agronomy

MDPI

Article

A Study on Hyperspectral Soil Moisture Content Prediction by Incorporating a Hybrid Neural Network into Stacking Ensemble Learning

Yuzhu Yang [1], Hongda Li [2], Miao Sun [1], Xingyu Liu [3] and Liying Cao [1,*]

[1] College of Information and Technology, Jilin Agricultural University, Changchun 130118, China; yangyuzhu@mails.jlau.edu.cn (Y.Y.)
[2] College of Information and Electrical Engineering, China Agricultural University, Beijing 100091, China
[3] College of Statistics and Data Science, Nankai University, Tianjin 300071, China
* Correspondence: z2112020114@mails.jlau.edu.cn

Abstract: The accurate prediction of soil moisture content helps to evaluate the quality of farmland. Taking the black soil in the Nanguan District of Changchun City as the research object, this paper proposes a stacking ensemble learning model integrating hybrid neural networks to address the issue that it is difficult to improve the accuracy of inversion soil moisture content by a single model. First, raw hyperspectral data are processed by removing edge noise and standardization. Then, the gray wolf optimization (GWO) algorithm is adopted to optimize a convolutional neural network (CNN), and a gated recurrent unit (GRU) and an attention mechanism are added to construct a hybrid neural network model (GWO–CNN–GRU–Attention). To estimate soil water content, the hybrid neural network model is integrated into the stacking model along with Bagging and Boosting algorithms and the feedforward neural network. Experimental results demonstrate that the GWO–CNN–GRU–Attention model proposed in this paper can better predict soil water content; the stacking method of integrating hybrid neural networks overcomes the limitations of a single model's instability and inferior accuracy. The relative prediction deviation (RPD), root mean square error (RMSE), and coefficient of determination (R^2) on the test set are 4.577, 0.227, and 0.952, respectively. The average R^2 and RPD increased by 0.056 and 1.418 in comparison to the base learner algorithm. The study results lay a foundation for the fast detection of soil moisture content in black soil areas and provide a data source for intelligent irrigation in agriculture.

Keywords: soil moisture content; remote sensing; predictive model; machine learning techniques

Citation: Yang, Y.; Li, H.; Sun, M.; Liu, X.; Cao, L. A Study on Hyperspectral Soil Moisture Content Prediction by Incorporating a Hybrid Neural Network into Stacking Ensemble Learning. *Agronomy* **2024**, *14*, 2054. https://doi.org/10.3390/agronomy14092054

Academic Editor: Paul Kwan

Received: 31 July 2024
Revised: 29 August 2024
Accepted: 5 September 2024
Published: 8 September 2024

1. Introduction

Research on quick, precise, and reliable methods to determine soil composition is essential for precision agriculture [1]. Adequate soil moisture is conducive to plant growth and is one of the key indicators for evaluating soil fertility. However, too high or too low moisture content can affect the uptake of nutrients in the soil by the plant roots, thus affecting the yield and quality of crops. Remote sensing is an effective tool for tracking both temporal and spatial variations in crop morphology and physiological conditions, and it is one of the precision farming techniques. Hyperspectral imaging, a cutting-edge method, can capture a precise spectral response of a target's features [2]. Therefore, it is crucial to precisely quantify the soil moisture content (SMC) and analyze the hyperspectral response pattern of SMC.

Hyperspectral analysis detection can be conducted directly without destroying the sample, which has the characteristics of easy detection and high efficiency. Nowadays, hyperspectral detection has been widely used to detect the internal related content of soil, food, crop plants, petroleum, etc. Hyperspectral remote sensing data have a high spatial dimension and high correlation [3], and it can be used to obtain fine spectral profiles of soil

in a certain range. It has a high spectral resolution that allows for a large number of bands and leads to a dramatic increase in the amount of spectral data. Hyperspectral data contain not only information on SMC but also a lot of redundant and invalid information, which affects the model prediction. Feature wavelength selection algorithms reduce data redundancy and multicollinearity by selecting the most informative wavelength combinations. The use of characteristic wavelengths for quantile regression modeling can reduce model complexity. Commonly used hyperspectral feature wavelength selection methods include the successive projection algorithm [4], competitive adaptive reweighted sampling [5,6], and uninformative variable elimination [7].

Many previous studies on the quantitative inversion of SMC chose linear models [8,9] for training, and most of them were based on partial least squares regression (PLSR) [10–12], multiple linear stepwise regression [13], support vector regression (SVR) [14], spectral transformations [15–17], etc. However, the model's robustness needs to be improved, and model redundancy should be reduced. In recent years, with the rapid development in machine learning and its superiority in solving nonlinear problems, researchers have employed machine learning techniques to categorize or regress hyperspectral data, thereby optimizing the inversion accuracy and model stability during the spectral inversion of soil property components. For instance, SMC was predicted using a neural network, Random Forest (RF), and extreme learning machine (ELM) [18], and the findings revealed that all three machine learning techniques obtained better prediction results than PLSR. The extremely randomized trees' (Extra Trees) method was used to build a regression model for SMC based on hyperspectral data [19], and the model obtained better prediction results than SVR. There are also related studies using RF and Extra Trees to predict soil salinity content, and good results have been obtained [20].

At present, a hot research area in machine learning is deep learning, which simulates the interaction between neurons in the human brain, recognizes patterns, and performs classification and prediction by training with a large amount of data. With the continuous advancement in deep learning technology and the improvement in data processing capabilities, studies have been conducted to use deep learning for remote sensing image classification [21–24], but there are very few applications in hyperspectral remote sensing for predicting SMC. Using hyperspectral data from soil samples in Xinjiang, Wang et al. [25] constructed a hybrid neural network model with a long short-term memory (LSTM) convolutional neural network (CNN) to identify the moisture and organic matter content; AHMED et al. [26] designed a hybrid model using GRU, adaptive noise full ensemble empirical modal decomposition, and CNN for daily time-step surface soil moisture forecasts. According to the investigation, convolutional neural networks and hybrid neural networks may accurately predict soil moisture and quantitatively estimate it.

The swarm intelligence algorithm has been widely applied extensively to numerous intricate optimization issues [27]. The optimization of neural network models using both a genetic algorithm (GA) and particle swarm optimization (PSO) has been used on soil moisture inversion problems [28,29]. In certain research, the GWO algorithm, however, is more capable of adapting to the issue [30]. When compared to PSO and GA, GWO is more adept at identifying the optimal solution, more robust, and less prone to falling into the local optimal solution. A GWO-SVM model was developed by Zhang et al. [31] to differentiate between the severity of cotton crown wilt. Better classification accuracy than the traditional machine learning model was also shown in a hyperspectral image classification study when the CNN model's hyperparameters were optimized using GWO [32]. It is evident that the GWO algorithm performs exceptionally well when it comes to image classification. In previous studies, the hybrid neural network with the GWO algorithm has rarely been utilized to forecast soil moisture, but this combination also has outstanding potential in prediction issues.

Ensemble learning, a recent development trend in machine learning, enhances a single method's generalization capacity by combining the findings of various learning approaches, thereby enhancing prediction accuracy and improving generalization performance.

In the past two years, studies have been conducted to investigate the hyperspectral remote sensing inversion of soil-related component content using Boosting and Bagging models [33–37]. Among them, both AdaBoost and GBRT algorithms outperformed PLSR and SVR in the inversion of soil fast-acting potassium [38]; The chlorophyll content of mangrove canopy and spring maize leaves can be detected with the stacking model [39,40], and remarkable results were obtained. Feng et al. [41] combined three commonly used machine learning methods, RF, SVR, and k-nearest neighbor, to develop a stacking model for alfalfa yield prediction. Ensemble learning has achieved high modeling accuracy and good prediction results in all the above studies, and it is a proven method in hyperspectral prediction models. Meanwhile, in the prediction of soil moisture, some studies have combined PLSR and SVR or Boosting and Bagging models by stacking to predict soil moisture in Vertisol soil [42] and grape-growing area soil [43]. Some studies have also achieved a better prediction of the soil moisture content of over-partial-vegetation-covered surfaces using CNN + LSTM [44], which also shows that deep learning can improve the prediction of SMC.

Therefore, this paper combines the designed GWO–CNN–GRU–Attention with the stacking model to build a model for estimating the SMC of the study area. Compared with previous related studies, the model proposed in this paper, which incorporates a deep learning model into the base learner, improves the type of base learner and enhances the stacking model's prediction ability. In particular, the approach used in this paper overcomes the drawbacks of a single inversion model and improves the model's generalizability as compared to earlier related works. After comparing the results of a feedforward neural network, GWO–CNN–GRU–Attention, and various Boosting and Bagging algorithms, the validity and stability of the model proposed in this paper are confirmed. Finally, the best prediction model for SMC in the study area is determined, further improving the hyperspectral prediction accuracy and stability of SMC.

2. Materials and Methods

2.1. Study Area

This experimental study was conducted on the terrace of the Yitong River in the hinterland of the Songliao Plain in Nanguan District, Changchun City, Jilin Province, China, in the mid-latitude northern temperate zone of the Northern Hemisphere. The area has a temperate continental semi-humid monsoon climate, characterized by a cold and dry winter, a cool and brief summer, and frequent cold waves in the spring. The annual average temperature is 4.9 °C, and the annual average precipitation is 593.8 mm. The study area (125°24′19.04″–125°24′27.33″ E, 43°49′24.50″–43°49′34.25″ N and 125°24′34.58″–125°24′39.45″ E, 43°49′26.05″–43°49′34.91″ N) is located at the teaching and research base of Jilin Agricultural University (Figure 1). The soil type is meadow black soil, and the main crops are corn, soybean, and rice.

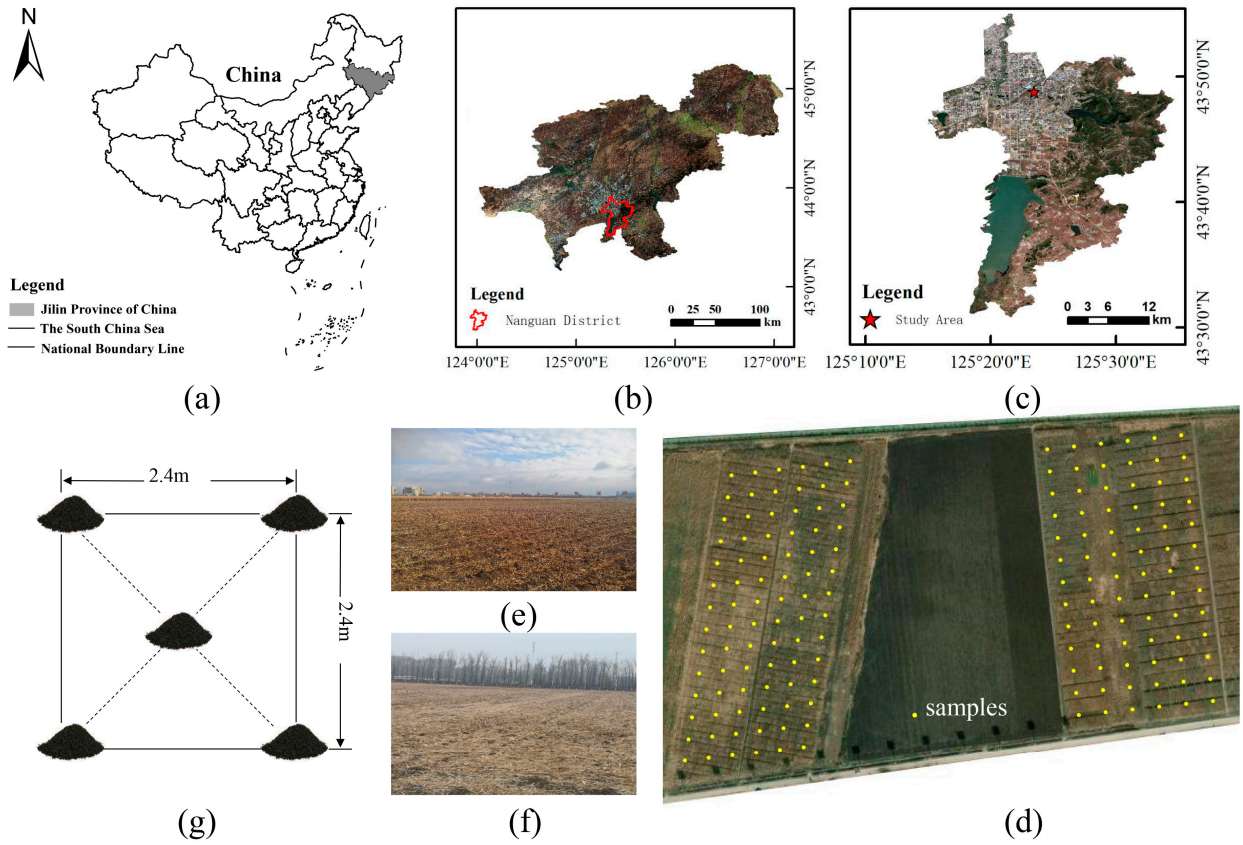

Figure 1. (**a**–**c**) The geographical location of Nanguan District, Changchun City, Jilin Province, China. (**d**) The distribution of 162 sample points in the study area. (**e**) Soil sampling reality of soybean-planted fields. (**f**) Soil sampling reality of corn-planted fields. (**g**) Five-point sampling method.

2.2. Soil Sample Collection and Determination

In the crop harvest period in October 2021, 162 soil samples were obtained by utilizing a five-point sampling technique. The sample areas were divided evenly by taking 6 ridges in each of the corn- and soybean-planting areas, with 13 sample points per ridge for corn and 14 sample points per ridge for soybean. In total, 84 samples of soil were taken from the soybean-planting region, and 78 soil samples were taken from the corn-planting region. Specifically, the soil at a depth of 5–20 cm was taken, the sampling point and the four corners of soil centered on it were mixed, and some of the soil was gathered and kept in sealed bags. The soil samples were dried under air naturally in a ventilated room for one week. The dried soil samples were placed in a mortar, then manually ground, and finally filtered through a 100-mesh sieve. The sieved soil was divided into two portions for soil spectral data collection and moisture determination (Figure 2). In the drying and weighing process, the relative moisture content of the sample soil was obtained. The samples' estimated SMC ranged from 2.25 to 6.95 percent. The SMC of the soybean-planting region was generally higher than that of the corn-planting region. This is because soybean consumes less water and utilizes less water than corn [45], and this is consistent with the findings of existing studies.

Figure 2. Soil sample processing flow. (**a**) Dried soil samples are ground, filtered, and divided into two portions. (**b**) Soil spectral data collection. (**c**) Soil moisture determination.

2.3. Measurement and Pre-Processing of Hyperspectral Data

The FieldSpec3 portable spectrometer, manufactured by ASD Inc. located in Santa Clara, CA, USA, was used for the spectral measurement of soil samples. It is a passive sensor. The wavelength range is 350–2500 nm. The indoor spectral measurements are characterized by easy control and high-quality data results. The soil was placed in a Petri dish with a black body, and a steel ruler was used to scrape the surface layer of the soil. In the dark room, the light source was a 50 W halogen bulb, with the irradiation direction at an angle of 45° from the vertical, and the soil sample's surface was about 100 cm away from the light source. The spectrometer probe was positioned vertically above the soil surface, and the probe was approximately 10 cm from the soil sample. A standard white shift correction was performed on every 5 samples during the measurement. The soil hyperspectral data measured by the spectrometer were processed and output using ViewSpecPro5.6.8 software. As illustrated in Figure 3, the spectral curves of each soil sample over five successive time periods were taken to reduce measurement data errors and instability. Then, the average spectrum was used as the soil's original spectrum.

Due to the instrument and the inevitable influence of the test environment, sample background, observation angle, sample roughness, stray light, and other factors in the spectral acquisition process, the noise in the edge band of the spectral curve is relatively large. To reduce the interference of external noise on the modeling inversion result, the noisy fringe bands 350–399 nm and 2400–2500 nm were removed for each soil sample, and the band from 400 to 2399 nm was reserved for the modeling analysis. Then, the data were standardized so that they obeyed a normal distribution with a mean of 0 and a variance of 1.

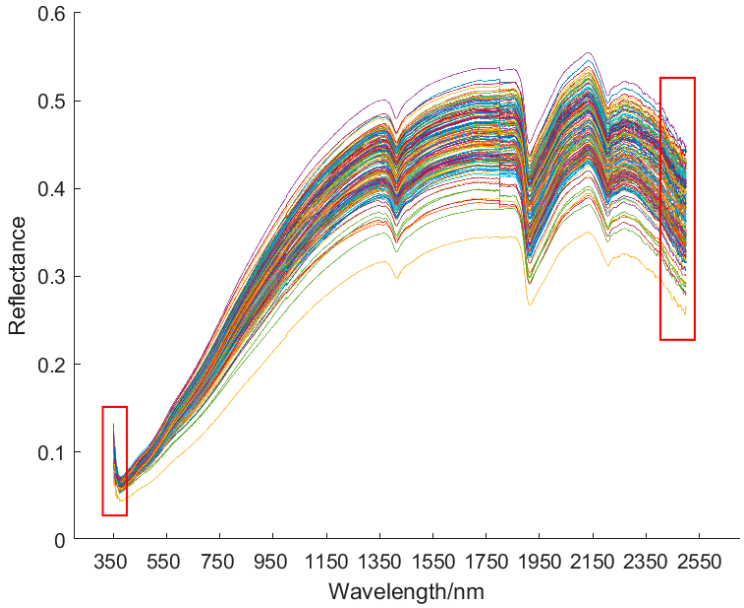

Figure 3. Original spectra. The red rectangles surround the noisy fringe bands 350–399 nm and 2400–2500 nm.

2.4. Research Methods

2.4.1. Overview of GWO–CNN–GRU–Attention Model

(1) Convolutional Neural Network

A convolutional neural network (CNN) is one of the most extensively studied and applied models in the recent history of deep learning. It alternates convolutional and pooling layers to provide effective representations of raw data using local connections and shared weights. By automatically extracting the local features of the data, a dense and comprehensive feature vector is created [46]. CNN is chosen in this paper to extract spectral data features. After inputting the spectral data into the model, the data dimensions are adjusted using a combination of convolution and pooling layers for deep feature extraction.

(2) Gray Wolf Optimization

The gray wolf optimization (GWO) algorithm involves information feedback mechanisms and convergence factors that can be adaptively changed to strike a balance between a local search and global search. Because of this, GWO in predicting soil moisture content not only ensures the prediction accuracy, but also will enable the model to achieve faster convergence. In this paper, GWO is used to determine the unit parameters and the number of epochs.

(3) Gated Recurrent Unit

The GRU model is a type of recurrent neural network (RNN) that is widely used to process sequential data. GRU mixes cell states and hidden states and adds some other changes. In this study, the GRU model can be used for SMC prediction. It was found experimentally that increasing the GRU network depth helps to improve the model's capacity for prediction, but it also increases the complexity of the model. Thus, the final model proposed in this paper has two GRU network layers.

Figure 4 shows the structure of the GRU model, which is composed of the reset gate and update gate. The reset gate (r_t) controls how the new input data are merged with the old memory and how much of the hidden state output from one moment goes into the candidate hidden state for the next. As the reset gate's value decreases, less information enters, and more prior information is forgotten. The update gate (z_t) controls how much state data from the previous instant are brought into the present state by defining the quantity of past memory saved to the current time step. As the update gate's value

increases, more state data from the previous moment are brought in. The specific calculation of the GRU model is shown in (1), (2), (3), and (4):

$$r_t = \sigma(w_r \cdot [h_{t-1}, x_t] + b_r) \quad (1)$$

$$z_t = \sigma(w_z \cdot [h_{t-1}, x_t] + b_z) \quad (2)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (3)$$

$$\tilde{h}_t = tanh\left(w_{\tilde{h}} \cdot [r_t \odot h_{t-1}, x_t]\right) \quad (4)$$

where σ denotes the sigmoid function, r denotes the reset gate vector, z denotes the update gate vector, tanh is the tanh activation function, t is the moment, x is the input vector, and h and $\tilde{h}$ denote the hidden state and the candidate hidden state, respectively.

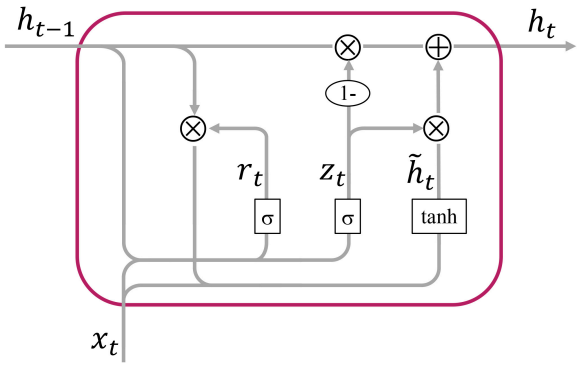

Figure 4. The structure of the GRU model.

(4) GWO–CNN–GRU–Attention

Figure 5 shows the structure of the GWO–CNN–GRU–Attention model proposed in this paper. The model mainly consists of a GRU and a CNN, where the CNN extracts features from the data, and the GRU is primarily responsible for predicting SMC. The two convolutional layers (Conv) of the CNN are configured with 32 and 64 convolutional kernels, respectively. By mining pertinent data in accordance with the size of the kernels, CNN can extract useful features from the input data. The activation function is ReLU, and the convolution kernel size is 3. After mapping through the features in the convolution layer, the output dimension size is reduced using maximum pooling. In the two pooling layers (Pool), the pooling window is set to 2, the step size is set to 2, and the fill method is selected with the input and output images of the same size. A 1 × 500 × 64 tensor is obtained after 2 convolutions and pooling, where 64 is the number of channels.

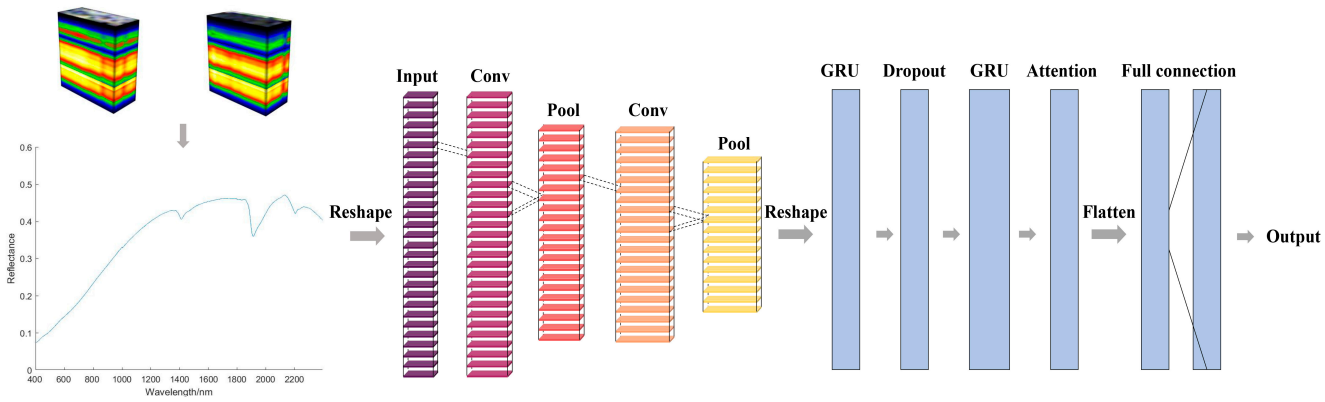

Figure 5. GWO–CNN–GRU–Attention model.

To avoid model overfitting, a random deactivation technique is applied between the two GRU network layers. After the GRU network layer, an attention mechanism is introduced to further extract features to enhance the precision of the prediction result. Next, a one-dimensional vector array with a length of 50 is output using a flattening operation. Finally, a fully connected layer outputs the predicted SMC.

2.4.2. Overview of the Stacking Model

(1) Boosting and Bagging algorithm

Boosting and Bagging algorithms in ensemble learning are widely used because they can contribute to better generalization performance and reduce the error rate by constructing multiple weak learners and combining multiple learners using some strategy. In this study, two different forms of Bagging algorithms, namely, Random Forest (RF) and Extra Trees, are employed. There are five different Boosting techniques in use: the extreme Gradient Boosting (XGBoost) algorithm, the light Gradient Boosting Machine (LightGBM) algorithm, the Gradient-Boosted regression tree (GBRT), and the Gradient Boosting + Categorical Features (CatBoost).

AdaBoost, GBRT, XGBoost, LightGBM, and CatBoost are all iterative Boosting-based algorithms. A decision tree serves as the weak learner in the Boosting method, while least squares is used as the loss function. AdaBoost and Gradient Boosting are the two primary implementations. Specifically, AdaBoost trains many weak learners by varying the sample weights; Gradient Boosting trains multiple weak learners by varying the objective function; finally, the serial technique is used to transform the weak learners into strong learners. GBRT, XGBoost, LightGBM, and CatBoost are all specific implementations of Gradient Boosting algorithms.

In Bagging ensemble learning, this study builds multiple weak learners in parallel to generate different training sets using sampling and thus train different models. Typically, the weak learners use decision trees, but other nonlinear algorithms are also possible. Finally, the results of multiple weak learners are combined for output. When a regression task is the objective of the ensemble learning process, the output is the average of the weak learner's outputs. Essentially, Bagging exploits the model's diversity to improve the algorithm's overall efficacy, and it focuses on the generation of different training sets. In this study, a method called Bootstrap (i.e., repeated random sampling with put-backs) is used, which generates different datasets. RF is fundamentally superior to single decision trees because of its exceptionally powerful learning capability, high algorithm complexity, and resistance to overfitting. The Extra Trees model introduces a greater degree of randomness in node division. A subset of features is randomly selected for feature selection for branching at each split or branching. Meanwhile, it does not need to choose the optimal threshold, and it uses a random threshold for branching.

(2) Feedforward neural network

The feedforward neural network, commonly referred to as multilayer perceptron (MLP), has input and output layers as well as one or more hidden layers.

A conventional MLP has a three-layer structure and only one hidden layer. The neurons have nonlinear activation functions and connect the layers in all directions. In the hidden layer, each neuron transmits parameters from the input layer to the output layer, and its weights are adjusted backward by using the gradient descent approach and the loss function to minimize the difference between the anticipated and actual values. The activation function chosen in this study is tanh, the hidden layer contains 100 neurons, the number of training iterations is 400, and the Adam gradient descent function is employed. The calculation of each neuron is shown in Equation (5), where n denotes the amount of input data, and I_i denotes the input data; β_j denotes the deviation, and $\omega_{i,j}$ denotes the weight of the connection.

$$f(x)_j = \sum_{i=1}^{n} \omega_{i,j} I_i + \beta_j \quad (5)$$

The tanh activation function is shown below:

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (6)$$

(3) Stacking

Since a single model can have different degrees of variation in model structure and prediction bias, and model prediction can be unstable, this study chooses the stacking ensemble learning model. The degree of difference in the model structure and the prediction bias of the combined learner are fully utilized to make up for the shortcomings of the base learner and ensure the effectiveness of ensemble learning.

Stacking is an ensemble learning fusion strategy that has been proven to be effective in improving model prediction accuracy. Multiple base learners are trained with the original data, and the prediction results on the resultant training set and test set are, respectively, used as the input training set and test set for the following layer of learners. Finally, the strong learner with better prediction performance is trained. The base learners' prediction results are compiled by the strong learner, also referred to as the meta-learner, and proportional weights are given to the base learners to increase the model's prediction accuracy.

In most cases, stacking involves heterogeneous weak learners, indicating that many machine learning methods are used as base learners. To create more precise and reliable models, each base learner is trained with data using K-fold cross-validation techniques.

In this paper, a two-layer stacking ensemble learning model is constructed, and an SMC inversion algorithm is designed with the above-mentioned Boosting, Bagging model, MLP, and GWO–CNN–GRU–Attention as base learners and linear regression as the meta-learner. The prediction of SMC is performed on the soil hyperspectral data of the study area. Figure 6 shows the structure of the designed framework in detail.

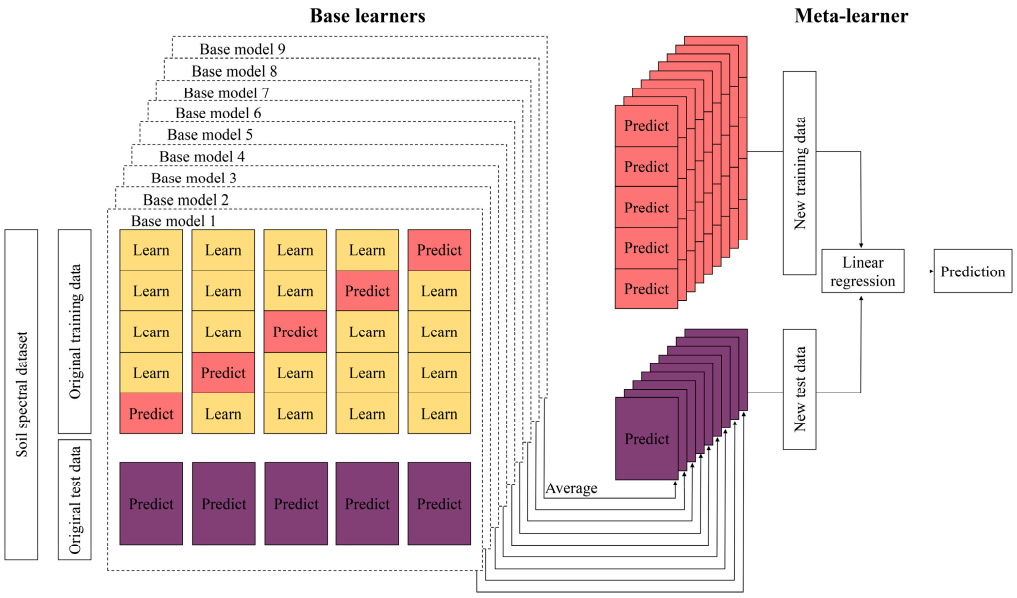

Figure 6. Structure diagram of stacking model.

The precise steps of the SMC inversion algorithm are as follows: (1) The initial dataset is proportionally split into training and test sets. (2) Five-fold cross-validation is applied to each base learner. One copy is used as the test set, and the remainder is used as the training set for each base learner. Then, the prediction results of all base learners are taken as the training set for the meta-learner. (3) A new test set is created by averaging the prediction results using the previous test set and the base learner. (4) To obtain the final prediction results, the new training set and the new test set are fed into the meta-learner for training.

2.5. Evaluation Method

In this study, the performance of the prediction model is evaluated using the coefficient of determination (R^2), root mean square error (RMSE), and relative prediction deviation (RPD). The calculation formulas of these evaluation metrics are shown in (7), (8), and (9), respectively:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (8)$$

$$RPD = \frac{SD}{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2} \Big/ \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (9)$$

where n represents the sample size; y_i represents the measured value of the i-th sample; $\hat{y}_i$ and $\overline{y}$ denote the model's predicted value for sample i and the sample's mean measured value, respectively; and SD represents the standard deviation of the sample's measured values. The model fits the data better when the R^2 is close to 1, and the model is more stable when the RMSE is smaller. When RPD is greater than 3.0, it indicates that the model has very high reliability and excellent prediction performance; when RPD ranges between 2.5 and 3.0, it indicates that the model has good prediction ability; when RPD ranges between 2.0 and 2.5, it indicates that the model has prediction ability and can make approximate predictions; when RPD ranges between 1.5 and 2.0, it indicates that the model has the potential to distinguish between high and low values and can be improved by other methods; when RPD is smaller than 1.5, it indicates that the model is unreliable, and the prediction fails [47]. As a result, models that perform well in this study should have large R^2 and RPD values along with small RMSE values.

3. Results

3.1. Sample Analysis

The collected 162 soil samples were randomly divided into a training set and a test set at a ratio of 8:2 using the joint x-y distance (SPXY) approach, with 130 soil samples in the training set and 32 soil samples in the test set. The moisture content of the soil at each sampling site is shown in Figure 7. It can be seen that the SMC varies in a gradient. The statistical characteristics of the SMC for the training set, the test set, and all samples are presented in Table 1. Three of the sample sets have relatively similar means and standard deviations, indicating that they have a similar data distribution. As a result, the distributional features of the entire dataset can be accurately represented in both the training and testing sets.

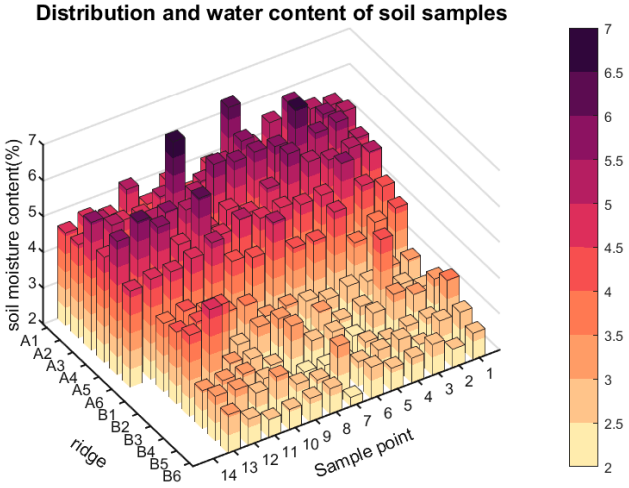

Figure 7. The SMC at the sampling sites.

Table 1. The statistical characteristics of the SMC for each dataset.

Dataset	Number of Samples	Maximum (%)	Minimum (%)	Mean (%)	Standard Deviation (%)	Coefficient of Variation
Total	162	6.951	2.254	4.099	1.171	0.286
Training set	130	6.951	2.254	4.168	1.154	0.277
Testing set	32	5.800	2.489	3.817	1.201	0.315

3.2. Prediction Results of the GWO–CNN–GRU–Attention Model

To construct the hybrid neural network model, the fitting effects of CNN, GWO-CNN, GWO-CNN-GRU, GWO–CNN–Attention, and GWO–CNN–GRU–Attention were compared and analyzed in the experiments, and the fitting plots of the prediction results were drawn using the predicted and measured values of the model (Figure 8).

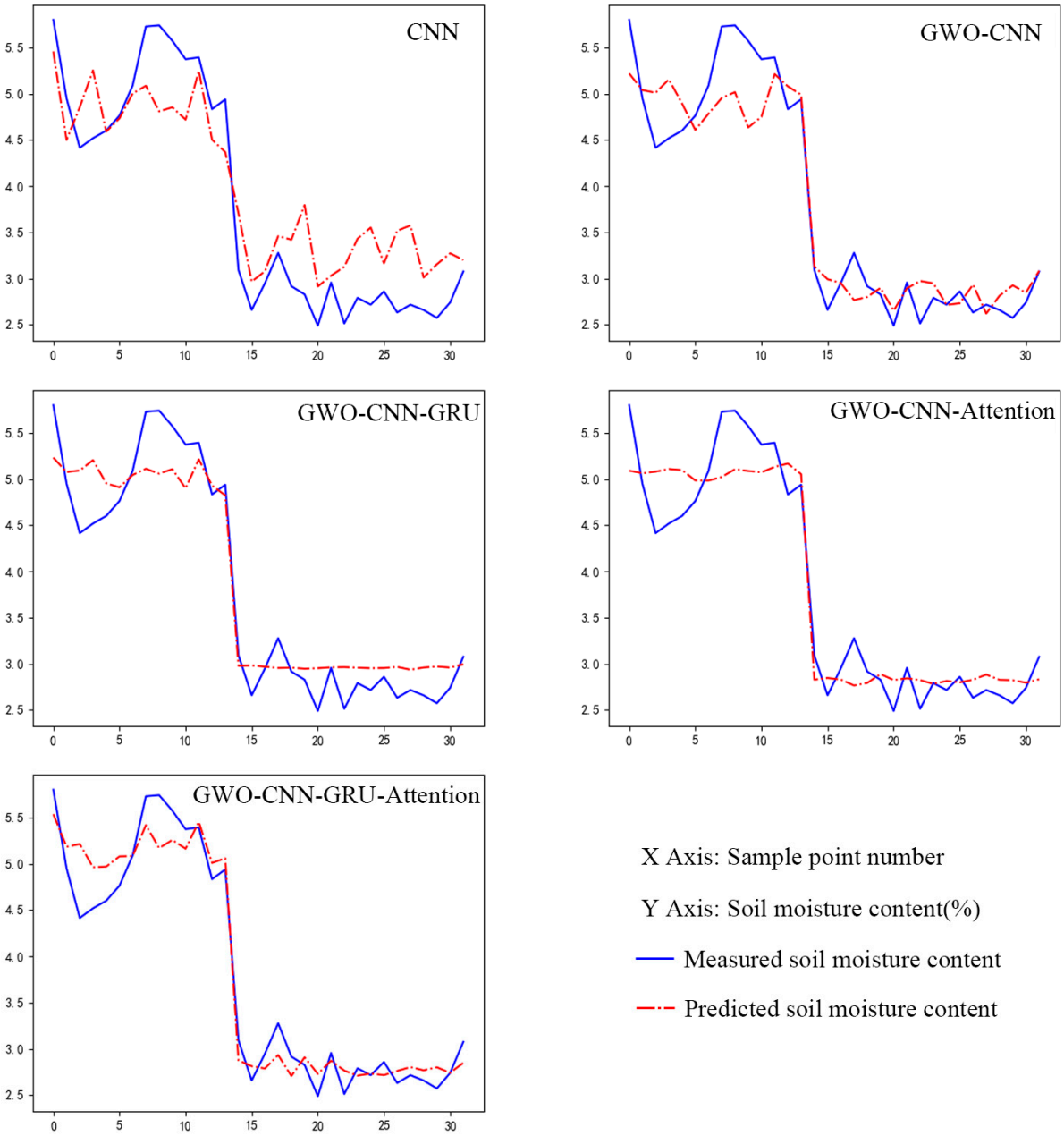

Figure 8. The predicted and measured values of GWO–CNN–GRU–Attention and ablation models.

Table 2 presents the prediction results for each model's estimation of SMC. On the test set, the prediction accuracies for the CNN and GWO-CNN models were 79.4% and 89.7%, respectively; the GWO-CNN model performed better than the CNN model. The GWO-CNN model's prediction accuracy increased to 91.4% after the GRU model was included. The GWO-CNN model's prediction accuracy increased to 91.6% after the attention mechanism was included. It can be seen that the model's performance was improved with the inclusion of the GRU model or the attention mechanism. Finally, when both the GRU and attention mechanism were included, the accuracy of the GWO-CNN model on the test set was 92.9%. The GWO–CNN–GRU–Attention model obtained the highest R2 and RPD as well as the lowest RMSE when compared to the other four models. It demonstrates that the GWO–CNN–GRU–Attention model developed in this study has improved precision and accuracy for SMC prediction.

Table 2. The prediction results of GWO–CNN–GRU–Attention and ablation models.

Model	Training Set			Testing Set		
	R^2	RMSE	RPD	R^2	RMSE	RPD
CNN	0.502	0.706	1.416	0.794	0.472	2.205
GWO-CNN	0.814	0.431	2.321	0.897	0.333	3.122
GWO-CNN-GRU	0.784	0.466	2.149	0.914	0.305	3.409
GWO–CNN–Attention	0.784	0.465	2.153	0.916	0.302	3.445
GWO–CNN–GRU–Attention	0.872	0.373	2.792	0.929	0.270	3.779

3.3. Prediction Results of the Stacking Model

In our experiments, the stacking model containing the hybrid neural network is called Stacking2, which has nine base learners. The stacking model that does not contain the hybrid neural network is called Stacking1, and it has eight base learners. Table 3 presents the prediction results for each base learner and stacking technique.

Table 3. The prediction results for stacking and each base learner.

Model	Training Set			Testing Set		
	R^2	RMSE	RPD	R^2	RMSE	RPD
XGBoost	0.999	0.013	79.478	0.895	0.338	3.082
MLP	0.863	0.370	2.702	0.915	0.303	3.430
RF	0.967	0.181	5.531	0.908	0.316	3.292
LightGBM	0.987	0.115	8.692	0.870	0.376	2.772
GBRT	0.999	0.003	35,718.721	0.849	0.405	2.571
AdaBoost	0.896	0.322	3.093	0.916	0.302	3.450
CatBoost	0.998	0.037	27.043	0.905	0.321	3.245
Extra Trees	1.0	0.001	7.417×10^{14}	0.873	0.371	2.807
GWO–CNN–GRU–Attention	0.872	0.373	2.792	0.929	0.270	3.779
Stacking1	0.892	0.328	3.047	0.920	0.294	3.536
Stacking2	0.923	0.301	3.590	0.952	0.227	4.577

According to the prediction results of the Stacking2 model, the R^2 of the nine base learner models on the test set are ranked in descending order: GWO–CNN–GRU–Attention, AdaBoost, MLP, RF, CatBoost, XGBoost, Extra Trees, LightGBM, and GBRT. GWO–CNN–GRU–Attention achieves the largest R^2 of 0.929 and GBRT obtains the smallest R^2 of 0.849. Stacking2 achieves the highest goodness of fit with an R^2 of 0.952 compared to the nine base learner models.

Among the nine base learner models, GWO–CNN–GRU–Attention obtains the smallest RMSE of 0.270 and is the only model with RMSE below 0.3, while GBRT obtains the largest RMSE. The RMSE of the Stacking2 model is 0.227, which is lower than that of the other models. Through the analysis of the RPD, it can be seen that only the LightGBM,

GBRT, and Extra Trees models have RPD values between 2.5 and 3.0, indicating the good prediction ability of these models. The RPD values of other models are above 3.0, and the Stacking2 model has excellent prediction ability with an RPD value of 4.577. The average R^2 and RPD of the Stacking2 model were improved by 0.056 and 1.418, respectively, and its average RMSE decreased by 0.107 compared with that of the nine base learner models on the test set. The prediction performance of Stacking2 is significantly improved, and it shows a strong prediction capability in this experiment.

The prediction results of Stacking1 and Stacking2 outperform the base learners in all three evaluation metrics, indicating that the stacking ensemble learning method used in this study is effective and improves the prediction accuracy. However, the Stacking2 model demonstrates a stronger enhancement effect. Compared with the Stacking1 model, Stacking2 improves R^2 and RPD by 3.2% and 1.041 and reduces RMSE by 0.067. It can be seen that adding GWO–CNN–GRU–Attention to stacking greatly improves the performance of the stacking model.

Figure 9 illustrates the comparison between the predicted and measured values of the two stacking models. Also, scatter plots were plotted for the two stacking models in Figure 9. The curve fit of the predicted and measured values is best for the Stacking2 model. Furthermore, the points created using the Stacking2 model's projected values are substantially closer to the grey regression line, indicating the model's higher predictive power. Figure 10 intuitively shows that Stacking2 obtains the highest R2 and RPD and lowest RMSE values, indicating that this model performs best in SMC prediction.

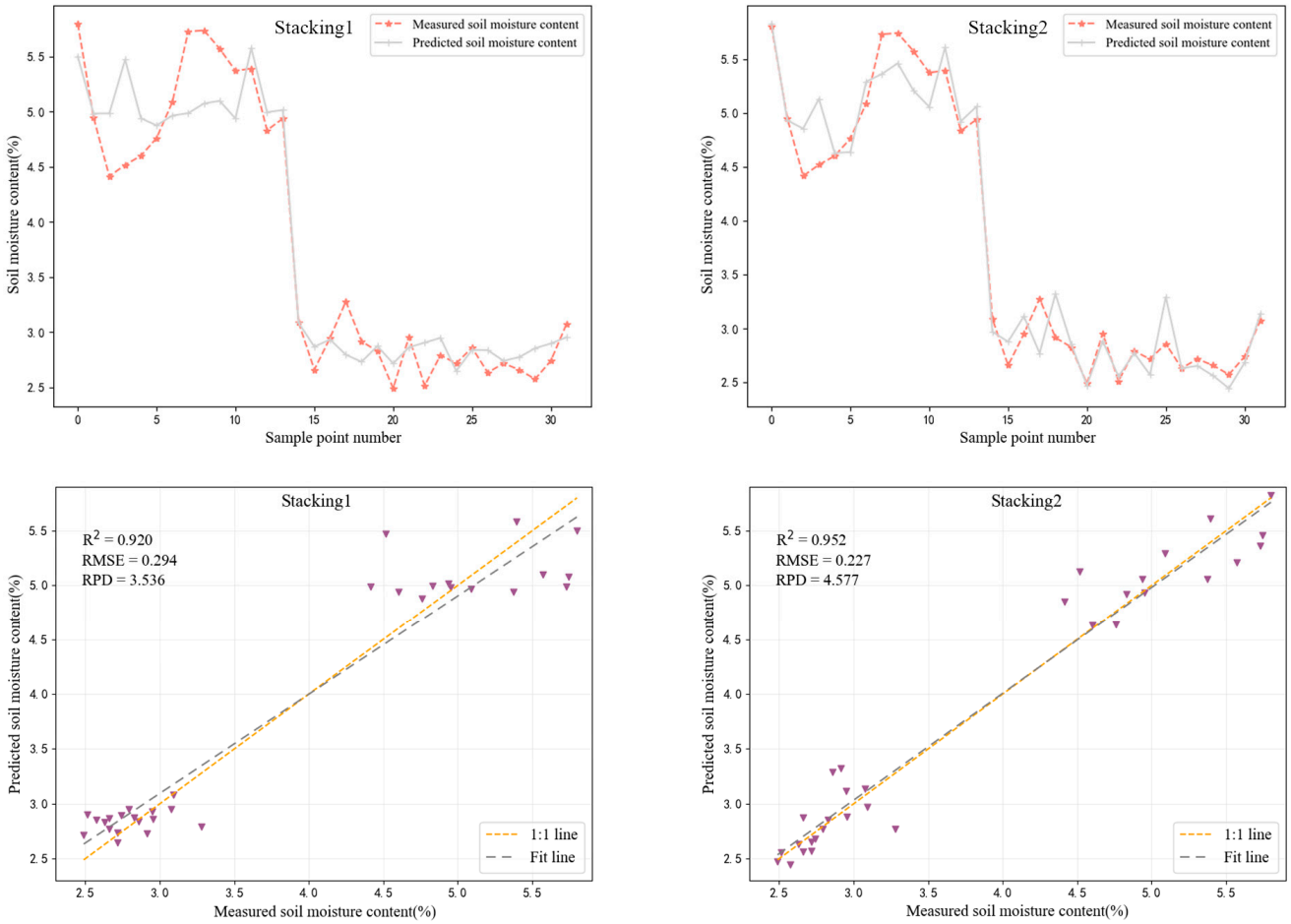

Figure 9. Measured and predicted values of stacking model.

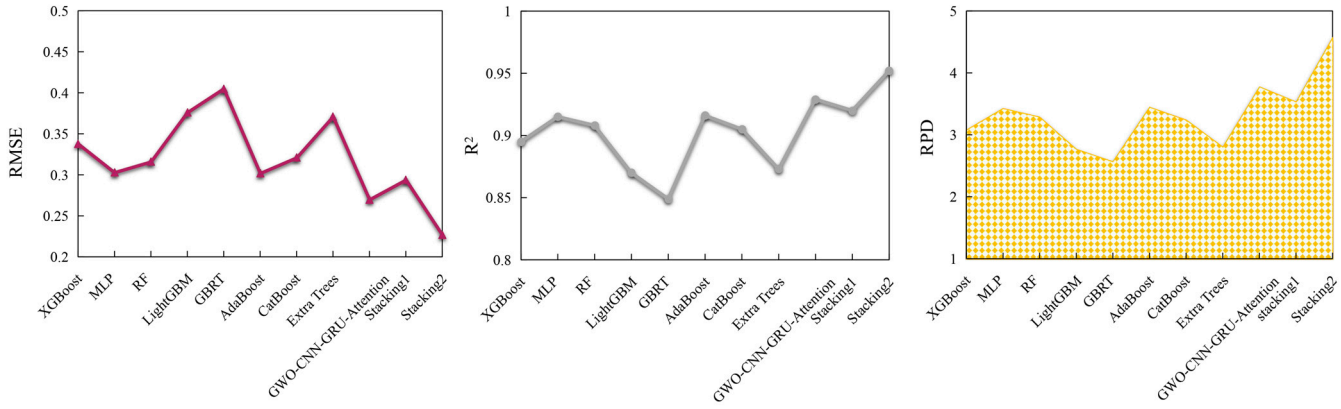

Figure 10. Comparison of evaluation indexes of each prediction model.

In the experiments, in addition to the stacking model's strong practical application capability, the base learners also obtained good prediction results, among which GWO–CNN–GRU–Attention and MLP both have potent predictive power. However, the use of the neural network model may somewhat hamper the performance due to the single soil type and the limited sampling area. The effectiveness of the neural network model lies in its requirement for a large amount of training data. Therefore, there is still a great opportunity for advancement in the prediction of SMC using neural networks, particularly CNNs and hybrid neural networks.

4. Discussion

In recent years, soil moisture inversion based on hyperspectral remote sensing has become a hot research topic [44,48–50]. One part of the study by Zhang et al. [51] is very similar to ours, which demonstrated the effectiveness of the 1DCNN model in predicting soil properties. Similar to what KARA et al. [52] found, combining RNN with an attention mechanism and utilizing optimization techniques to tune the hyperparameters is a great way to predict soil moisture. As envisioned in the above-mentioned article, this study integrates CNN with this method to achieve excellent performance in soil moisture prediction.

Prior research has also focused a great deal of attention on the stacking model employed in this work. Wang et al. [53] demonstrated in their article that the stacking model outperformed the single GBRT model. In addition, Ge et al. [54] used a hyperspectral sensor carried by a UAV to acquire hyperspectral images of the study area, and used Boosting and Bagging algorithms to construct a soil moisture prediction and analysis model for the 70 pieces of data acquired. Using R^2 and RPD as evaluation metrics, the model achieved an excellent performance of 92.6% and 2.556. In related works, the stacking method is utilized to incorporate machine learning models to predict soil moisture content and is also gaining popularity. The stacking approach was used to integrate CatBoost, RF, and GBRT models to predict the soil of planted grapes [43], and the R2 reached 0.86, which is a superior prediction compared to other models. However, GWO–CNN–GRU–Attention is a deep learning model, and the deep learning model added to the stacking model design in this study enriches the type of base learner, retains richer spectral properties in the prediction, and is a more robust model. A more advanced moisture prediction model was constructed for more than 160 data values and significantly outperformed the above methods in terms of R^2 and RPD. This result is attributed to the hybrid neural network and the design approach of using it as one of the stacking model-based learners in this paper, which effectively reduces the risk of overfitting in a single model and enhances the tolerance to noise and anomalous data.

In addition, Wang et al. [55] proposed an innovative method of SSA-CNN to invert soil moisture in farmland areas using 200 multi-source remote sensing data values. There is also a study that designed an SCSANet deep learning network combined with a self-attention mechanism for hyperspectral feature extraction, and established a prediction model for soil

metal content for 500 spectral data values [56]. In contrast, although the method proposed in this study yielded more reliable results, it should be noted that this paper is not as adequate as the above studies in terms of the amount of data, which also constitutes a limitation of this paper. And the efficiency of the deep learning model is also an issue worth exploring and improving. This is because the inclusion of a hybrid neural network increases the computational load.

5. Conclusions

Based on the field collection of soil samples and various modeling approaches, this study focused on using soil reflectance spectra to realize the better inversion of soil moisture content, and the conclusions are presented below:

(1) The GWO–CNN–GRU–Attention method improves the accuracy of hyperspectral soil water content inversion, compared to CNN, GWO-CNN, GWO-CNN-GRU, and GWO–CNN–Attention.

(2) Nine models from Boosting, the Bagging ensemble learning algorithm, the feed-forward neural network, and GWO–CNN–GRU–Attention are utilized as the base learners of the stacking model. This model exhibits strong stability and predictive performance in this study. The R^2, RMSE, and RPD values of the stacking model on the test set are 0.952, 0.227, and 4.577, respectively. It achieves the best evaluation index and the highest inversion accuracy when compared with other models. In conclusion, the stacking ensemble learning model incorporating the hybrid neural network greatly improves the prediction of hyperspectral soil moisture content.

The findings demonstrate that the stacking model proposed in this study can meet actual prediction needs, significantly increase the precision of soil moisture content inversion using hyperspectral data, and serve as a foundation for the inversion of soil moisture content in areas with black soil. In future work, we will research multi-region and multi-species soil moisture content prediction models to provide technical support for large-scale irrigation monitoring on farmland.

Author Contributions: Y.Y.: Data curation, Conceptualization, Methodology, Writing—original draft. H.L.: Conceptualization, Methodology, Writing—review and editing. M.S.: Data curation, Writing—review and editing. X.L.: Data curation, Writing—review and editing. L.C.: Writing—review and editing, Supervision, Project administration. All authors have read and agreed to the published version of the manuscript.

Funding: This article was supported by the National Natural Science Foundation of China (Grant No. U19A2061); agricultural image recognition and processing team, Jilin Provincial Science and Technology Department of young and middle-aged scientific and technological innovation and entrepreneurship excellence talent (team) project (innovation category) (Grant No. 20220508133RC); and Jilin Province Science and Technology Development Plan Project (Grant No. 20210404020NC).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Ge, J.; Thomasson, A.; Sui, R. Remote Sensing of Soil Properties in Precision Agriculture: A Review. *Front. Earth Sci.* **2011**, *5*, 229–238. [CrossRef]
2. Bing, L.; Phuong, D.; Liu, J.; He, Y.; Shang, J. Recent Advances of Hyperspectral Imaging Technology and Applications in Agriculture. *Remote Sens.* **2020**, *12*, 2659. [CrossRef]
3. Li, H.; Cui, J.; Zhang, X.; Han, Y.; Cao, L. Dimensionality Reduction and Classification of Hyperspectral Remote Sensing Image Feature Extraction. *Remote Sens.* **2022**, *14*, 4579. [CrossRef]
4. Guo, P.; Li, T.; Gao, H.; Chen, X.; Cui, Y.; Huang, Y. Evaluating Calibration and Spectral Variable Selection Methods for Predicting Three Soil Nutrients Using Vis-Nir Spectroscopy. *Remote Sens.* **2021**, *13*, 4000. [CrossRef]
5. Liu, J.; Dong, Z.; Xia, J.; Wang, H.; Meng, T.; Zhang, R.; Han, J.; Wang, N.; Xie, J. Estimation of Soil Organic Matter Content Based on Cars Algorithm Coupled with Random Forest. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2021**, *258*, 119823. [CrossRef]

6. Tinghui, W.; Yu, J.; Lu, J.; Zou, X.; Zhang, W. Research on Inversion Model of Cultivated Soil Moisture Content Based on Hyperspectral Imaging Analysis. *Agriculture* **2020**, *10*, 292. [CrossRef]
7. Jia, L.S.; Le, Z. Quantitative Analysis of Soil Total Nitrogen Using Hyperspectral Imaging Technology with Extreme Learning Machine. *Sensors* **2019**, *19*, 4355. [CrossRef] [PubMed]
8. Wang, Q.; Li, P.; Pu, Z.; Chen, X. Calibration and Validation of Salt-Resistant Hyperspectral Indices for Estimating Soil Moisture in Arid Land. *J. Hydrol.* **2011**, *408*, 276–285. [CrossRef]
9. Yin, Y.B.; Li, X.; Zhao, Z.; Dong, D.R. Predict Model and Analysis of the Sandy Soil Moisture with Hyperspectral. *Remote Sens. Technol. Appl.* **2011**, *26*, 355–359.
10. Lim, H.H.; Cheon, E.; Lee, D.-H.; Jeon, J.-S.; Lee, S.-R. Soil Water Content Measurement Technology Using Hyperspectral Visible and near-Infrared Imaging Technique. *J. Korean Geotech. Soc.* **2019**, *35*, 51–62.
11. Yu, L.; Zhu, Y.; Hong, Y.; Xia, T.; Liu, M.; Zhou, Y. Determination of Soil Moisture Content by Hyperspectral Technology with Cars Algorithm. *Trans. Chin. Soc. Agric. Eng.* **2016**, *32*, 138–145.
12. Wang, W.; Gao, M.; Wang, J. Hyperspectral Parameters and Prediction Model of Soil Moisture in Apple Orchards. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *687*, 012085. [CrossRef]
13. Xu, C.; Zeng, W.; Huang, J.; Wu, J.; Van Leeuwen, W.J. Prediction of Soil Moisture Content and Soil Salt Concentration from Hyperspectral Laboratory and Field Data. *Remote Sens.* **2016**, *8*, 42. [CrossRef]
14. Lv, J.; Zhang, R.; Tu, J.; Liao, M.; Pang, J.; Yu, B.; Li, K.; Xiang, W.; Fu, Y.; Liu, G. A Gnss-Ir Method for Retrieving Soil Moisture Content from Integrated Multi-Satellite Data That Accounts for the Impact of Vegetation Moisture Content. *Remote Sens.* **2021**, *13*, 2442. [CrossRef]
15. Cai, L.; Ding, J. Wavelet Transformation Coupled with Cars Algorithm Improving Prediction Accuracy of Soil Moisture Content Based on Hyperspectral Reflectance. *Trans. Chin. Soc. Agric. Eng.* **2017**, *33*, 144–151.
16. Ke, X.; Xia, S.; Shen, Q.; Yang, B.; Song, Q.; Xu, Y.; Zhang, S.; Zhou, X.; Zhou, Y. Moisture Spectral Characteristics and Hyperspectral Inversion of Fly Ash-Filled Reconstructed Soil. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2021**, *253*, 119590.
17. Shen, L.; Gao, M.; Yan, J.; Li, Z.-L.; Leng, P.; Yang, Q.; Duan, S.-B. Hyperspectral Estimation of Soil Organic Matter Content Using Different Spectral Preprocessing Techniques and Plsr Method. *Remote Sens.* **2020**, *12*, 1206. [CrossRef]
18. Ge, J.D.; Wang, J.; Fei, W.; Sun, H. Estimation of Soil Moisture Content Based on Competitive Adaptive Reweighted Sampling Algorithm Coupled with Machine Learning. *Acta Opt. Sin.* **2018**, *38*, 1030001.
19. Lobato, M.; Norris, W.R.; Nagi, R.; Soylemezoglu, A.; Nottage, D. Machine Learning for Soil Moisture Prediction Using Hyperspectral and Multispectral Data. In Proceedings of the 2021 IEEE 24th International Conference on Information Fusion (FUSION), Sun City, South Africa, 1–4 November 2021.
20. Jia, P.; Zhang, J.; He, W.; Yuan, D.; Hu, Y.; Zamanian, K.; Jia, K.; Zhao, X. Inversion of Different Cultivated Soil Types' Salinity Using Hyperspectral Data and Machine Learning. *Remote Sens.* **2022**, *14*, 5639. [CrossRef]
21. Hong, D.; Gao, L.; Yokoya, N.; Yao, J.; Chanussot, J.; Du, Q.; Zhang, B. More Diverse Means Better: Multimodal Deep Learning Meets Remote-Sensing Imagery Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 4340–4354. [CrossRef]
22. Kussul, N.; Lavreniuk, M.; Skakun, S.; Shelestov, A. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 778–782. [CrossRef]
23. Xu, F.; Hu, C.; Li, J.; Plaza, A.; Datcu, M. Special Focus on Deep Learning in Remote Sensing Image Processing. *Sci. China-Inf. Sci.* **2020**, *63*, 140300. [CrossRef]
24. Yang, X.; Ye, Y.; Li, X.; Lau, R.Y.K.; Zhang, X.; Huang, X. Hyperspectral Image Classification with Deep Learning Models. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5408–5423. [CrossRef]
25. Wang, H.; Zhang, L.; Zhao, J.; Hu, X.; Ma, X. Application of Hyperspectral Technology Combined with Genetic Algorithm to Optimize Convolution Long- and Short-Memory Hybrid Neural Network Model in Soil Moisture and Organic Matter. *Appl. Sci.* **2022**, *12*, 10333. [CrossRef]
26. Ahmed, A.M.; Deo, R.C.; Raj, N.; Ghahramani, A.; Feng, Q.; Yin, Z.; Yang, L. Deep Learning Forecasts of Soil Moisture: Convolutional Neural Network and Gated Recurrent Unit Models Coupled with Satellite-Derived Modis, Observations and Synoptic-Scale Climate Index Data. *Remote Sens.* **2021**, *13*, 554. [CrossRef]
27. Tang, J.; Liu, G.; Pan, Q. A Review on Representative Swarm Intelligence Algorithms for Solving Optimization Problems: Applications and Trends. *IEEE/CAA J. Autom. Sin.* **2021**, *8*, 1627–1643. [CrossRef]
28. Tao, L.; Wang, G.; Chen, X.; Li, J.; Cai, Q. Soil Moisture Retrieval Using Modified Particle Swarm Optimization and Back-Propagation Neural Network. *Photogramm. Eng. Remote Sens.* **2019**, *85*, 789–798. [CrossRef]
29. Liang, Y.-J.; Ren, C.; Wang, H.-Y.; Huang, Y.-B.; Zheng, Z.-T. Research on Soil Moisture Inversion Method Based on Ga-Bp Neural Network Model. *Int. J. Remote Sens.* **2019**, *40*, 2087–2103. [CrossRef]
30. Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey Wolf Optimizer. *Adv. Eng. Softw.* **2014**, *69*, 46–61. [CrossRef]
31. Zhang, N.; Zhang, X.; Shang, P.; Ma, R.; Yuan, X.; Li, L.; Bai, T. Detection of Cotton Verticillium Wilt Disease Severity Based on Hyperspectrum and Gwo-Svm. *Remote Sens.* **2023**, *15*, 3373. [CrossRef]
32. Ladi, S.K.; Panda, G.K.; Dash, R.; Ladi, P.K.; Dhupar, R. A Novel Grey Wolf Optimisation Based Cnn Classifier for Hyperspectral Image Classification. *Multimed. Tools Appl.* **2022**, *81*, 28207–28230. [CrossRef]
33. Ge, X.; Ding, J.; Jin, X.; Wang, J.; Chen, X.; Li, X.; Liu, J.; Xie, B. Estimating Agricultural Soil Moisture Content through Uav-Based Hyperspectral Images in the Arid Region. *Remote Sens.* **2021**, *13*, 1562. [CrossRef]

34. Lin, N.; Jiang, R.; Li, G.; Yang, Q.; Li, D.; Yang, X. Estimating the Heavy Metal Contents in Farmland Soil from Hyperspectral Images Based on Stacked Adaboost Ensemble Learning. *Ecol. Indic.* **2022**, *143*, 109330. [CrossRef]
35. Xie, B.; Ding, J.; Ge, X.; Li, X.; Han, L.; Wang, Z. Estimation of Soil Organic Carbon Content in the Ebinur Lake Wetland, Xinjiang, China, Based on Multisource Remote Sensing Data and Ensemble Learning Algorithms. *Sensors* **2022**, *22*, 2685. [CrossRef]
36. Cheng, Y.; Li, Y.; Li, F.; He, L. Soil Moisture Retrieval Using Extremely Randomized Trees over the Shandian River Basin. *Natl. Remote Sens. Bull.* **2021**, *25*, 941–951. [CrossRef]
37. Han, Y.; Li, S.; Zheng, W.; Shi, S.; Zhu, X.; Jin, X. Regression Prediction of Soil Available Nitrogen near-Infrared Spectroscopy Based on Boosting Algorithm. *Laser Optoelectron. Prog.* **2021**, *58*, 555–565.
38. Jin, X.; Li, S.; Zhang, W.; Zhu, J.; Sun, J. Prediction of Soil-Available Potassium Content with Visible near-Infrared Ray Spectroscopy of Different Pretreatment Transformations by the Boosting Algorithms. *Appl. Sci.* **2020**, *10*, 1520. [CrossRef]
39. Huang, X.; Guan, H.; Bo, L.; Xu, Z.; Mao, X. Hyperspectral Proximal Sensing of Leaf Chlorophyll Content of Spring Maize Based on a Hybrid of Physically Based Modelling and Ensemble Stacking. *Comput. Electron. Agric.* **2023**, *208*, 107745. [CrossRef]
40. Fu, B.; Deng, L.; Zhang, L.; Qin, J.; Liu, M.; Jia, M.; He, H.; Deng, T.; Gao, E.; Fan, D. Estimation of Mangrove Canopy Chlorophyll Content Using Hyperspectral Image and Stacking Ensemble Regression Algorithm. *Natl. Remote Sens. Bull.* **2022**, *26*, 1182–1205. [CrossRef]
41. Feng, L.; Zhang, Z.; Ma, Y.; Du, Q.; Williams, P.; Drewry, J.; Luck, B. Alfalfa Yield Prediction Using Uav-Based Hyperspectral Imagery and Ensemble Learning. *Remote Sens.* **2020**, *12*, 2028. [CrossRef]
42. Wang, Z.; Hang, Z.; He, C.; Cai, T.; Feng, Y.; Lu, N.; Dou, H. Hyperspectral inversion study of Vertisol soil moisture content based on ensemble learning. *J. Agric. Resour. Environ.* **2023**, *40*, 1426–1434.
43. Tao, S.; Zhang, X.; Chen, J.; Zhang, Z.; Kang, X.; Qi, W.; Wang, Y.; Gao, Y. Generating surface soil moisture at the 30 m resolution in grape-growing areas based on stacked ensemble learning. *Int. J. Remote Sens.* **2024**, *45*, 5385–5424. [CrossRef]
44. Zhang, F.; Wu, S.; Liu, J.; Wang, C.; Guo, Z.; Xu, A.; Pan, K.; Pan, X. Predicting soil moisture content over partially vegetation covered surfaces from hyperspectral data with deep learning. *Soil Sci. Soc. Am. J.* **2021**, *85*, 989–1001. [CrossRef]
45. Yang, Y.; Zhang, H.; Gao, L.L.; Wang, J.Y.; Yang, J.H. Study on Soil Moisture Utilization in Different Rotation Patterns of Maize and Soybean. *Anhui Agric. Sci. Bull.* **2019**, *25*, 38–40.
46. Yann, L.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-Based Learning Applied to Document Recognition. *Proc. IEEE* **1998**, *86*, 2278–2324.
47. Malley, D.F.; Williams, P.C. Use of near-Infrared Reflectance Spectroscopy in Prediction of Heavy Metals in Freshwater Sediment by Their Association with Organic Matter. *Environ. Sci. Technol.* **1997**, *31*, 3461–3467. [CrossRef]
48. Yuan, J.; Wang, X.; Yan, C.-X.; Wang, S.-R.; Ju, X.-P.; Li, Y. Soil Moisture Retrieval Model for Remote Sensing Using Reflected Hyperspectral Information. *Remote Sens.* **2019**, *11*, 366. [CrossRef]
49. Zhang, Y.; Tan, K.; Wang, X.; Chen, Y. Retrieval of Soil Moisture Content Based on a Modified Hapke Photometric Model: A Novel Method Applied to Laboratory Hyperspectral and Sentinel-2 Msi Data. *Remote Sens.* **2020**, *12*, 2239. [CrossRef]
50. Mireguli, A.; Ding, J.; Kasim, N.; Wang, J.; Wang, J. Regional Scale Soil Moisture Content Estimation Based on Multi-Source Remote Sensing Parameters. *Int. J. Remote Sens.* **2020**, *41*, 3346–3367.
51. Zhang, F.; Wang, C.; Pan, K.; Guo, Z.; Liu, J.; Xu, A.; Ma, H.; Pan, X. The Simultaneous Prediction of Soil Properties and Vegetation Coverage from Vis-Nir Hyperspectral Data with a One-Dimensional Convolutional Neural Network: A Laboratory Simulation Study. *Remote Sens.* **2022**, *14*, 397. [CrossRef]
52. Kara, A.; Pekel, E.; Ozcetin, E.; Yıldız, G.B. Genetic Algorithm Optimized a Deep Learning Method with Attention Mechanism for Soil Moisture Prediction. *Neural Comput. Appl.* **2024**, *36*, 1761–1772. [CrossRef]
53. Wang, S.; Wu, Y.; Li, R.; Wang, X. Remote Sensing-Based Retrieval of Soil Moisture Content Using Stacking Ensemble Learning Models. *Land Degrad. Dev.* **2023**, *34*, 911–925. [CrossRef]
54. Ge, X.; Ding, J.; Wang, J.; Sun, H.; Zhu, Z. A New Method for Predicting Soil Moisture Based on UAV Hyperspectral Image. *Spectrosc. Spectr. Anal.* **2020**, *40*, 602–609.
55. Wang, R.; Zhao, J.; Yang, H.; Li, N. Inversion of Soil Moisture on Farmland Areas Based on SSA-CNN Using Multi-Source Remote Sensing Data. *Remote Sens.* **2023**, *15*, 2515. [CrossRef]
56. Zhang, T.; Fu, Q.; Tian, R.; Zhang, Y.; Sun, Z. A spectrum contextual self-attention deep learning network for hyperspectral inversion of soil metals. *Ecol. Indic.* **2023**, *152*, 110351. [CrossRef]