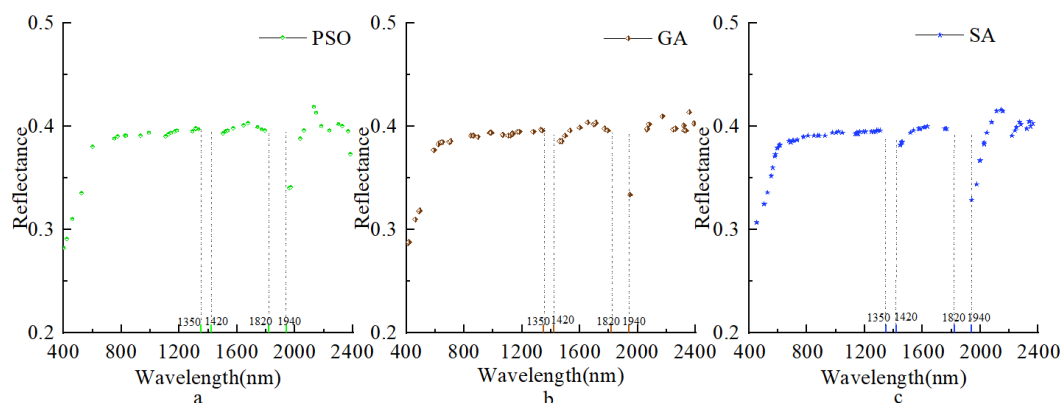
Table 2 provides the descriptive statistics of SS in the study area. To ensure the robustness and generalization ability of the model, the sample data set ($n = 135$) was divided into a calibration set of 90 samples and a validation set of 45 samples using the Kennard–Stone (K-S) algorithm [38]. The maximum SS in all samples was 109.24 g kg^{-1} , the minimum was 27.01 g kg^{-1} , and the average value was 68.74 g kg^{-1} , all of which were heavily salinized [39]. There were varying degrees of salinization in the study area, and the overall salinity content was high. The SS was highly banded throughout the study area. The maximum soil salt in the calibration set was 109.24 g kg^{-1} , and the minimum was 27.01 g kg^{-1} . The maximum SS in the validation set was 109.18 g kg^{-1} , and the minimum was 28.07 g kg^{-1} . The distribution of SS in the calibration and validation sets was relatively uniform with little variation. The coefficients of variation were 26.71–26.90%, which were medium variability. This indicates that the variability is not significant in the spatial distribution of SS.

Table 2. Descriptive statistics of SS (g kg^{-1}).

Dataset	Number	Mean	Max	Min	SD	CV (%)
Calibration	90	68.72	109.24	27.01	18.40	26.77
Validation	45	68.78	109.18	28.07	18.50	26.90
Entire	135	68.74	109.24	27.01	18.36	26.71

3.2. Feature Band Selection Based on the Different Methods

The feature bands selection method is important in regression analysis to improve model prediction and enhance model robustness. It can minimize the effect of uncorrelated or noisy variables by taking into account the interaction information between effective spectral variables. In this study, we selected the characteristic wavelengths of SS of PSO, GA, and SA. The 401–2400 nm band of Vis-NIR in situ spectroscopy was selected as the full wavelength for the feature bands selection. Figure 4 presented the distribution of the characteristic wavelengths selected by PSO, GA, and SA. In this study, 41 feature variables were screened using the PSO method. Using GA, the corresponding 52 best wavelengths were obtained as input variables after 100 iterations. With the SA method, 70 characteristic wavelengths for SS were finally selected. These wavelengths were then used as new datasets to construct ELM, BPNN, and CNN prediction models. GA and PSO have the same characteristics, and they both belong to global optimization techniques. Both algorithms search in the global solution space and focus their search on the high-performance part [40]. However, the results show that compared to PSO and GA methods, SA can retain more useful information.

**Figure 4.** Result of (a) PSO, (b) GA, (c) SA characteristic wavelength screening.

Using the CARS method, 15 SS-characteristic bands focus on 400–962 nm were selected by Zhu et al. [41]. Wang et al. reported that the reflectance spectral curves of salt content of three typical salt kinds in the west of China were relatively similar, and all three salt types were distributed in the range of 400–2400 nm [1]. Mahajan revealed that soil salts exhibit specific absorption properties at wavelengths of 427, 487, 950, 1414, 1917, 2206, 2380, and 2460 nm [8]. Sidike et al. found that saline soils exhibit absorption properties in the visible (400–770 nm) and near infrared (900–1030 nm, 1270–1700 nm, 1900–2150 nm, 2150–2310 nm, 2320–2400 nm) spectra with significant absorption properties [42]. However, there were some differences in the bands selected by PSO, GA, and SA in this paper. However, the characteristic bands selected by the three methods are basically consistent with the results of Sidike et al. [42]. As presented in Figure 4, no feature bands in the 1350–1420 nm and 1820–1940 nm ranges, which were highly disturbed by water vapor in the atmosphere, were screened out. It is well known that the spectral features were highly correlated with moisture in the ranges of 1350–1420 nm and 1820–1940 nm. After processing the soil Vis-NIR in situ spectral data using the feature band selection method, no feature bands highly correlated with moisture were filtered out. This indicates that the feature band

selection method could effectively reject the interference of atmospheric water vapor on the soil salt inversion. In this study, the SS-sensitive wavelengths covered most of the spectrum. After processing by the three algorithms, the number of rejected wavelengths reached more than 90%. It can obviously reduce the amount of model inputs and improved the efficiency of model operations. However, there were differences in the selection of SS-characteristic wavebands by different selection methods.

3.3. Predictive Regression Models

Feature band selection method is an effective means to improve model prediction accuracy. To promote the accuracy of the model performance and decline the interference of atmospheric water vapor, we used characteristic band modeling. The Table 3 presented the prediction results of the SS models with the datasets selected by different feature band selection methods and the full band dataset. All the feature band selection methods substantially improved the model prediction capability compared with full-band modeling (Table 3). The R^2 of ELM, BPNN, and CNN models built by the feature bands selected by PSO, GA, and SA, respectively, were all improved by more than 0.06. This indicates that the feature selection methods could effectively eliminate the redundant information of the spectra, select effective feature bands, and make the prediction accuracy of the models more accurate. Among the three feature bands selection methods, SA had the largest promotion in model accuracy, followed by GA and PSO.

Table 3. Comparison of prediction accuracy of different SS models (g kg^{-1}).

FEA	Model	Calibration		Validation			
		R^2	RMSE	R^2	RMSE	RPD	RPIQ
R	ELM	0.50	13.21	0.46	13.57	1.25	1.40
	BPNN	0.59	11.96	0.51	12.86	1.32	1.46
	CNN	0.63	11.40	0.57	11.62	1.46	1.49
GA	ELM	0.64	11.24	0.55	11.90	1.42	1.63
	BPNN	0.70	10.44	0.60	11.26	1.51	1.79
	CNN	0.76	9.85	0.68	11.00	1.54	1.82
PSO	ELM	0.59	12.25	0.52	12.69	1.34	1.48
	BPNN	0.68	11.05	0.61	11.18	1.52	1.75
	CNN	0.73	10.25	0.65	11.06	1.53	1.81
SA	ELM	0.65	11.13	0.57	11.61	1.46	1.66
	BPNN	0.71	10.97	0.63	11.16	1.52	1.85
	CNN	0.84	8.64	0.79	9.41	1.81	2.37

The PSO-ELM model (validation set $R^2 = 0.52$, $RPD = 1.34$) built based on the dataset selected by PSO showed the worst performance for predicting SS effectively in this study. Wei et al. found that the PSO-BPNN model was significantly higher than the BPNN model, indicating that PSO can significantly improve the model's accuracy [43]. The PSO-BPNN model in this study was significantly higher than the BPNN model, which is consistent with Wei's findings. The inversion accuracy of the CNN model was significantly higher than that of the BPNN and ELM models for both the SS inversion models built using full-band spectral data and feature-band spectral data. The optimal hybrid models (Figure 5) established by different modeling methods combined with the feature bands selection method was SA-ELM, SA-BPNN, and SA-CNN models, respectively. Among them, the SA-CNN hybrid model ($R^2 = 0.79$, $RMSE = 9.41 \text{ g kg}^{-1}$, $RPD = 1.81$, $RPIQ = 2.37$) had the highest accuracy and can accurately invert SS.

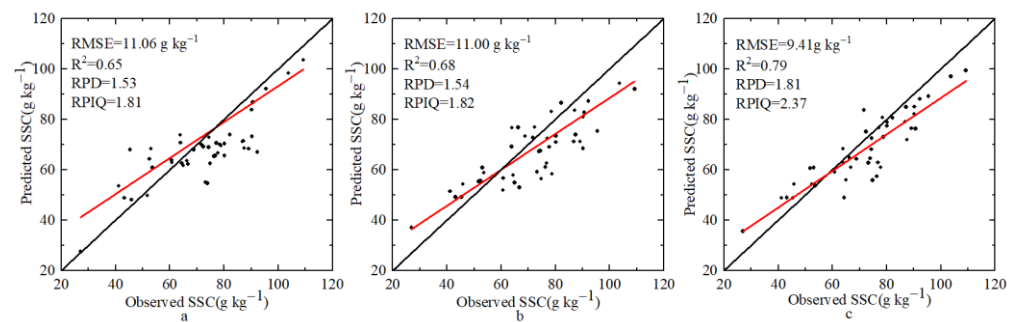


Figure 5. Scatter plot of the three best hybrid models. (a) PSOCNN model; (b) GA-CNN model; (c) SA-CNN model.

4. Discussion

4.1. Source of Uncertainty of Predicting SS Using Field In Situ Spectroscopy

The accuracy of in situ spectroscopy predicting of salinity is affected by many factors, and thus, its predictive power still has room for improvement. The sources of uncertainty are mainly soil moisture content, light factor, SS type, water vapor in the atmosphere, and vegetation cover [25,44–48]. Ben-Dor et al. predicted soil properties by using the PLSR method—soil moisture, organic matter, carbonate, and iron oxide are some of them—and the results showed that in situ Vis-NIR spectroscopy has good potential for characterizing soil properties [49]. Janik et al. performed in situ spectroscopic measurements under dry conditions in summer and autumn. They found that in situ measurements under dry conditions exhibited greater potential [50]. Some researchers have also developed EPO-PLSR models for different soil types to predict soil moisture content by reducing the effect of soil type through the external parameter orthogonalization (EPO) algorithm [51]. Wang et al. confirmed that the spectral of different salinization types of soils differ significantly. The differences due to different salinization types are even greater than the differences due to different salt contents [52]. The samples in this study were collected in autumn, which is a dry season with little precipitation and low soil moisture content. In particular, the surface soil was very dry with a moisture content of less than 10%; thus, soil moisture has little effect on in situ spectroscopy.

In current research, the sampling time was chosen from 13:00 to 15:00, during which the sunlight shines directly on the ground and there is no shadow and sufficient light, which has less influence on the spectral acquisition. In this study, soil samples were collected from the same study area, which caused less impact on the in situ spectroscopy. The vegetation cover rate in the study area is low. Although there were a few salt-tolerant plants growing, the in situ spectral collection was performed in the bare soil area, which effectively avoided the effect from the difference in vegetation cover. Atmospheric water vapor is identified as the most influential factor on model prediction. The SS inversion model, which was established by filtering the characteristic bands of SS through the feature bands selection method, can predict SS better. It is proven that the redundant information and error information in Vis-NIR in situ spectroscopy, including the noise generated by atmospheric water vapor and the anomalous spectral bands, can be reduced by the feature bands selection method.

4.2. Performance Comparison of Different Feature Band Selection Methods

Selecting effective spectral feature bands is a key step to ensure the accuracy of SS prediction models, but different feature band selection methods have different effects on model performance. The feature bands selection algorithm can remove redundant information [53]. Our results revealed that SA was better than GA and PSO, and the R^2 of the combined SA-CNN model reached 0.79, and the RPD was 1.81, which could better predict the SS. In contrast, the R^2 of the models constructed by the feature bands selected by GA and PSO were both lower than 0.65 and the RPD were both lower than 1.55. The PSO obtains an effective subset of variables by establishing a suitable objective function.

With these subsets of variables, a highly accurate model with good robustness can be obtained [38]. Based on this, we concluded that the PSO method is as appropriate of a method as the PSO-CNN model (validation set $R^2 = 0.65$) developed in this study and that it still yielded acceptable results for predicting SS content. The wave modeling selected using different feature bands selection methods all obviously improved the predictive power of these models. However, our results indicated that the predictive ability of the model built by PSO was significantly inferior to that of GA and SA, mainly because the PSO algorithm converges quickly when the weights are not added, and premature convergence tends to fall into local optimal solutions [54]. In contrast, GA and SA have strong global search capability and are able to obtain optimal solutions. Therefore, based on the three feature bands selection methods, GA and SA perform better.

4.3. Comparison of ELM, BPNN, and CNN Estimation Models

The ELM is a spectral prediction model and commonly shows strong predictive power [38]. However, in this study, the prediction ability of ELM is not as good as that of the CNN model. In general, the model input parameters are the key factors affecting the prediction accuracy of the model. Since the input parameters of ELM are randomly generated (possibly zero), this may cause negative effects on the hidden layer neurons. To achieve the desired accuracy, ELM usually requires many hidden layer neurons. However, the use of a large number of hidden layer neurons can increase the complexity of the network and reduce the generalization ability of the model [26], which is probably the main reason for the poor prediction results of the model in this study.

The BPNN method shows great advantages in its capability of capturing nonlinear relationships, adaptability, and generalization ability [55]. Here, the BPNN method was able to predict soil salt content better. In its optimal model, SA-BPNN, its prediction set R^2 reached 0.63, which basically met our requirement for in situ spectroscopy predicting soil salt content. Deep learning methods have achieved great success in recent years. It is considered an effective regression tool in quantifying soil properties [56]. Xiao et al. constructed a salt detection model using the CNN-GRC-ELM method. The R^2 of the CNN-GRC-ELM model was 0.90, and the RMSE was 1.55 g kg^{-1} . This proved that it could achieve feasible online fast salt content detection [57].

Moreover, the CNN showed good prediction performance in this study, with R^2 of 0.79 for SA-CNN. The deep learning method outperformed the machine learning method among the three modeling methods and was able to predict SS more accurately. However, the ELM and BPNN methods can still be used to construct SS inversion models and combining with other methods may be able to better improve model prediction accuracy. The transferability of CNN models across the globe has been analyzed in several studies. Nonetheless, considering the spatial heterogeneity of SS types and the influence of uncertainties in in situ spectral predictions, a highly accurate CNN model may exhibit poor model effectiveness when migrating to other regions [33,58]. For the SA-CNN method, it is limited by various uncertainties of in situ measurements, and although we use the SA method to attenuate their influence on SS prediction, we cannot completely eliminate their interference; we must make improvements based on the SA-CNN method, such as adding a residual module to the CNN model. A major problem of SA is that when the initial temperature is too high, the convergence the speed is slow, so it is more appropriate to reduce the initial parameters and take the subset of the result notification from multiple runs as the optimal features. In the future, we will consider comparing CNN methods to other deep learning methods, such as LSTM, DBN, GRU, etc. We will study it to a deeper extent from two perspectives, such as contact and non-contact. Later, we will explore this method for modelling in situ spectral of different SS content gradients on this basis, and we will further explore the method of in situ spectroscopy for predicting SS with higher accuracy.

4.4. Interpretability of the Selected Bands and Limitations of the Study

In in situ spectroscopy, soil moisture is a major influence when using contact measurement methods, and it has been demonstrated that EPO, DS, and PDS algorithms can remove soil moisture interference very well [59]. However, when using non-contact measurement methods, in addition to soil moisture, environmental factors also become a major influence factor. In this study, we collected soil samples during the dry season, a period when the surface soil is very dry and the soil moisture content is below 10%, effectively avoiding the influence of soil moisture on in situ spectral data. In data acquisition, environmental factors (atmospheric water vapor, light) become the main factors affecting the spectral data. It is well known that the feature bands selection method can remove noise and reduce the interference of invalid information, and in this study, we applied the feature band selection method to select SS features that are not sensitive to environmental factors, effectively eliminating the influence of environmental factors on in situ spectral prediction SS. Saline soils exhibit absorption properties in the 400–770 nm, 900–1030 nm, 1270–1700 nm, 1900–2150 nm, 2150–2310 nm, and 2320–2400 nm bands [42]. Mahajan et.al found that soil salts exhibited specific absorption characteristics at wavelengths of 427, 487, 950, 1414, 1917, 2206, 2380, and 2460 nm [8]. In this study, the characteristic bands of soil salts obtained by using the feature bands selection method are similar to Mahajan and are also expressed in these bands. The shortcomings of this study are that we are limited by the traffic conditions; the spatial distribution of the collected samples is not uniform, and the samples we collected mainly occurred on both sides of the road (S215) that crosses the KPNR in the northwest–southwest direction, with a small number of samples crossing the central area of the KPNR in the southwest–northeast direction, and some high or low salt samples were not collected, which may be beyond our model’s prediction range. The dominant soil type in the study area is saline soil, so different soil types would also have less impact. In this study, good results have been achieved for high-salinity saline soils in desert areas, and further tests are necessary for low- and medium-salinity soils.

5. Conclusions

Our results showed that different feature bands selection methods were effective in filtering out redundant SS feature bands and improving the accuracy of Vis-NIR in situ spectral modeling. Compared with full-band modeling, the absolute R^2 of the ELM, BPNN, and CNN models established by the feature bands selected by PSO, GA, and SA, respectively, were improved by 0.06, 0.09, and 0.11, respectively. Among the three feature band selection methods, SA outperforms than others two methods. Among the three methods used for predicting SS, the CNN model has the highest prediction accuracy. The optimal combination of the feature band selection method and the modeling method is SA-CNN, and the hybrid model has $R^2 = 0.79$, RMSE = 9.41 g kg^{−1}, RPD = 1.81, RPIQ = 2.37 is used for predicting SS, which can invert the SS.

Author Contributions: All authors contributed in a substantial way to the manuscript. Methodology, M.X. and Y.W.; formal analysis, Q.J. and Y.W.; data curation, Y.W.; writing—original draft preparation, Y.W.; writing—review and editing, J.P.; conceptualization, Z.S.; investigation, Y.H. and Y.W.; project administration, J.P.; funding acquisition, J.P.; visualization, B.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by grants from the National Science Foundation of China (Grant Nos.42261016, 42071068 and 41061031), the Tarim University President’s Fund (Grant Nos. TDZKCX202205, TDZKSS202227), the National Key Research and Development Program of China (Grant Nos. 2018YFE0107000), the Open foundation from Key Laboratory of Environment Remediation and Ecological Health (Zhejiang University), Ministry of Education (EREH202206), Chinese Universities Scientific Fund (Grant Nos. ZNLH201904), Program of China Geological Survey (Grant Nos. ZD20220127), and Tarim University Graduate Research Innovation Program (Grant Nos. TDGRI202116).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Wang, Q.; Li, P.; Chen, X. Modeling salinity effects on soil reflectance under various moisture conditions and its inverse application: A laboratory experiment. *Geoderma* **2012**, *170*, 103–111. [\[CrossRef\]](#)
- Haj-Amor, Z.; Araya, T.; Kim, D.-G.; Bouri, S.; Lee, J.; Ghiloufi, W.; Yang, Y.; Kang, H.; Jhariya, M.K.; Banerjee, A.; et al. Soil salinity and its associated effects on soil microorganisms, greenhouse gas emissions, crop yield, biodiversity and desertification: A review. *Sci. Total Environ.* **2022**, *843*, 156946. [\[CrossRef\]](#)
- Singh, A. Soil salinization management for sustainable development: A review. *J. Environ. Manag.* **2021**, *277*, 111383. [\[CrossRef\]](#)
- Metternicht, G.I.; Zinck, J.A. Remote sensing of soil salinity: Potentials and constraints. *Remote Sens. Environ.* **2003**, *85*, 1–20. [\[CrossRef\]](#)
- Liu, L.-P.; Long, X.-H.; Shao, H.-B.; Liu, Z.-P.; Ya, T.; Zhou, Q.-Z.; Zong, J.-Q. Ameliorants improve saline-alkaline soils on a large scale in northern Jiangsu Province, China. *Ecol. Eng.* **2015**, *81*, 328–334.
- Peng, J.; Ji, W.; Ma, Z.; Li, S.; Chen, S.; Zhou, L.; Shi, Z. Predicting total dissolved salts and soluble ion concentrations in agricultural soils using portable visible near-infrared and mid-infrared spectrometers. *Biosyst. Eng.* **2016**, *152*, 94–103. [\[CrossRef\]](#)
- Hu, B.; Chen, S.; Hu, J.; Xia, F.; Xu, J.; Li, Y.; Shi, Z. Application of portable XRF and VNIR sensors for rapid assessment of soil heavy metal pollution. *PLoS ONE* **2017**, *12*, e0172438. [\[CrossRef\]](#) [\[PubMed\]](#)
- Mahajan, G.R.; Das, B.; Gaikwad, B.; Murganekar, D.; Desai, A.; Morajkar, S.; Patel, K.P.; Kulkarni, R.M. Monitoring properties of the salt-affected soils by multivariate analysis of the visible and near-infrared hyperspectral data. *Catena* **2021**, *198*, 105041. [\[CrossRef\]](#)
- Bai, Z.; Xie, M.; Hu, B.; Luo, D.; Wan, C.; Peng, J.; Shi, Z. Estimation of Soil Organic Carbon Using Vis-NIR Spectral Data and Spectral Feature Bands Selection in Southern Xinjiang, China. *Sensors* **2022**, *22*, 6124. [\[CrossRef\]](#) [\[PubMed\]](#)
- Biney, J.K.M.; Blöcher, J.R.; Bell, S.M.; Borůvka, M.; Vašát, R. Can in situ spectral measurements under disturbance-reduced environmental conditions help improve soil organic carbon estimation? *Sci. Total Environ.* **2022**, *838*, 156304. [\[CrossRef\]](#) [\[PubMed\]](#)
- Taghdis, S.; Farpoor, M.H.; Mahmoodabadi, M. Pedological assessments along an arid and semi-arid transect using soil spectral behavior analysis. *Catena* **2022**, *214*, 106288. [\[CrossRef\]](#)
- Farifteh, J.; Van der Meer, F.; Atzberger, C.; Carranza, E.J.M. Quantitative analysis of salt-affected soil reflectance spectra: A comparison of two adaptive methods (PLSR and ANN). *Remote Sens. Environ.* **2007**, *110*, 59–78. [\[CrossRef\]](#)
- Mashimbye, Z.E.; Cho, M.A.; Nell, J.P.; De Clercq, W.P.; Van Niekerk, A.; Turner, D.P. Model-Based Integrated Methods for Quantitative Estimation of Soil Salinity from Hyperspectral Remote Sensing Data: A Case Study of Selected South African Soils. *Pedosphere* **2012**, *22*, 640–649. [\[CrossRef\]](#)
- Zhang, X.; Huang, B. Prediction of soil salinity with soil-reflected spectra: A comparison of two regression methods. *Sci. Rep.* **2019**, *9*, 5067. [\[CrossRef\]](#) [\[PubMed\]](#)
- Khosravi, V.; Ardejani, F.D.; Yousefi, S.; Aryafar, A. Monitoring soil lead and zinc contents via combination of spectroscopy with extreme learning machine and other data mining methods. *Geoderma* **2018**, *318*, 29–41. [\[CrossRef\]](#)
- Zhang, J.; Zhang, Z.; Chen, J.; Chen, H.; Jin, J.; Han, J.; Wang, X.; Song, Z.; Wei, G. Estimating soil salinity with different fractional vegetation cover using remote sensing. *Land Degrad. Dev.* **2021**, *32*, 597–612. [\[CrossRef\]](#)
- Ding, J.; Yu, D. Monitoring and evaluating spatial variability of soil salinity in dry and wet seasons in the Werigan-Kuqa Oasis, China, using remote sensing and electromagnetic induction instruments. *Geoderma* **2014**, *235–236*, 316–322. [\[CrossRef\]](#)
- Aldabaa, A.A.A.; Weindorf, D.C.; Chakraborty, S.; Sharma, A.; Li, B. Combination of proximal and remote sensing methods for rapid soil salinity quantification. *Geoderma* **2015**, *239–240*, 34–46. [\[CrossRef\]](#)
- Wang, Y.; Li, M.; Ji, R.; Wang, M.; Zheng, L. A deep learning-based method for screening soil total nitrogen characteristic wavelengths. *Comput. Electron. Agric.* **2021**, *187*, 106228. [\[CrossRef\]](#)
- Yang, J.; Wang, X.; Wang, R.; Wang, H. Combination of Convolutional Neural Networks and Recurrent Neural Networks for predicting soil properties using Vis-NIR spectroscopy. *Geoderma* **2020**, *380*, 114616. [\[CrossRef\]](#)
- Padarian, J.; Minasny, B.; McBratney, A.B. Using deep learning to predict soil properties from regional spectral data. *Geoderma Reg.* **2019**, *16*, e00198. [\[CrossRef\]](#)
- Wang, J.; Peng, J.; Li, H.; Yin, C.; Liu, W.; Wang, T.; Zhang, H. Soil Salinity Mapping Using Machine Learning Algorithms with the Sentinel-2 MSI in Arid Areas, China. *Remote Sens.* **2021**, *13*, 305. [\[CrossRef\]](#)
- Peng, J.; Biswas, A.; Jiang, Q.; Zhao, R.; Hu, J.; Hu, B.; Shi, Z. Estimating soil salinity from remote sensing and terrain data in southern Xinjiang Province, China. *Geoderma* **2019**, *337*, 1309–1319. [\[CrossRef\]](#)
- Wang, N.; Peng, J.; Xue, J.; Zhang, X.; Huang, J.; Biswas, A.; He, Y.; Shi, Z. A framework for determining the total salt content of soil profiles using time-series Sentinel-2 images and a random forest-temporal convolution network. *Geoderma* **2022**, *409*, 115656. [\[CrossRef\]](#)
- Stevens, A.; van Wesemael, B.; Bartholomeus, H.; Rosillon, D.; Tychon, B.; Ben-Dor, E. Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils. *Geoderma* **2008**, *144*, 395–404. [\[CrossRef\]](#)

26. Jian, H.; Lin, Q.; Wu, J.; Fan, X.; Wang, X. Design of the color classification system for sunglass lenses using PCA-PSO-ELM. *Measurement* **2022**, *189*, 110498. [\[CrossRef\]](#)
27. Ong, P.; Tung, I.-C.; Chiu, C.-F.; Tsai, I.-L.; Shih, H.-C.; Chen, S.; Chuang, Y.-K. Determination of aflatoxin B1 level in rice (*Oryza sativa* L.) through near-infrared spectroscopy and an improved simulated annealing variable selection method. *Food Control* **2022**, *136*, 108886. [\[CrossRef\]](#)
28. Sajadi, S.; Fathi, A. Genetic algorithm based local and global spectral features extraction for ear recognition. *Expert Syst. Appl.* **2020**, *159*, 113639. [\[CrossRef\]](#)
29. Hong, Y.; Chen, S.; Zhang, Y.; Chen, Y.; Yu, L.; Liu, Y.; Liu, Y.; Cheng, H.; Liu, Y. Rapid identification of soil organic matter level via visible and near-infrared spectroscopy: Effects of two-dimensional correlation coefficient and extreme learning machine. *Sci. Total Environ.* **2018**, *644*, 1232–1243. [\[CrossRef\]](#)
30. Ma, Q.; Teng, Y.; Li, C.; Jiang, L. Simultaneous quantitative determination of low-concentration ternary pesticide mixtures in wheat flour based on terahertz spectroscopy and BPNN. *Food. Chem.* **2022**, *377*, 132030. [\[CrossRef\]](#)
31. Wang, X.; Zhang, F.; Ding, J.; Kung, H.-T.; Latif, A.; Johnson, V.C. Estimation of soil salt content (SSC) in the Ebinur Lake Wetland National Nature Reserve (ELWNNR), Northwest China, based on a Bootstrap-BP neural network model and optimal spectral indices. *Sci. Total Environ.* **2018**, *615*, 918–930. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Xiao, C.; Wang, X.; Dou, H.; Li, H.; Lv, R.; Wu, Y.; Song, G.; Wang, W.; Zhai, R. Non-Uniform Synthetic Aperture Radiometer Image Reconstruction Based on Deep Convolutional Neural Network. *Remote Sens.* **2022**, *14*, 2359. [\[CrossRef\]](#)
33. Chen, Y.; Li, L.; Whiting, M.; Chen, F.; Sun, Z.; Song, K.; Wang, Q. Convolutional neural network model for soil moisture prediction and its transferability analysis based on laboratory Vis-NIR spectral data. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *104*, 102550. [\[CrossRef\]](#)
34. Cui, C.; Fearn, T. Modern practical convolutional neural networks for multivariate regression: Applications to NIR calibration. *Chemom. Intell. Lab. Syst.* **2018**, *182*, 9–20. [\[CrossRef\]](#)
35. Hong, Y.; Chen, Y.; Chen, S.; Shen, R.; Hu, B.; Peng, J.; Wang, N.; Guo, L.; Zhuo, Z.; Yang, Y.; et al. Data mining of urban soil spectral library for estimating organic carbon. *Geoderma* **2022**, *426*, 116102. [\[CrossRef\]](#)
36. Hu, B.; Xue, J.; Zhou, Y.; Shao, S.; Fu, Z.; Li, Y.; Chen, S.; Qi, L.; Shi, Z. Modelling bioaccumulation of heavy metals in soil-crop ecosystems and identifying its controlling factors using machine learning. *Environ. Pollut.* **2020**, *262*, 114308. [\[CrossRef\]](#)
37. Gholizadeh, A.; Žižala, D.; Saberioon, M.; Borůvka, L. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. *Remote Sens. Environ.* **2018**, *218*, 89–103. [\[CrossRef\]](#)
38. Lao, C.; Chen, J.; Zhang, Z.; Chen, Y.; Ma, Y.; Chen, H.; Gu, X.; Ning, J.; Jin, J.; Li, X. Predicting the contents of soil salt and major water-soluble ions with fractional-order derivative spectral indices and variable selection. *Comput. Electron. Agric.* **2021**, *182*, 106031. [\[CrossRef\]](#)
39. Wang, J.; Liu, Y.; Wang, S.; Liu, H.; Fu, G.; Xiong, Y. Spatial distribution of soil salinity and potential implications for soil management in the Manas River watershed. *China. Soil Use Manag.* **2020**, *36*, 93–103. [\[CrossRef\]](#)
40. Li, M.; Feng, Y.; Yu, Y.; Zhang, T.; Yan, C.; Tang, H.; Sheng, Q.; Li, H. Quantitative analysis of polycyclic aromatic hydrocarbons in soil by infrared spectroscopy combined with hybrid variable selection strategy and partial least squares. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2021**, *257*, 119771. [\[CrossRef\]](#)
41. Zhu, C.M.; Ding, J.L.; Zhang, Z.P.; Wang, Z. Exploring the potential of UAV hyperspectral image for estimating SS: Effects of optimal band combination algorithm and random forest. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2022**, *279*, 121416. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Sidike, A.; Zhao, S.; Wen, Y. Estimating soil salinity in Pingluo County of China using QuickBird data and soil reflectance spectra. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *26*, 156–175. [\[CrossRef\]](#)
43. Wei, Q.; Nurmemet, I.; Gao, M.; Xie, B. Inversion of Soil Salinity Using Multisource Remote Sensing Data and Particle Swarm Machine Learning Models in Keriya Oasis, Northwestern China. *Remote Sens.* **2022**, *14*, 512. [\[CrossRef\]](#)
44. Yu, W.; Hong, Y.; Chen, S.; Chen, Y.; Zhou, L. Comparing Two Different Development Methods of External Parameter Orthogonalization for Estimating Organic Carbon from Field-Moist Intact Soils by Reflectance Spectroscopy. *Remote Sens.* **2022**, *14*, 1303. [\[CrossRef\]](#)
45. Li, S.; Shi, Z.; Chen, S.; Ji, W.; Zhou, L.; Yu, W.; Webster, R. In situ measurements of organic carbon in soil profiles using vis-NIR spectroscopy on the Qinghai-Tibet plateau. *Environ. Sci. Technol.* **2015**, *49*, 4980–4987. [\[CrossRef\]](#)
46. Farifteh, J.; Van der Meer, F.; Van der Meijde, M.; Atzberger, C. Spectral characteristics of salt-affected soils: A laboratory experiment. *Geoderma* **2008**, *145*, 196–206. [\[CrossRef\]](#)
47. Piekarczyk, J.; Kaźmierowski, C.; Królewicz, S.; Cierniewski, J. Effects of soil surface roughness on soil reflectance measured in laboratory and outdoor conditions. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *9*, 827–834. [\[CrossRef\]](#)
48. Nocita, M.; Kooistra, L.; Bachmann, M.; Müller, A.; Powell, M.; Weel, S. Predictions of soil surface and topsoil organic carbon content through the use of laboratory and field spectroscopy in the Albany Thicket Biome of Eastern Cape Province of South Africa. *Geoderma* **2011**, *167–168*, 295–302. [\[CrossRef\]](#)
49. Rossel, R.A.V.; Cattle, S.R.; Ortega, A.; Fouad, Y. In situ measurements of soil colour, mineral composition and clay content by vis-NIR spectroscopy. *Geoderma* **2009**, *150*, 253–266. [\[CrossRef\]](#)

50. Vohland, M.; Ludwig, B.; Seidel, M.; Hutengs, C. Quantification of soil organic carbon at regional scale: Benefits of fusing vis-NIR and MIR diffuse reflectance data are greater for in situ than for laboratory-based modelling approaches. *Geoderma* **2022**, *405*, 115426. [\[CrossRef\]](#)
51. Liu, J.; Zhang, D.; Yang, L.; Ma, Y.; Cui, T.; He, X.; Du, Z. Developing a generalized vis-NIR prediction model of soil moisture content using external parameter orthogonalization to reduce the effect of soil type. *Geoderma* **2022**, *419*, 115877. [\[CrossRef\]](#)
52. Jin, P.; Li, P.; Wang, Q.; Pu, Z. Developing and applying novel spectral feature parameters for classifying soil salt types in arid land. *Ecol. Indic.* **2015**, *54*, 116–123. [\[CrossRef\]](#)
53. Wang, J.; Ding, J.; Yu, D.; Ma, X.; Zhang, Z.; Ge, X.; Teng, D.; Li, X.; Liang, J.; Lizaga, I.; et al. Capability of Sentinel-2 MSI data for monitoring and mapping of soil salinity in dry and wet seasons in the Ebinur Lake region, Xinjiang, China. *Geoderma* **2019**, *353*, 172–187. [\[CrossRef\]](#)
54. Eappen, G.; Shankar, T. Hybrid PSO-GSA for energy efficient spectrum sensing in cognitive radio network. *Phys. Commun.* **2020**, *40*, 101091. [\[CrossRef\]](#)
55. Wan, T.; Bai, Y.; Wang, T.; Wei, Z. BPNN-based optimal strategy for dynamic energy optimization with providing proper thermal comfort under the different outdoor air temperatures. *Appl. Energy* **2022**, *313*, 118899. [\[CrossRef\]](#)
56. Tsakiridis, N.L.; Keramaris, K.D.; Theocharis, J.B.; Zalidis, G.C. Simultaneous prediction of soil properties from VNIR-SWIR spectra using a localized multi-channel 1-D convolutional neural network. *Geoderma* **2020**, *367*, 114208. [\[CrossRef\]](#)
57. Xiao, D.; Vu, Q.H.; Le, B.T. Salt content in saline-alkali soil detection using visible-near infrared spectroscopy and a 2D deep learning. *Microchem. J.* **2021**, *165*, 106182. [\[CrossRef\]](#)
58. Romero, D.J.; Ben-Dor, E.; Demattê, J.A.; e Souza, A.B.; Vicente, L.E.; Tavares, T.R.; Martello, M.; Strabeli, T.F.; da Silva Barros, P.P.; Fiorio, P.R.; et al. Internal soil standard method for the Brazilian soil spectral library: Performance and proximate analysis. *Geoderma* **2018**, *312*, 95–103. [\[CrossRef\]](#)
59. Liu, Y.; Deng, C.; Lu, Y.; Shen, Q.; Zhao, H.; Tao, Y.; Pan, X. Evaluating the characteristics of soil vis-NIR spectra after the removal of moisture effect using external parameter orthogonalization. *Geoderma* **2020**, *376*, 114568. [\[CrossRef\]](#)