



# Article Estimation of Coastal Wetland Soil Organic Carbon Content in Western Bohai Bay Using Remote Sensing, Climate, and Topographic Data

Yongbin Zhang <sup>1,†</sup>, Caiyao Kou <sup>1,†</sup>, Mingyue Liu <sup>1,2,3,4,†</sup>, Weidong Man <sup>1,2,3,4,\*</sup>, Fuping Li <sup>1,2,3,4</sup>, Chunyan Lu <sup>5</sup>, Jingru Song <sup>1</sup>, Tanglei Song <sup>1</sup>, Qingwen Zhang <sup>1</sup>, Xiang Li <sup>1</sup> and Di Tian <sup>1</sup>

- <sup>1</sup> College of Mining Engineering, North China University of Science and Technology, Tangshan 063210, China; zhangyb@ncst.edu.cn (Y.Z.); koucaiyao@stu.ncst.edu.cn (C.K.); liumy917@ncst.edu.cn (M.L.); lifuping@ncst.edu.cn (F.L.); songjingru@stu.ncst.edu.cn (J.S.); songtanglei@stu.ncst.edu.cn (T.S.); zhangqingwen@stu.ncst.edu.cn (Q.Z.); lixiang@stu.ncst.edu.cn (X.L.); td@stu.ncst.edu.cn (D.T.)
- <sup>2</sup> Tangshan Key Laboratory of Resources and Environmental Remote Sensing, Tangshan 063210, China
- <sup>3</sup> Hebei Industrial Technology Institute of Mine Ecological Remediation, Tangshan 063210, China
- <sup>4</sup> Collaborative Innovation Center of Green Development and Ecological Restoration of Mineral Resources, Tangshan 063210, China
- <sup>5</sup> College of Computer and Information Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, China; luchunyan@fafu.edu.cn
- Correspondence: manwd@ncst.edu.cn; Tel.: +86-315-8805408
- These authors contributed equally to this work.

Abstract: Coastal wetland soil organic carbon (CW-SOC) is crucial for wetland ecosystem conservation and carbon cycling. The accurate prediction of CW-SOC content is significant for soil carbon sequestration. This study, which employed three machine learning (ML) methods, including random forest (RF), gradient boosting machine (GBM), and extreme gradient boosting (XGBoost), aimed to estimate CW-SOC content using 98 soil samples, SAR images, optical images, and climate and topographic data. Three statistical metrics and leave-one-out cross-validation were used to evaluate model performance. Optimal models using different ML methods were applied to predict the spatial distribution of CW-SOC content. The results showed the following: (1) The models built using optical images had higher predictive accuracy than models built using synthetic aperture radar (SAR) images. The model that combined SAR images, optical images, and climate data demonstrated the highest prediction accuracy. Compared to the model using only optical images and SAR images, the prediction accuracy was improved by 0.063 and 0.115, respectively. (2) Regardless of the combination of predictive variables, the XGBoost method achieved higher prediction accuracy than the RF and GBM methods. (3) Optical images were the main explanatory variables for predicting CW-SOC content, explaining more than 65% of the variability. (4) The CW-SOC content predicted by the three ML methods showed similar spatial distribution characteristics. The central part of the study area had higher CW-SOC content, while the southern and northern regions had lower levels. This study accurately predicted the spatial distribution of CW-SOC content, providing data support for ecological environmental protection and carbon neutrality of coastal wetlands.

**Keywords:** random forest; gradient boosting machine; extreme gradient boosting; Sentinel-1A; Sentinel-2A; CW-SOC content inversion

#### 1. Introduction

The soil organic carbon (SOC) pool contains twice as much organic carbon as the atmosphere and three times as much as vegetation [1]. It plays a critical role in mitigating global warming [2,3]. Wetlands are referred to as "the kidneys of earth"; despite covering only 6–8% of the global land surface [4], wetlands store 20–30% of terrestrial SOC and play a major role in the global carbon cycle [5,6]. Coastal wetlands are an essential type of



Citation: Zhang, Y.; Kou, C.; Liu, M.; Man, W.; Li, F.; Lu, C.; Song, J.; Song, T.; Zhang, Q.; Li, X.; et al. Estimation of Coastal Wetland Soil Organic Carbon Content in Western Bohai Bay Using Remote Sensing, Climate, and Topographic Data. *Remote Sens.* **2023**, *15*, 4241. https://doi.org/10.3390/ rs15174241

Academic Editor: Deepak R. Mishra

Received: 18 August 2023 Accepted: 26 August 2023 Published: 29 August 2023



**Copyright:** © 2023 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/). wetlands. They have a strong carbon sink capacity due to the periodic tidal inundation of seawater they experience [7]. However, because coastal wetlands are located at the junction of land and sea and are affected by various factors from both the land and sea, the changes in the soil environment are relatively complex, increasing the difficulty of predicting coastal wetland soil organic carbon (CW-SOC) content. Therefore, an accurate estimation of CW-SOC content is crucial for assessing coastal wetlands' carbon sequestration capacity and promoting the implementation of China's carbon-neutral strategy [8].

Traditional methods for predicting the spatial distribution of SOC require a great deal of field sampling data [9]. These methods are relatively time-consuming, costly, and highly destructive to wetlands. It is difficult to predict the SOC content in large areas. Remote sensing technology has the advantages of short update cycles, large monitoring ranges, and fewer restrictions by environmental factors, playing an important role in predicting the spatial distribution of SOC content [10,11]. Optical images, as the main source of remote sensing data, are widely used in the inversion of soil characteristics by providing information such as spectral reflectance and remote sensing indices [12–14]. However, it is often cloudy and rainy in coastal areas; optical images are easily affected by cloud cover and cannot obtain effective information in these conditions, which can lead to a decrease in the accuracy of predicting CW-SOC content. Synthetic aperture radar (SAR) images are not constrained by weather and lighting conditions; SAR possesses strong cloud-penetration capabilities, which can effectively overcome the issue of optical images being unable to obtain information under cloud coverage [15]. Some studies conducted by Yang et al. indicated that SAR images contribute to the prediction of soil properties [16,17]. However, SAR images contain a lot of noise, require complex processing, and are not conducive to wide application [18]. Could the combination of both optical and SAR imagery be employed for soil property prediction? Zhou and Azizi conducted research in the Heihe River Basin in China and in western Iran, discovering that the combination of the two yielded promising results in predicting soil properties [19,20]. Currently, research on this method is primarily focused on inland wetlands, with limited studies concerning coastal wetlands. Taking into account the complex environment of coastal wetlands, we incorporate terrain and climatic variables in addition to combining optical and SAR images to predict CW-SOC content.

The use of appropriate models to accurately predict the spatial distribution of CW-SOC content based on SAR images, optical images, climate data, and topographic data has been a subject of concern among scholars. Currently, methods for predicting SOC mainly include linear and nonlinear models. Linear models can intuitively describe the relationship between independent and dependent variables. However, remote sensing data and SOC content mostly have a nonlinear relationship. Nonlinear models better identify the nonlinear relationship between predictor variables and CW-SOC content, solve the problem of spatial autocorrelation [14,21], and have good generalization performance. This is especially the case for machine learning methods based on decision tree models, such as random forest (RF), gradient boosting machine (GBM), and extreme gradient boosting (XGBoost) [22]. They show good performance in predicting SOC content, especially when dealing with complex and high-dimensional data. Related studies have found that ML methods have higher reliability and better robustness in predicting soil properties [23].

The coastal wetlands of western Bohai Bay are affected by factors such as vegetation, climate, and hydro-geographical conditions, making the environment complex and changeable. Therefore, accurately predicting CW-SOC content is of paramount importance. To ensure the effective acquisition of CW-SOC content, we constructed four CW-SOC content prediction models using RF, GBM, and XGBoost methods based on SAR images, optical images, and climate and topographic data. Therefore, the purpose of this paper is to (1) compare the prediction performances of four CW-SOC content models constructed using three different ML methods; (2) explore the relative importance of the predictor variables in the RF, GBM, and XGBoost methods; and (3) utilize the optimal model from each method to predict the spatial distribution of CW-SOC content. This provides data support for wetlands conservation and carbon cycling in coastal areas.

#### 2. Materials and Methods

## 2.1. Study Area

The coastal area of western Bohai Bay includes Tianjin Binhai New District, Huanghua, and Haixing (Figure 1). The geographical coordinates of the study area are 117°27′29″–118°03′33″E and 38°22′22″–39°19′33″N, and the total area is 5475.5 km<sup>2</sup>. It is an integral component of the Bohai Sea Economic Circle and has abundant wetland resources, including Nandagang wetlands reserve and Beidagang wetlands reserve. The region has a warm temperate semi-humid and semiarid monsoonal continental climate with marine climatic features. The mean annual precipitation (MAP) ranges from 567.8 to 782.6 mm, with the majority of precipitation concisely concentrated in the summer, accounting for more than 60% of the annual precipitation. Referring to the Bohai Rim coastal wetland map and the comprehensive classification system of Chinese coastal wetlands proposed by Liu [24,25], we defined the coastal wetlands in this study as the wetlands in the coastal area of the western Bohai Bay, including natural coastal wetlands (tidal flats, marshes, and water bodies) and artificial coastal wetlands (salt fields and mariculture ponds).



**Figure 1.** Location of the study area and sample points distribution: (**a**) Binhai New District; (**b**) Huanghua; (**c**) Haixing.

## 2.2. Soil Sampling and Analysis

In October 2021, the low tide period and mariculture ponds drainage period were selected for soil sample collection (0–30 cm) in the coastal area of western Bohai Bay (Figure 1). The sampling points covered various land use types. Three soil samples were collected from each sampling point and thoroughly mixed to form a composite sample [26]. Considering accessibility and flood discharge, the straight-line interval between design sampling points is about 3 km, with a total of 98 soil samples collected. The collected

soil samples were air-dried in the laboratory, ground, and purified to remove impurities. Subsequently, the samples were filtered using a 100-mesh sieve with an aperture of 0.15 mm, and their SOC content was measured using the potassium dichromate capacity.

The variation range of CW-SOC content was 2.198–18.835 g·kg<sup>-1</sup>, and the mean and SD values of CW-SOC content were 6.116 g·kg<sup>-1</sup> and 3.614 g·kg<sup>-1</sup>, respectively. The CV value is 59.091%, which belongs to the medium degree of variation (Table 1) [27].

Table 1. Description statistics of CW-SOC content.

|        | Max/(g·kg $^{-1}$ ) | $Min/(g \cdot kg^{-1})$ | Mean/(g·kg⁻                           | (g·kg <sup>-1</sup> ) SD/(g·kg <sup>-1</sup> ) | CV/(%) |
|--------|---------------------|-------------------------|---------------------------------------|--|--------|
| CW-SOC | 18.835              | 2.198                   | 6.116                                 | 3.614  | 59.091 |
|        | 14                  | CD + 1 + 1              | · · · · · · · · · · · · · · · · · · · | (C <sup>1</sup> · · · · · · ·                  |        |

Notes: Max, maximum; Min, minimum; SD, standard deviation; CV, coefficient of variation.

#### 2.3. Predictor Variables

2.3.1. Remote Sensing Variables and Processing

The remote sensing data includes SAR images (Sentinel-1A) and optical images (Sentinel-2A) obtained from Google Earth Engine (GEE). In order to extract remote sensing data values corresponding to soil samples, the time filtering function of GEE, sentinel-1A, and Sentinel-2A images close to the sampling period were selected (October 2021). Sentinel-1A is a satellite equipped with C-band SAR and is widely used in soil property research [28]. We chose interferometric wide swath mode ground range detected data with VV and VH dual polarization modes. The images have a spatial resolution of 10 m  $\times$  10 m [28]. We selected the VV polarization backscattering coefficient (VV) and VH polarization backscattering coefficient (VV) and VH polarization modes: difference of VV and VH (D), sum of VV and VH (S), quotient of VV and VH (Q) and difference sum ratio of VV and VH (DSR) were used as remote sensing variables for predicting CW-SOC content (Table 2).

Sentinel-2A contains 13 optical harmonic bands and is the only optical satellite with three bands in the red-edge range [29]. The images used in this study were obtained from the zenith reflectance dataset on the GEE and underwent radiometric calibration and topographic correction. All images underwent a de-cloud process through quality assessment to obtain images with less than 5% cloud cover and then were cropped to fit the study area. Reflectance information of the Sentinel-2A image bands was extracted, and remote sensing indices were constructed through mathematical calculation. The construction of remote sensing indices helps to overcome the influence of illumination and atmosphere, and is conducive to the prediction of CW-SOC content. Additionally, the red-edge band of Sentinel-2A and the corresponding remote sensing indices are highly sensitive to vegetation growth and can provide valuable information on soil properties [30]. Therefore, they are frequently employed as evaluation indices to assist in the prediction of soil characteristics [31]. Based on previous research [32–43], this study selected 10 spectral bands and 16 remote sensing indices from Sentinel-2A images as predictor variables (Table 2).

 Table 2. Remote sensing variables for modeling.

| Sources | Category       | Variables | Calculation Formula | Literature |
|---------|----------------|-----------|---------------------|------------|
|         |                | VV, VH    | -                   | [32]       |
| CAD     | Polarization   | D         | VV - VH             | [32]       |
| images  | backscattering | S         | VV + VH             | [32]       |
|         | coefficient    | Q         | VV/VH               | [32]       |
|         |                | DSR       | (VV - VH)/(VV + VH) |            |

| Sources | Category            | Variables   | Calculation Formula  | Literature   |
|---------|---------------------|---|--|--------------|
|         | Band<br>reflectance | B2 (490 nm), B3 (560 nm)<br>B4 (665 nm), B5 (705 nm)<br>B6 (740 nm), B7 (783 nm)<br>B8 (842 nm), B8A (865 nm)<br>B11 (1610 nm), B12 (2190 nm) | _  | [32]         |
|         |                     | NDVI<br>NDWI  | (B8 - B4)/(B8 + B4)<br>(B3 - B8)/(B3 + B8)                     | [33]<br>[34] |
| -       |                     | NDBI  | (B11 - B8)/(B11 + B8)  | [35]         |
|         | Remote              | SAVI  | $1.5 \times (B8 - B4)/(B8 + B4 + 0.5)$                         | [36]         |
| optical |                     | RVI   | B8/B4  | [37]         |
| imagas  |                     | DVI   | B8 - B4  | [38]         |
| images  |                     | EVI   | $2.5 \times (B8 - B4)/(B8 + 6 \times B4) - 7.5 \times B2 + 1)$ | [39]         |
|         |                     | BSI   | 1 + ((B4 + B11) - (B8 + B2))/((B4 + B11) + (B8 + B2))          | [40]         |
|         | sensing             | NDRE1   | (B6 - B5)/(B6 + B5)  | [41]         |
|         | indices             | NDRE2   | (B7 - B5)/(B7 + B5)  | [41]         |
|         |                     | CIRE1   | (B8/B5) - 1  | [42]         |
|         |                     | CIRE2   | (B8/B6) - 1  | [42]         |
|         |                     | CIRE3   | (B8/B7) - 1  | [42]         |
|         |                     | NDVIRE1   | (B8 - B5)/(B8 + B5)  | [43]         |
|         |                     | NDVIRE2   | (B8 - B6)/(B8 + B6)  | [43]         |
|         |                     | NDVIRE3   | (B8 - B7)/(B8 + B7)  | [43]         |

Table 2. Cont.

Notes: B2, B3, B4, B5, B6, B7, B8, B8A, B11, and B12 represent the bands reflectance values of Sentinel-2A images.

#### 2.3.2. Environmental Variables

The MAP, mean annual temperature (MAT), and mean annual relative humidity (MARH), which are climate variables, were obtained by collating and calculating observational data from 26 meteorological stations in Tianjin and Cangzhou. These variables are interpolated using the inverse distance weighting methods and have a spatial resolution of 10 m. This is a useful and commonly used method for obtaining continuous soil and climate variables in space [44].

For this study, elevation (H), slope (i), and aspect ( $\alpha$ ) were chosen as the topographic variables in order to ensure that the DEM data takes into account both full coverage and high resolution. Therefore, the shuttle radar topographic mapping version 3.0 (SRTM V3) data product with a resolution of 30 m was obtained on the GEE. To reduce the impact caused by different resolutions, the SRTM V3 data product was resampled to 10 m resolution. Slope and aspect were calculated from the elevation data using the topography analysis command available on the GEE.

## 2.4. Boruta

Some predictor variables may exhibit redundancy and high autocorrelation, failing to provide effective information for predicting CW-SOC content [26]. Therefore, it is necessary to perform variable selection to enhance model accuracy [45]. Boruta, as a commonly used variable selection method, has been widely employed [46–48]. Its main principle involves randomly shuffling the original variables to create shadow variables, inputting the shadow and original variables into the random forest classifier to calculate Z-scores. The maximum Z-score among the shadow variables is denoted as  $Z_{max}$ , and variables with Z-scores higher than  $Z_{max}$  are retained, while those with Z-scores lower than  $Z_{max}$  are removed, this process is repeated until all variables are screened [48–51]. In this study, variable selection was implemented using the 'Boruta' package in the R environment.

We applied the Boruta method for variable selection on Model A, Model B, Model C, and Model D, using the selected variables to predict CW-SOC content. The results of variable selection are presented in Table 3.

Table 3. Variables screening results.

| No          | Model       | Variables                          | Screening Variables                                    |
|-------------|-------------|------------------------------------|--|
| Ι           | Model A     | SAR images                         | VV, VH, D, S, Q, and DSR                               |
| П           | Model B     | Optical images                     | B2, B3, B4, CIRE1, NDVI, NDEI, RVI, DVI, NDVIRE1,      |
| II WIOdel D |             | optical mages                      | NDRE1, NDRE2, EVI, SAVI                                |
| Ш           | Model C     | SAR and optical images             | VH, D, B2, B3, B4, CIRE1, NDVI, NDEI, RVI, DVI,        |
|             |             |                                    | NDVIRE1, NDRE1, NDRE2, EVI, SAVI                       |
| IV          | Model D     | SAR images, optical images,        | VH, D, B2, B3, B4, CIRE1, NDVI, NDEI, RVI, DVI, NDRE1, |
| 1 V         | initiater D | topographic, and climate variables | NDRE2, EVI, SAVI, MARH                                 |

## 2.5. Modeling Methods

## 2.5.1. Random Forest

The random forest (RF) is an ML method based on decision trees, which is suitable for classification and regression tasks [52]. The algorithm constructs a large number of uncorrelated decision trees, each tree corresponds to independent original sample data and models each sample data using decision tree methods [9]. Then, these decision trees are combined into a prediction model of multiple decision trees, providing a single prediction target for data prediction [53]. Finally, the prediction result is the average value of all single-tree predictions [54]. Because of its advantages in processing multi-variate nonlinear data, this method has become the preferred choice for soil property research [55].

#### 2.5.2. Gradient Boosting Machine

The gradient boosting machine (GBM) is a typical algorithm for boosting. The core of the GBM is to construct a more powerful model by iteratively combining multiple weak models. The loss function of the previous model is reduced through each iteration so that the overall loss function of the model decreases and the model improves continuously [56]. The loss function describes the degree of the unreliability of the model. If the model's loss function continues declining, it indicates that it is continuously optimized. When the loss of the model continues to decline in its gradient direction, the speed of model optimization is increased [57].

#### 2.5.3. Extreme Gradient Boosting

The extreme gradient boosting (XGBoost) is a decision-tree-based model, which achieves high accuracy in practical applications. The fundamental idea behind the XGBoost method is to use an additive learning approach and iteratively combine multiple models with lower accuracy into a higher-accuracy model [58,59]. By minimizing uncertainty, overfitting is controlled, and the model's generalization ability is improved. Compared to the traditional GBM, the extreme gradient boosting model has improved and optimized running speed and accuracy [60].

The above three ML methods are implemented using the "train" function of the "caret" package in the R environment. The main parameters in the ML method were optimized using grid search to reduce model errors and improve model performance.

#### 2.6. Model Performance Evaluation

Four predictor models were constructed using three ML methods, RF, GBM, and XGBoost, based on 98 soil samples and multi-source data (Figure 2). Model A and Model B consisted of SAR images and optical images predictor variables, respectively; Model C consisted of both SAR images and optical images predictor variables; and Model D consisted of all predictor variables (Figure 2). The model performance was validated by using a leave-one-out cross-validation method [9]. The method divides the original data

into *n* subsets, where n is 98 in this study. It selects n- subsets for training and 1 subset for testing, repeating this process n times [9]. This approach effectively avoids overfitting and underfitting, ensuring the accuracy of predictions. The coefficient of determination ( $R^2$ ), mean absolute error (*MAE*), and root mean square error (*RMSE*) were used to evaluate the model prediction accuracy. The closer the  $R^2$  value is to 1 and the lower the *MAE* and *RMSE* values, the better the model performance [61]. These evaluation metrics are calculated using the following formula:

$$R^{2} = \frac{\sum_{i=1}^{n} (\hat{y}_{i} - \overline{y})^{2}}{\sum_{i=1}^{n} (y_{i} - \overline{y})^{2}}, \overline{y} = \frac{\sum_{i=1}^{n} y_{i}}{n}$$
(1)

$$MAE = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n}$$
(2)

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

where  $y_i$  denotes the measured values of CW-SOC,  $\hat{y}_i$  denotes the predicted values of CW-SOC,  $\overline{y}$  denotes the measured mean value of SOC, *n* denotes the number of samples, and *i* = 1, 2, 3..., *n*.



Figure 2. Workflow diagram of this study.

## 3. Results

## 3.1. Model Performance Comparison

The model performance of RF, GBM, and XGBoost methods in predicting the CW-SOC content was evaluated through *R*<sup>2</sup>, *MAE*, and *RMSE* values. When using the RF, GBM, and XGBoost methods to establish a CW-SOC model based on multi-source data, it was found that the combination of multi-source data can effectively improve the prediction accuracy of the model and minimize its variability (Figure 3). Model D, based on the combination of multi-source data, can achieve the best prediction performance using the RF, GBM, and XGBoost methods. Compared to Model A, Model D improves the model performance by 22.9% (RF), 34.9% (GBM), and 18.7% (XGBoost). Model D's performance has been improved by 10.7% (RF), 11.4% (GBM), and 7.8% (XGBoost) compared to Model B. Compared to model C, model D shows a smaller degree of performance improvement, with improvements of 7% (RF), 6% (GBM), and 2.2% (XGBoost).



Figure 3. RF, GBM, and XGBoost model results: (A1–A4), RF; (B1–B4), GBM; (C1–C4), XGBoost.

The  $R^2$  value of the CW-SOC content prediction model constructed by the RF method exceeds 0.4, with the highest being 0.505. The  $R^2$  value of Model C based on the RF method is 14.8% and 3.5% higher than Model A and Model B, respectively. Furthermore, the *MAE/RMSE* value of Model C is 0.125/0.216 g·kg<sup>-1</sup> and 0.048/0.078 g·kg<sup>-1</sup> lower than Model A and Model B, respectively. After adding environmental variables, the  $R^2$  value of Model D reached 0.505. While the *MAE* and *RMSE* values decrease to 1.092 g·kg<sup>-1</sup> and 1.479 g·kg<sup>-1</sup>, respectively.

Scatterplots of four models were constructed by analyzing the RF and GBM method (Figure 3(A1–B4)). It is evident that when SAR images are utilized as the input predictor variable, the GBM method's predictive ability is weaker than that of the RF method. For other models, the GBM method performs better than the RF method. Model D, based on the GBM method, exhibits 34.9% and 11.4% higher prediction accuracy than Model A and Model B, respectively, and the optimization is greater than that of the RF method. Compared to Model C, the  $R^2$  value of Model D, constructed by using the GBM method, only increased by 0.029, which is lower than the 0.033 observed with the RF method.

Figure 3 displays the modeling results of the XGBoost method in the four models. When compared to the RF and GBM methods, although the XGBoost method does not exhibit as impressive optimization as the GBM method, it achieves the optimal model performance among the three ML methods. The  $R^2$  values of the four models constructed based on the XGBoost method are all greater than 0.6, the highest value is 0.730 and the *MAE* values are all less than 1 g·kg<sup>-1</sup>, indicating that the model instability is reduced.

After conducting a comparative analysis of the performance of models constructed using three different ML methods, the results indicate that the four models constructed using the XGBoost method had the highest accuracy when compared to those constructed using the RF and GBM methods (Table 4). Except for Model A, the models constructed using the GBM method demonstrate better performance than those constructed using the RF method.

| Methods Technique | Model | <i>R</i> <sup>2</sup> | $M\!AE$ (g·kg $^{-1}$ ) | $RMSE$ (g·kg $^{-1}$ ) |
|-------------------|-------|-----------------------|-------------------------|------------------------|
|                   | А     | 0.411                 | 1.304                   | 1.760                  |
| DE                | В     | 0.456                 | 1.227                   | 1.621                  |
| KF                | С     | 0.472                 | 1.179                   | 1.543                  |
|                   | D     | 0.505                 | 1.092                   | 1.479                  |
|                   | А     | 0.378                 | 1.644                   | 2.455                  |
| CDM               | В     | 0.458                 | 1.487                   | 2.006                  |
| GBM               | С     | 0.481                 | 1.314                   | 1.841                  |
|                   | D     | 0.510                 | 1.224                   | 1.800                  |
|                   | А     | 0.615                 | 0.823                   | 1.162                  |
| YCP               | В     | 0.677                 | 0.661                   | 0.994                  |
| AGBoost           | С     | 0.714                 | 0.571                   | 0.939                  |
|                   | D     | 0.730                 | 0.554                   | 0.899                  |

Table 4. Evaluation and comparison of different models.

The accuracy of predicting CW-SOC content was influenced by both the ML methods and predictor variables. Using a combination of multi-source data as predictor variables, three ML methods were effective for modeling CW-SOC content, with the XGBoost method having the highest accuracy. The  $R^2$  value was as high as 0.730, and the *MAE* and *RMSE* values were both less than 0.9 g·kg<sup>-1</sup> (Table 4). This study found that the prediction accuracy ranking of the models constructed by the three ML methods was Model D > Model C > Model B > Model A. The combination of SAR and optical images improves prediction accuracy when compared with using a single type of remote sensing images. Adding climate variables further improved the accuracy of the model, though the improvement was not substantial.

#### 3.2. Relative Importance of Predictor Variables

The analysis of relative importance is used to reveal the contributions of relevant predictor variables in the predictive model and estimate their importance [62]. To obtain the relative importance of each variable in predicting CW-SOC content, the "variable importance" feature of the RF, GBM, and XGBoost models was used to calculate the relative importance of each predictor variable. Both RF and GBM calculate the reduction in weighted impurity of all non-leaf nodes when splitting variables. The greater the reduction, the more important the variable [63]. XGBoost utilizes the average gain brought by variables when used as split attributes. The larger the value, the stronger the importance [64]. The relative importance of different variables in Model D for predicting CW-SOC content was obtained (Figure 4). To enhance the comparability of predictor variables, they were normalized to 100%. Since topographic variables were screened out during the variable selection process, the relative importance is ranked only for SAR images, optical images, and climate variables. This also applies to the discussion of variable relative importance. Optical images were found to be the main explanatory variables for CW-SOC content prediction, accounting for over 65% of the relative importance, followed by SAR images, with climate variables having the lowest relative importance. In models with high prediction accuracy, the relative importance of band reflectance tends to be higher. In the XGBoost method, the relative importance of band reflectance exceeds 39%.



**Figure 4.** Using the RF, GBM, and XGBoost methods, we obtained the relative importance of different data in Model D for predicting the CW-SOC content; (a) RF; (b) GBM; (c) XGBoost.

Compared to the relative importance ranking of the top ten predictive variables in Model D obtained by RF, GBM, and XGBoost methods (Figure 5), the results showed that the three ML methods exhibited different characteristics in terms of importance ranking. The most important predictive variables for RF, GBM, and XGBoost methods were, respectively MARH (12.2%), DVI (18.1%), and B2 (37.6%). MARH was the only climate variable predicting CW-SOC content, ranked among the top three in all three ML methods, even becoming the most important predictor variables in the RF method. In any ML method, the number of optical images predictive variables was always among the top ten. SAR images explained 14.7%, 23.6%, and 18.5% of the variability in the RF, GBM, and XGBoost methods, respectively. Overall, in the three ML methods, optical images contributed the most, while SAR images and climate variables also made a certain degree of contribution.



Figure 5. The relative importance of the top 10 predictive variables in Model D was obtained using the RF, GBM, and XGBoost methods (normalized to 100%); (a) RF; (b) GBM; (c) XGBoost.

Optical images

Climate

#### 3.3. Spatial Distribution Prediction of the CW-SOC Content

SAR images

D 7.3%

Based on RF, GBM, and XGBoost, the model with the best performance in each method was selected for CW-SOC content prediction and drew the spatial distribution map of the CW-SOC content (Figure 6). The predicted range of CW-SOC content using the RF method was from  $3.174 \text{ g} \cdot \text{kg}^{-1}$  to  $14.079 \text{ g} \cdot \text{kg}^{-1}$ , with a mean and SD values of  $8.001 \text{ g} \cdot \text{kg}^{-1}$ and 1.681 g·kg<sup>-1</sup>, respectively. The GBM method predicted the CW-SOC content ranging from 0.455 g·kg<sup>-1</sup> to 14.923 g·kg<sup>-1</sup>, with a mean value of 6.857 g·kg<sup>-1</sup> and an SD value of 1.565 g·kg<sup>-1</sup>. The range of CW-SOC content predicted by the XGBoost method was 1.208 g·kg<sup>-1</sup> to 17.645 g·kg<sup>-1</sup>, with a mean and SD values of 6.236 g·kg<sup>-1</sup> and 1.862 g·kg<sup>-1</sup>, respectively (Table 5). The range, mean, and SD value of CW-SOC content predicted by the XGBoost method was closest to the measured value of CW-SOC content, which can better reflecting the spatial distribution of CW-SOC content.



Figure 6. Spatial distribution prediction of CW-SOC content based on the RF, GBM, and XGBoost methods.

| Methods<br>Technique | Area                 | Max<br>(g∙kg <sup>-1</sup> ) | Min<br>(g∙kg <sup>-1</sup> ) | Mean<br>(g∙kg <sup>-1</sup> ) | SD<br>(g·kg <sup>−1</sup> ) | CV<br>(%) |
|----------------------|----------------------|------------------------------|------------------------------|-------------------------------|-----------------------------|-----------|
| RF                   | Study area<br>Binhai | 14.079                       | 3.174                        | 8.001                         | 1.681                       | 21.01     |
|                      | New<br>District      | 14.079                       | 3.276                        | 8.629                         | 1.449                       | 16.79     |
|                      | Huanghua             | 13.149                       | 3.396                        | 7.660                         | 1.586                       | 20.70     |
|                      | Haixing              | 12.846                       | 3.174                        | 6.392                         | 1.284                       | 20.09     |
| GBM                  | Study area<br>Binhai | 14.923                       | 0.455                        | 6.857                         | 1.565                       | 22.82     |
|                      | New<br>District      | 14.844                       | 0.752                        | 7.463                         | 1.366                       | 18.30     |
|                      | Huanghua             | 14.923                       | 0.455                        | 6.575                         | 1.457                       | 22.16     |
|                      | Haixing              | 12.906                       | 0.709                        | 5.217                         | 0.952                       | 18.25     |
| XGBoost              | Study area<br>Binhai | 17.645                       | 1.208                        | 6.236                         | 1.862                       | 29.86     |
|                      | New<br>District      | 17.645                       | 1.379                        | 6.621                         | 1.815                       | 27.41     |
|                      | Huanghua             | 16.086                       | 1.208                        | 6.192                         | 1.829                       | 29.54     |
|                      | Haixing              | 17.645                       | 1.337                        | 4.984                         | 1.478                       | 29.65     |

Table 5. Summary statistics for predicted CW-SOC content.

The spatial distribution maps of CW-SOC content predicted by three ML methods (Figure 6) had similar spatial distribution characteristics. The CW-SOC content showed a gradually increasing trend from coastal to inland areas, which is most evident in the spatial distribution maps predicted by the RF and GBM methods. The southern and northern parts of the study area had lower CW-SOC content, while the central region had higher content, mainly near the Beidagang Wetlands Reserve and Nandagang Wetlands Reserve (Figure 6). According to the summary statistics of CW-SOC content in different counties and cities (Table 5), it was found that the CW-SOC content in Binhai new district was higher than those in Huanghua and Haixing.

## 4. Discussion

#### 4.1. Prediction Accuracy Comparison of Machine Learning Methods

By comparing this study to other related research, we aim to explore the model performance patterns of different data combinations and the performance of machine learning methods under different land-use types. This study found that the choice of ML methods and various combinations of predictor variables have a substantial impact on the accurate prediction of CW-SOC contents (Table 4). This study has showed that regardless of the combination of predictor variables used, the XGBoost method has the highest prediction accuracy for CW-SOC content. Xie et al. compared the prediction accuracy of the RF, GBM, and XGBoost methods for SOC content in the Ebinur Lake wetlands in Xinjiang and found that the XGBoost method had better prediction accuracy than RF and GBM methods (Table 6) [65]. However, when Zhang et al. predicted the SOC content in the dry land of Northeast China using the RF and XGBoost methods and found that the RF method outperformed XGBoost [23]. This implies that there is no single ML method that is suitable for all ecosystems [26]. At the same time, there are uncertainties in parameter adjustment and optimization of the three machine learning methods [66]. Additionally, the sample selection in the study can also have an impact on the model performance. Although machine learning methods are effective in estimating CW-SOC content, the presence of unknown internal nonlinear processes can introduce additional uncertainty to the model [67]. Therefore, it is crucial to evaluate the performance of different ML methods under different combinations of predictor variables.

| Land Cover | Depth   | Data                                  | Method  | R <sup>2</sup> | Literature |
|------------|---------|---------------------------------------|---------|----------------|------------|
|            |         |                                       | RF      | 0.583          |            |
|            |         | Landsat 8 (6band)                     | GBM     | 0.531          |            |
|            |         |                                       | XGBoost | 0.600          |            |
|            |         |                                       | RF      | 0.633          |            |
|            |         | Landsat 8 (6band) + Spectral index    | GBM     | 0.689          |            |
|            |         | _                                     | XGBoost | 0.677          |            |
|            |         | Landsat 8 (6band) + Spectral index +  | RF      | 0.627          |            |
|            |         | Climate variables + Topographic       | GBM     | 0.670          |            |
|            |         | variables                             | XGBoost | 0.693          |            |
|            |         | Landsat 8 (6band) + Spectral index +  | RF      | 0.681          |            |
|            |         | Climate variables + Topographic       | GBM     | 0.671          |            |
|            |         | variables + Sentinel-1A               | XGBoost | 0.701          |            |
|            |         |                                       | RF      | 0.615          |            |
| Wetland    | 0–10 cm | Sentinel-2A (6band)                   | GBM     | 0.626          |            |
|            |         |                                       | XGBoost | 0.685          |            |
|            |         |                                       | RF      | 0.632          |            |
|            |         | Sentinel-2A (6band) + Spectral index  | GBM     | 0.649          |            |
|            |         |                                       | XGBoost | 0.693          |            |
|            |         |                                       | RF      | 0.569          | [65]       |
|            |         | Sentinel-2A (6band) + Spectral index  | GBM     | 0.681          |            |
|            |         | + Climate + Topographic variables     | XGBoost | 0.712          |            |
|            |         | Sentinel-2A (6band) + Spectral index  | RF      | 0.701          |            |
|            |         | + Climate + Topographic variables +   | GBM     | 0.708          |            |
|            |         | Sentinel-1A                           | XGBoost | 0.735          |            |
|            |         |                                       | RF      | 0.615          |            |
|            |         | Sentinel-2A (10band)                  | GBM     | 0.659          |            |
|            |         | ( ,                                   | XGBoost | 0.694          |            |
|            |         |                                       | RF      | 0.693          |            |
|            |         | Sentinel-2A (10band) + Spectral index | GBM     | 0.663          |            |
|            |         | + Red-edge index                      | XGBoost | 0.715          |            |
|            |         | Sentinel-2A (10band) + Spectral index | RF      | 0.640          |            |
|            |         | + Red-edge index + Climate +          | GBM     | 0.687          |            |
|            |         | Topographic variables                 | XGBoost | 0.726          |            |
|            |         | Sentinel-2A (10band) + Spectral index | RF      | 0.705          |            |
|            |         | + Red-edge index + Climate +          | GBM     | 0.751          |            |
|            |         | Topographic variables +Sentinel-1A    | XGBoost | 0.771          |            |
|            |         |                                       | RF      | 0.411          |            |
|            |         | SAR images                            | GBM     | 0.378          |            |
|            |         | 0                                     | XGBoost | 0.615          |            |
|            |         |                                       | RF      | 0.456          |            |
|            |         | Optical images                        | GBM     | 0.458          |            |
|            |         | - F                                   | XGBoost | 0.677          |            |
|            | 0–30 cm |                                       | RF      | 0.472          | This study |
|            |         | SAR and optical images                | GBM     | 0.481          |            |
|            |         |                                       | XGBoost | 0.714          |            |
|            |         |                                       | RF      | 0.505          |            |
|            |         | SAR images, optical images, and       | GBM     | 0.510          |            |
|            |         | climate data                          | XGBoost | 0.730          |            |
|            |         |                                       |         |                |            |

 Table 6. Comparison of model accuracy between this study and related studies.

| Land Cover  | Depth    | Data  | Method  | <b>R</b> <sup>2</sup> | Literature |
|-------------|----------|---|---------|-----------------------|------------|
|             |          | SAR images  |         | 0.190                 |            |
|             |          | Optical images  |         | 0.500                 |            |
|             |          | SAR and optical images  |         | 0.560                 | [19]       |
|             | 0–20 cm  | Land use + climate + topography +<br>optical images               | RF      | 0.740                 |            |
|             |          | Land use + climate + topography +<br>SAR images + optical images) |         | 0.750                 |            |
| Dryland     | 0–10 cm  |   |         | 0.580                 |            |
|             | 10–20 cm |   | RF      | 0.710                 |            |
|             | 20–30 cm |   |         | 0.730                 |            |
|             | 30–40 cm | Soil and parent material, climate,                                |         | 0.740                 | [22]       |
|             | 0–10 cm  | organism, relief and remote sensing                               | XGBoost | 0.530                 | [23]       |
|             | 10–20 cm | variables   |         | 0.670                 |            |
|             | 20–30 cm |   |         | 0.700                 |            |
|             | 30–40 cm |   |         | 0.710                 |            |
|             |          | SAR images  |         | 0.160                 |            |
|             |          | Optical images  |         | 0.200                 |            |
| Forest land | 0–20 cm  | SAR and optical images  | RF      | 0.250                 | [13]       |
|             |          | Sentinel-1/2-derived predictors and DEM derivatives               |         | 0.400                 |            |

Table 6. Cont.

We found that among the three machine learning methods, the performance of Model B is consistently better than that of Model A, indicating that the prediction accuracy of optical images is higher than those of SAR images. Further, it showed that optical images provide more relevant information for predicting the CW-SOC content [19]. When optical images and SAR images are combined, the model's performance further improves. This was expected as the combination of the two types of images can complement each other and provide more effective information for building models. Previous research also confirmed the great potential of combining optical and SAR images in predicting SOC content [13,19]. For example, in Model C, which uses optical and SAR images as predictor variables, incorporating climate data into the model can improve its performance. However, the optimization effect of the model is not ideal, with an optimization amplitude of only 2.2–7%. In the study of Zhou and Xie et al., the optimal optimization amplitude reached 23.2% and 56%, respectively [13,65]. This could be caused by our study area being located in a coastal region where the effects of climate variables are not significant, thus their impact on the model accuracy is not significant.

It was found that the same ML method performs differently in different land use types. When Zhou et al. used the RF method to predict SOC content in drylands, their model had a higher prediction accuracy than this study [19]. However, when SAR images were used as predictor variables, the model performance of this study was better than the former. This was mainly because Zhou et al. only used VV and VH as predictor variables, while this study also took D as predictor variables, helping to improve the model's performance. When using the RF method to predict the SOC content of forest land, the prediction accuracy was lower than in this study. Perhaps it may be because Zhou et al. had fewer predictor variables, which cannot meet the needs of large-scale prediction, and did not consider the influence of climate factors.

#### 4.2. Influence of Predictor Variables on CW-SOC Content Prediction

In the three ML methods, optical images and SAR images were the main variables for predicting CW-SOC content. The relative importance of optical images and SAR images reached over 65% and 14%, respectively, indicating that optical images and SAR images could effectively explain the spatial variability of CW-SOC content [68]. Sentinel-2A was the most important factor for predicting CW-SOC content [69]. This was mainly because the

band reflectance and remote sensing indices of Sentinel-2A, especially the remote sensing indices involving the red-edge band, could provide characteristics that are related to the correlation between vegetation and soil properties [70]. Among the models with high prediction accuracy, the relative importance of the remote sensing indices in the red-edge band and red-edge band participating is relatively high. The relative importance of SAR imagery are 14.7%, 23.6%, and 18.5% in the RF, GBM, and XGBoost method, respectively. However, research by Zhou et al. indicates that the relative importance of SAR images is only 9% [19], which could be due to the contribution of SAR images to the model being dependent on the sensitivity of the backscatter coefficient to surface humidity and on this study area being coastal wetlands with a relatively high surface humidity [15]. In addition, some studies have mentioned that D obtained from mathematical calculations based on the VV and VH backscattering coefficients have higher relative importance than VV and VH [16].

In addition to remote sensing images, climate variables are important factors affecting the distribution of CW-SOC content. As a crucial factor influencing soil formation, climate impacts SOC content by affecting soil water and heat conditions, as well as the decomposition and transformation of SOC by microorganisms [71]. Previous studies have emphasized the importance of precipitation and temperature in influencing SOC distribution [72,73]. However, our results showed that regardless of which ML method is used for CW-SOC content prediction, MARH is always the most important predictor variable among climate variables, while MAT and MAP are both excluded during the variable selection process. This is mainly due to our study area being situated in a small coastal region where there are minor variations in precipitation and temperature. Hence, the CW-SOC content is not considerably affected. We observed a significant relationship between MARH and soil moisture. Any modifications in soil moisture levels can affect the exchange of water and energy between the land and the atmosphere [74], which can further impact plant growth and net primary productivity [72].

## 4.3. Spatial Distribution Characteristics of CW-SOC Content

Comparing the CW-SOC content predicted by the three ML methods, we found that the RF methods predicted a smaller range of CW-SOC content. This is due to the RF method deriving its prediction results from the average output values of multiple independent trees, making it more conservative in predicting CW-SOC content and less sensitive to extreme values [26].

The predicted range and spatial distribution patterns of CW-SOC content in this study are similar to the results obtained by Luo et al. [75]. The CW-SOC content predicted by the three ML methods had similar spatial distribution characteristics. The central part of the study area has higher CW-SOC content, mainly including Beidagang Wetlands Reserve and Nandagang Wetlands Reserve (Figure 6), while the northern and southern parts have lower content. The reason for this spatial distribution is that the central part of the study area has higher vegetation coverage than the southern and northern parts, especially in the Beidagang Wetlands Nature Reserve and Nandagang Wetlands Nature Reserve and Nandagang Wetlands Nature Reserve. High vegetation coverage can effectively protect soil, converting atmospheric carbon dioxide into organic carbon through photosynthesis and promoting the accumulation of CW-SOC [76,77]. On the other hand, the decomposition of vegetation litter by soil microorganisms can also promote the accumulation of CW-SOC [78]. We found that the CW-SOC content showed an increasing trend from the sea to the inland direction. This is mainly because the coastal areas are mostly tidal flats, and the vegetation coverage is lower than that of the inland areas [13].

Previous studies on this study area have some differences from the results compared to ours. Mao and Hao et al. studied the CW-SOC content in Tianjin and found that the average CW-SOC content was 16.304 g·kg<sup>-1</sup> and 8.55 g·kg<sup>-1</sup> [79,80], which is higher than the predicted CW-SOC content in this study. Whereas Li et al. found through the study of the salinity response and influencing factors of SOC in Tianjin coastal wetlands that

the average SOC content in the study area was  $5.40 \text{ g}\cdot\text{kg}^{-1}$  [81], which is lower than the results in this study. This discrepancy may result from the sampling points of Mao and Hao et al. being distributed in forests, grasslands, towns, harbors and tidal flat areas, while Li et al. only collected typical coastal wetlands soil samples with salinity differences in Tianjin. Our sampling points are distributed in coastal wetlands, and they are relatively evenly distributed. The different sampling areas and densities of soil samples resulted in different research results.

## 5. Conclusions

This study compared the performance of three ML methods based on decision tree models in predicting CW-SOC content and the prediction of the spatial distribution of CW-SOC content using SAR images, optical images, and climate and topographic data. The main conclusions of the study are summarized as follows:

- (1) Combining SAR images and optical images can effectively improve the prediction accuracy of the model. After adding climate variables, the performance of the model is further improved, but the optimization effect is not obvious, and the prediction accuracy is only increased by 7% (RF), 6% (GBM), and 2.2% (XGBoost).
- (2) XGBoost method exhibits better prediction ability than the RF and GBM method. The optimal model is built using the XGBoost method, with the  $R^2$  as high as 0.730, and the *MAE* and *RMSE* as low as 0.554 g·kg<sup>-1</sup> and 0.899 g·kg<sup>-1</sup>, respectively.
- (3) Remote sensing variables are the primary explanatory variables for predicting CW-SOC content, with optical images being the most prominent contributor, explaining more than 65% of the variability. The most important predictor variables for the RF, GBM, and XGBoost method were MARH (12.2%), DVI (18.1%), and B2 (37.6%), respectively.
- (4) CW-SOC content gradually increase from the coast to the inland. The CW-SOC content is lower in the south and north of the study area and higher in the central area. The mean value of CW-SOC content in Binhai New District is higher than those in Huanghua and Haixing.

Author Contributions: Conceptualization, Y.Z., C.K., M.L., F.L., C.L. and W.M.; methodology, Y.Z., C.K., M.L., F.L., J.S., T.S. and W.M.; software, Y.Z., C.K., Q.Z. and X.L.; validation, M.L. and W.M.; formal analysis, M.L. and W.M.; investigation, J.S., T.S., X.L. and D.T.; resources, Y.Z. and C.K.; data curation, C.K., Q.Z., X.L. and D.T.; writing—original draft preparation, Y.Z., C.K., M.L. and W.M.; writing—review and editing, M.L., C.L. and W.M.; visualization, C.K., Q.Z., X.L. and D.T.; supervision, Y.Z.; project administration, W.M.; funding acquisition, Y.Z., M.L. and W.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (Grant No. 41901375, 42101393 and 52274166); the Natural Science Foundation of Hebei Province, China (Grant No. D2022209005); the Science and Technology Project of Hebei Education Department (Grant No. BJ2020058); the Key Research and Development Program of Science and Technology Plan of Tangshan, China (Grant No. 22150221J); the North China University of Science and Technology Foundation (Grant No. BS201824 and BS201825); the Fostering Project for Science and Technology Research and Development Platform of Tangshan, China (No. 2020TS003b).

Data Availability Statement: Not applicable.

**Acknowledgments:** The authors would like to thank Hua Fang, Huifeng Gao, and Cheng Guan for processing soil samples. The authors are deeply grateful to the anonymous reviewers and the editor for their helpful comments on the manuscript.

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

## References

- 1. Rumpel, C.; Amiraslani, F.; Koutika, L.S.; Smith, P.; Whitehead, D.; Wollenberg, E. Put more carbon in soils to meet Paris climate pledges. *Nature* 2018, 564, 32–34. [CrossRef] [PubMed]
- Dharumarajan, S.; Kalaiselvi, B.; Suputhra, A.; Lalitha, M.; Vasundhara, R.; Kumar, K.S.A.; Nair, K.M.; Hegde, R.; Singh, S.K.; Lagacherie, P. Digital soil mapping of soil organic carbon stocks in Western Ghats, South India. *Geoderma Reg.* 2021, 25, e00387. [CrossRef]
- 3. Fernández-Martínez, M.; Peñuelas, J.; Chevallier, F.; Ciais, P.; Obersteiner, M.; Rödenbeck, C.; Sardans, J.; Vicca, S.; Yang, H.; Sitch, S.; et al. Diagnosing destabilization risk in global land carbon sinks. *Nature* **2023**, *615*, 848–853. [CrossRef]
- 4. Mao, D.; Luo, L.; Wang, Z.; Wilson, M.C.; Zeng, Y.; Wu, B.; Wu, J. Conversions between natural wetlands and farmland in China: A multiscale geospatial analysis. *Sci. Total Environ.* **2018**, *634*, 550–560. [CrossRef]
- 5. Xia, S.; Song, Z.; Van Zwieten, L.; Guo, L.; Yu, C.; Wang, W.; Li, Q.; Hartley, I.; Yang, Y.; Liu, H.; et al. Storage, patterns and influencing factors for soil organic carbon in coastal wetlands of China. *Glob. Chang. Biol.* **2022**, *28*, 6065–6085. [CrossRef]
- Lausch, A.; Baade, J.; Bannehr, L.; Borg, E.; Bumberger, J.; Chabrilliat, S.; Dietrich, P.; Gerighausen, H.; Glässer, C.; Hacker, J.M.; et al. Linking Remote Sensing and Geodiversity and Their Traits Relevant to Biodiversity—Part I: Soil Characteristics. *Remote Sens.* 2019, 11, 2356. [CrossRef]
- Wang, F.; Lu, X.; Sanders, C.J.; Tang, J. Tidal wetland resilience to sea level rise increases their carbon sequestration capacity in United States. *Nat. Commun.* 2019, 10, 5434. [CrossRef] [PubMed]
- 8. Mao, D.; Yang, H.; Wang, Z.; Song, K.; Thompson, J.R.; Flower, R.J. Reverse the hidden loss of China's wetlands. *Science* 2022, 376, 1061. [CrossRef]
- Song, J.; Gao, J.; Zhang, Y.; Li, F.; Man, W.; Liu, M.; Wang, J.; Li, M.; Zheng, H.; Yang, X.; et al. Estimation of Soil Organic Carbon Content in Coastal Wetlands with Measured VIS-NIR Spectroscopy Using Optimized Support Vector Machines and Random Forests. *Remote Sens.* 2022, 14, 4372. [CrossRef]
- 10. Zhang, T.; Zhang, W.; Yang, R.; Liu, Y.; Jafari, M. CO<sub>2</sub> capture and storage monitoring based on remote sensing techniques: A review. *J. Clean. Prod.* **2021**, *281*, 124409. [CrossRef]
- Li, Z.; Liu, F.; Peng, X.; Hu, B.; Song, X. Synergetic use of DEM derivatives, Sentinel-1 and Sentinel-2 data for mapping soil properties of a sloped cropland based on a two-step ensemble learning method. *Sci. Total Environ.* 2023, 866, 161421. [CrossRef] [PubMed]
- 12. Lin, C.; Zhu, A.; Wang, Z.; Wang, X.; Ma, R. The refined spatiotemporal representation of soil organic matter based on remote images fusion of Sentinel-2 and Sentinel-3. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *89*, 102094. [CrossRef]
- Zhou, T.; Geng, Y.; Chen, J.; Pan, J.; Haase, D.; Lausch, A. High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms. *Sci. Total Environ.* 2020, 729, 138244. [CrossRef] [PubMed]
- 14. Were, K.; Bui, D.T.; Dick, Ø.B.; Singh, B.R. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. *Ecol. Indic.* **2015**, *52*, 394–403. [CrossRef]
- 15. Chen, S.; Zhang, W.; Li, Z.; Wang, Y.; Zhang, B. Cloud Removal with SAR-Optical Data Fusion and Graph-Based Feature Aggregation Network. *Remote Sens.* 2022, 14, 3374. [CrossRef]
- Yang, R.-M.; Guo, W.-W. Using time-series Sentinel-1 data for soil prediction on invaded coastal wetlands. *Environ. Monit. Assess.* 2019, 191, 462. [CrossRef] [PubMed]
- 17. Yang, R.-M.; Guo, W.-W. Modelling of soil organic carbon and bulk density in invaded coastal wetlands using Sentinel-1 imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *82*, 101906. [CrossRef]
- Veloso, A.; Mermoz, S.; Bouvet, A.; Le Toan, T.; Planells, M.; Dejoux, J.-F.; Ceschia, E. Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. *Remote Sens. Environ.* 2017, 199, 415–426. [CrossRef]
- 19. Zhou, T.; Geng, Y.; Chen, J.; Liu, M.; Haase, D.; Lausch, A. Mapping soil organic carbon content using multi-source remote sensing variables in the Heihe River Basin in China. *Ecol. Indic.* **2020**, *114*, 106288. [CrossRef]
- 20. Azizi, K.; Garosi, Y.; Ayoubi, S.; Tajik, S. Integration of Sentinel-1/2 and topographic attributes to predict the spatial distribution of soil texture fractions in some agricultural soils of western Iran. *Soil Tillage Res.* **2023**, *229*, 105681. [CrossRef]
- 21. van der Westhuizen, S.; Heuvelink, G.B.M.; Hofmeyr, D.P. Multivariate random forest for digital soil mapping. *Geoderma* **2023**, 431, 116365. [CrossRef]
- 22. Akinci, H.; Zeybek, M.; Dogan, S. Evaluation of landslide susceptibility of Şavşat District of Artvin Province (Turkey) using machine learning techniques. In *Landslides*; IntechOpen: London, UK, 2021.
- 23. Zhang, X.; Xue, J.; Chen, S.; Wang, N.; Shi, Z.; Huang, Y.; Zhuo, Z. Digital Mapping of Soil Organic Carbon with Machine Learning in Dryland of Northeast and North Plain China. *Remote Sens.* **2022**, *14*, 2504. [CrossRef]
- Sun, S.; Zhang, Y.; Song, Z.; Chen, B.; Zhang, Y.; Yuan, W.; Chen, C.; Chen, W.; Ran, X.; Wang, Y. Mapping Coastal Wetlands of the Bohai Rim at a Spatial Resolution of 10 m Using Multiple Open-Access Satellite Data and Terrain Indices. *Remote Sens.* 2020, 12, 4114. [CrossRef]
- 25. Mou, X.; Liu, X.; Yan, B.; Cui, B. Classification system of coastal wetlands in China. Wetl. Sci. 2015, 13, 19–26. [CrossRef]
- Zhang, Q.; Liu, M.; Zhang, Y.; Mao, D.; Li, F.; Wu, F.; Song, J.; Li, X.; Kou, C.; Li, C.; et al. Comparison of Machine Learning Methods for Predicting Soil Total Nitrogen Content Using Landsat-8, Sentinel-1, and Sentinel-2 Images. *Remote Sens.* 2023, 15, 2907. [CrossRef]

- 27. Goovaerts, P. Geostatistical modelling of uncertainty in soil science. Geoderma 2001, 103, 3–26. [CrossRef]
- 28. Navarro, A.; Rolim, J.; Miguel, I.; Catalão, J.; Silva, J.; Painho, M.; Vekerdy, Z. Crop Monitoring Based on SPOT-5 Take-5 and Sentinel-1A Data for the Estimation of Crop Water Requirements. *Remote Sens.* **2016**, *8*, 525. [CrossRef]
- 29. Yang, J.; Fan, J.; Lan, Z.; Mu, X.; Wu, Y.; Xin, Z.; Miping, P.; Zhao, G. Improved Surface Soil Organic Carbon Mapping of SoilGrids250m Using Sentinel-2 Spectral Images in the Qinghai–Tibetan Plateau. *Remote Sens.* **2023**, *15*, 114. [CrossRef]
- Cui, Z.; Kerekes, J.P. Potential of Red Edge Spectral Bands in Future Landsat Satellites on Agroecosystem Canopy Green Leaf Area Index Retrieval. *Remote Sens.* 2018, 10, 1458. [CrossRef]
- Wang, X.; Li, Y.; Gong, X.; Niu, Y.; Chen, Y.; Shi, X.; Li, W. Storage, pattern and driving factors of soil organic carbon in an ecologically fragile zone of northern China. *Geoderma* 2019, 343, 155–165. [CrossRef]
- Nguyen, T.T.; Pham, T.D.; Nguyen, C.T.; Delfos, J.; Archibald, R.; Dang, K.B.; Hoang, N.B.; Guo, W.; Ngo, H.H. A novel intelligence approach based active and ensemble learning for agricultural soil organic carbon prediction using multispectral and SAR data fusion. *Sci. Total Environ.* 2022, *804*, 150187. [CrossRef] [PubMed]
- Rouse, J., Jr.; Haas, R.; Schell, J.; Deering, D. Monitoring vegetation systems in the Great Plains with ERTS. In Proceedings of the Third Earth Resources Technology Satellite-1 Symposium, Washington, DC, USA, 10–14 December 1973; NASA: Washington, DC, USA, 1974; pp. 309–317.
- 34. Gao, B.-C. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* **1996**, *58*, 257–266. [CrossRef]
- 35. Zha, Y.; Gao, J.; Ni, S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *Int. J. Remote Sens.* **2003**, *24*, 583–594. [CrossRef]
- 36. Huete, A.R. A soil-adjusted vegetation index (SAVI). Remote Sens. Environ. 1988, 25, 295–309. [CrossRef]
- 37. Birth, G.S.; McVey, G.R. Measuring the Color of Growing Turf with a Reflectance Spectrophotometer1. *Agron. J.* **1968**, *60*, 640–643. [CrossRef]
- Richardsons, A.J.; Wiegand, A. Distinguishing vegetation from soil background information. *Photogramm. Eng. Remote Sens.* 1977, 43, 1541–1552.
- 39. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [CrossRef]
- Rikimaru, A. Landsat T M Data Processing Guide for Forest Canopy Density Mapping and Monitoring Model. In Proceedings of the ITTO Workshop on Utilization of Remote Sens-ing in Site Assessment and Planning for Rehabilitation of Logged-Over Forest, Bangkok, Thailand, 30 July–1 August 1996; pp. 1–8.
- Gitelson, A.; Merzlyak, M.N. Spectral Reflectance Changes Associated with Autumn Senescence of Aesculus hippocastanum L. and Acer platanoides L. Leaves. Spectral Features and Relation to Chlorophyll Estimation. J. Plant Physiol. 1994, 143, 286–292. [CrossRef]
- Gitelson, A.A.; Viña, A.; Ciganda, V.; Rundquist, D.C.; Arkebauer, T.J. Remote estimation of canopy chlorophyll content in crops. *Geophys. Res. Lett.* 2005, 32, L08403. [CrossRef]
- Gitelson, A.A.; Merzlyak, M.N. Remote estimation of chlorophyll content in higher plant leaves. Int. J. Remote Sens. 1997, 18, 2691–2697. [CrossRef]
- 44. Chen, L.; Ren, C.; Li, L.; Wang, Y.; Zhang, B.; Wang, Z.; Li, L. A Comparative Assessment of Geostatistical, Machine Learning, and Hybrid Approaches for Mapping Topsoil Organic Carbon Content. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 174. [CrossRef]
- Guo, Z.; Li, Y.; Wang, X.; Gong, X.; Chen, Y.; Cao, W. Remote Sensing of Soil Organic Carbon at Regional Scale Based on Deep Learning: A Case Study of Agro-Pastoral Ecotone in Northern China. *Remote Sens.* 2023, 15, 3846. [CrossRef]
- 46. Wang, C.; Zhao, L.; Fang, H.; Wang, L.; Xing, Z.; Zou, D.; Hu, G.; Wu, X.; Zhao, Y.; Sheng, Y.; et al. Mapping Surficial Soil Particle Size Fractions in Alpine Permafrost Regions of the Qinghai–Tibet Plateau. *Remote Sens.* **2021**, *13*, 1392. [CrossRef]
- Huang, T.; Ou, G.; Wu, Y.; Zhang, X.; Liu, Z.; Xu, H.; Xu, X.; Wang, Z.; Xu, C. Estimating the Aboveground Biomass of Various Forest Types with High Heterogeneity at the Provincial Scale Based on Multi-Source Data. *Remote Sens.* 2023, 15, 3550. [CrossRef]
- Liu, Y.; Yue, Q.; Wang, Q.; Yu, J.; Zheng, Y.; Yao, X.; Xu, S. A Framework for Actual Evapotranspiration Assessment and Projection Based on Meteorological, Vegetation and Hydrological Remote Sensing Products. *Remote Sens.* 2021, 13, 3643. [CrossRef]
- Zhang, N.; Chen, M.; Yang, F.; Yang, C.; Yang, P.; Gao, Y.; Shang, Y.; Peng, D. Forest Height Mapping Using Feature Selection and Machine Learning by Integrating Multi-Source Satellite Data in Baoding City, North China. *Remote Sens.* 2022, 14, 4434. [CrossRef]
- 50. Raj, N.; Brown, J. An EEMD-BiLSTM Algorithm Integrated with Boruta Random Forest Optimiser for Significant Wave Height Forecasting along Coastal Areas of Queensland, Australia. *Remote Sens.* **2021**, *13*, 1456. [CrossRef]
- Zhang, Y.; Liu, J.; Li, W.; Liang, S. A Proposed Ensemble Feature Selection Method for Estimating Forest Aboveground Biomass from Multiple Satellite Data. *Remote Sens.* 2023, 15, 1096. [CrossRef]
- Shafizadeh-Moghadam, H.; Weng, Q.; Liu, H.; Valavi, R. Modeling the spatial variation of urban land surface temperature in relation to environmental and anthropogenic factors: A case study of Tehran, Iran. *GIScience Remote Sens.* 2020, 57, 483–496. [CrossRef]
- 53. Tamiru, B.; Soromessa, T.; Warkineh, B.; Legese, G. Mapping Soil Parameters with Environmental Covariates and Land Cover Projection in Tropical Rainforest, Hangadi Watershed, Ethiopia. *Sustainability* **2023**, *15*, 1066. [CrossRef]

- 54. Zhang, H.; Wu, P.; Yin, A.; Yang, X.; Zhang, M.; Gao, C. Prediction of soil organic carbon in an intensively managed reclamation zone of eastern China: A comparison of multiple linear regressions and the random forest model. *Sci. Total Environ.* **2017**, *592*, 704–713. [CrossRef] [PubMed]
- 55. Zhou, T.; Geng, Y.; Ji, C.; Xu, X.; Wang, H.; Pan, J.; Bumberger, J.; Haase, D.; Lausch, A. Prediction of soil organic carbon and the C:N ratio on a national scale using machine learning and satellite data: A comparison between Sentinel-2, Sentinel-3 and Landsat-8 images. *Sci. Total Environ.* 2021, 755, 142661. [CrossRef] [PubMed]
- 56. Yang, L.; Zhang, X.; Liang, S.; Yao, Y.; Jia, K.; Jia, A. Estimating Surface Downward Shortwave Radiation over China Based on the Gradient Boosting Decision Tree Method. *Remote Sens.* **2018**, *10*, 185. [CrossRef]
- 57. Lu, Q.; Tian, S.; Wei, L. Digital mapping of soil pH and carbonates at the European scale using environmental variables and machine learning. *Sci. Total Environ.* 2023, *856*, 159171. [CrossRef]
- 58. Mahmoudzadeh, H.; Matinfar, H.R.; Taghizadeh-Mehrjardi, R.; Kerry, R. Spatial prediction of soil organic carbon using machine learning techniques in western Iran. *Geoderma Reg.* 2020, *21*, e00260. [CrossRef]
- 59. Stojić, A.; Stanić, N.; Vuković, G.; Stanišić, S.; Perišić, M.; Šoštarić, A.; Lazić, L. Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition. *Sci. Total Environ.* **2019**, 653, 140–147. [CrossRef]
- 60. Emadi, M.; Taghizadeh-Mehrjardi, R.; Cherati, A.; Danesh, M.; Mosavi, A.; Scholten, T. Predicting and Mapping of Soil Organic Carbon Using Machine Learning Algorithms in Northern Iran. *Remote Sens.* **2020**, *12*, 2234. [CrossRef]
- Siqueira, R.G.; Moquedace, C.M.; Francelino, M.R.; Schaefer, C.E.G.R.; Fernandes-Filho, E.I. Machine learning applied for Antarctic soil mapping: Spatial prediction of soil texture for Maritime Antarctica and Northern Antarctic Peninsula. *Geoderma* 2023, 432, 116405. [CrossRef]
- 62. Mizumoto, A. Calculating the Relative Importance of Multiple Regression Predictor Variables Using Dominance Analysis and Random Forests. *Lang. Learn.* 2023, 73, 161–196. [CrossRef]
- An, R.; Tong, Z.; Ding, Y.; Tan, B.; Wu, Z.; Xiong, Q.; Liu, Y. Examining non-linear built environment effects on injurious traffic collisions: A gradient boosting decision tree analysis. *J. Transp. Health* 2022, 24, 101296. [CrossRef]
- Shi, X.; Wong, Y.D.; Li, M.Z.-F.; Palanisamy, C.; Chai, C. A feature learning approach based on XGBoost for driving assessment and risk prediction. *Accid. Anal. Prev.* 2019, 129, 170–179. [CrossRef] [PubMed]
- 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] [PubMed]
- 66. Chen, S.; Liu, W.; Feng, P.; Ye, T.; Ma, Y.; Zhang, Z. Improving Spatial Disaggregation of Crop Yield by Incorporating Machine Learning with Multisource Data: A Case Study of Chinese Maize Yield. *Remote Sens.* **2022**, *14*, 2340. [CrossRef]
- 67. Han, J.; Zhang, Z.; Cao, J.; Luo, Y.; Zhang, L.; Li, Z.; Zhang, J. Prediction of Winter Wheat Yield Based on Multi-Source Data and Machine Learning in China. *Remote Sens.* **2020**, *12*, 236. [CrossRef]
- 68. Wang, S.; Gao, J.; Zhuang, Q.; Lu, Y.; Gu, H.; Jin, X. Multispectral Remote Sensing Data Are Effective and Robust in Mapping Regional Forest Soil Organic Carbon Stocks in a Northeast Forest Region in China. *Remote Sens.* **2020**, *12*, 393. [CrossRef]
- 69. 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]
- Castaldi, F.; Hueni, A.; Chabrillat, S.; Ward, K.; Buttafuoco, G.; Bomans, B.; Vreys, K.; Brell, M.; van Wesemael, B. Evaluating the capability of the Sentinel 2 data for soil organic carbon prediction in croplands. *ISPRS J. Photogramm. Remote Sens.* 2019, 147, 267–282. [CrossRef]
- 71. Wang, Z.; Zhang, Y.; Govers, G.; Tang, G.; Quine, T.A.; Qiu, J.; Navas, A.; Fang, H.; Tan, Q.; Van Oost, K. Temperature effect on erosion-induced disturbances to soil organic carbon cycling. *Nat. Clim. Chang.* **2023**, *13*, 174–181. [CrossRef]
- 72. Wang, B.; Waters, C.; Orgill, S.; Gray, J.; Cowie, A.; Clark, A.; Liu, D.L. High resolution mapping of soil organic carbon stocks using remote sensing variables in the semi-arid rangelands of eastern Australia. *Sci. Total Environ.* **2018**, 630, 367–378. [CrossRef]
- Zhang, Y.; Jiang, Y.; Jia, Z.; Qiang, R.; Gao, Q. Identifying the scale-controlling factors of soil organic carbon in the cropland of Jilin Province, China. *Ecol. Indic.* 2022, 139, 108921. [CrossRef]
- Materia, S.; Ardilouze, C.; Prodhomme, C.; Donat, M.G.; Benassi, M.; Doblas-Reyes, F.J.; Peano, D.; Caron, L.-P.; Ruggieri, P.; Gualdi, S. Summer temperature response to extreme soil water conditions in the Mediterranean transitional climate regime. *Clim. Dyn.* 2022, *58*, 1943–1963. [CrossRef]
- 75. Luo, M.; Guo, L.; Zhang, H.; Wang, S.; Liang, P. Characterization of Spatial Distribution of Soil Organic Carbon in China Based on Environmental Variables. *Acta Pedol. Sin.* **2020**, *57*, 48–59. [CrossRef]
- 76. Liu, X.; Lu, X.; Yu, R.; Sun, H.; Li, X.; Li, X.; Qi, Z.; Liu, T.; Lu, C. Distribution and storage of soil organic and inorganic carbon in steppe riparian wetlands under human activity pressure. *Ecol. Indic.* **2022**, *139*, 108945. [CrossRef]
- Bao, T.; Jia, G.; Xu, X. Weakening greenhouse gas sink of pristine wetlands under warming. *Nat. Clim. Chang.* 2023, 13, 462–469. [CrossRef]
- Li, J.; Zhang, T.; Meng, B.; Rudgers, J.A.; Cui, N.; Zhao, T.; Chai, H.; Yang, X.; Sternberg, M.; Sun, W. Disruption of fungal hyphae suppressed litter-derived C retention in soil and N translocation to plants under drought-stressed temperate grassland. *Geoderma* 2023, 432, 116396. [CrossRef]
- 79. Mao, T.; Shi, T.; Li, Y. Capacity estimation of soil organic carbon pools in the intertidal zone of the Bohai Bay. *IOP Conf. Ser. Earth Environ. Sci.* **2018**, 128, 012140. [CrossRef]

- 80. Hao, C.; Li, H.; Li, S.; Meng, W.; Wu, X.; Wang, X. Analysis of Soil Organic Carbon Storage and Influencing Factors in the Soil of Binhai Wetland in Tianjin. *Res. Environ. Sci.* 2011, 24, 1276–1282. [CrossRef]
- Li, S.; Guan, D.; Li, X.; Zhang, J.; Teng, H. Changes in Response to Salinity and Influencing Factors of Soil Organic Carbon and Available Phosphorus in Tianjin Coastal Wetland. *Chin. J. Ecol.* 2023. *in press* (In Chinese). Available online: https: //kns.cnki.net/kcms/detail/21.1148.Q.20230309.1047.006.html (accessed on 13 March 2023).

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.