

Article

Estimation of Maize Residue Cover Using Remote Sensing Based on Adaptive Threshold Segmentation and CatBoost Algorithm

Nan Lin ^{1,2,*}, Xunhu Ma ¹, Ranzhe Jiang ³, Menghong Wu ^{1,4} and Wenchun Zhang ¹

¹ College of Surveying and Exploration Engineering, Jilin Jianzhu University, Changchun 130118, China; maxunhu@student.jlju.edu.cn (X.M.); wumenghong@jlju.edu.cn (M.W.); zhangwenchun@jlju.edu.cn (W.Z.)

² Jilin Province Natural Resources Remote Sensing Information Technology Innovation Laboratory, Changchun 130118, China

³ College of Biological and Agricultural Engineering, Jilin University, Changchun 130012, China; jiangrz23@mails.jlu.edu.cn

⁴ College of Resource and Environmental Science, Jilin Agricultural University, Changchun 130118, China

* Correspondence: linnan@jlju.edu.cn

Abstract: Maize residue cover (MRC) is an important parameter to quantify the degree of crop residue cover in the field and its spatial distribution characteristics. It is also a key indicator of conservation tillage. Rapid and accurate estimation of maize residue cover (MRC) and spatial mapping are of great significance to increasing soil organic carbon, reducing wind and water erosion, and maintaining soil and water. Currently, the estimation of maize residue cover in large areas suffers from low modeling accuracy and poor working efficiency. Therefore, how to improve the accuracy and efficiency of maize residue cover estimation has become a research hotspot. In this study, adaptive threshold segmentation (Yen) and the CatBoost algorithm are integrated and fused to construct a residue coverage estimation method based on multispectral remote sensing images. The maize planting areas in and around Sihe Town in Jilin Province, China, were selected as typical experimental regions, and the unmanned aerial vehicle (UAV) was employed to capture maize residue cover images of sample plots within the area. The Yen algorithm was applied to calculate and analyze maize residue cover. The successive projections algorithm (SPA) was used to extract spectral feature indices from Sentinel-2A multispectral images. Subsequently, the CatBoost algorithm was used to construct a maize residue cover estimation model based on spectral feature indices, thereby plotting the spatial distribution map of maize residue cover in the experimental area. The results show that the image segmentation based on the Yen algorithm outperforms traditional segmentation methods, with the highest Dice coefficient reaching 81.71%, effectively improving the accuracy of maize residue cover recognition in sample plots. By combining the spectral index calculation with the SPA algorithm, the spectral features of the images are effectively extracted, and the spectral feature indices such as NDTI and STI are determined. These indices are significantly correlated with maize residue cover. The accuracy of the maize residue cover estimation model built using the CatBoost model surpasses that of traditional machine learning models, with a maximum determination coefficient (R^2) of 0.83 in the validation set. The maize residue cover estimation model constructed based on the Yen and CatBoost algorithms effectively enhances the accuracy and reliability of estimating maize residue cover in large areas using multispectral imagery, providing accurate and reliable data support and services for precision agriculture and conservation tillage.

Keywords: maize residue cover (MRC); conservation tillage; multispectral remote sensing images; adaptive threshold segmentation; Yen algorithm; CatBoost algorithm; spectral feature indices; spatial mapping



Citation: Lin, N.; Ma, X.; Jiang, R.; Wu, M.; Zhang, W. Estimation of Maize Residue Cover Using Remote Sensing Based on Adaptive Threshold Segmentation and CatBoost Algorithm. *Agriculture* **2024**, *14*, 711. <https://doi.org/10.3390/agriculture14050711>

Academic Editor: Jiehao Li

Received: 26 March 2024

Revised: 26 April 2024

Accepted: 29 April 2024

Published: 30 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Crop residue management is one of the most important conservation measures in modern conservation tillage techniques [1]. As an end product of crop production, maize

residue helps reduce water evaporation and retain moisture in the topsoil layer. In addition, residue cover can slow the weathering and erosion of agricultural soils. Returning residues to the field increases straw mineralization, enhances soil fertility, and provides natural organic matter for crop growth [2–4]. Estimation of maize residue cover is an important task in crop residue management, as rapid and accurate access to straw cover information can not only master the spatial distribution of conservation tillage but also macroscopically monitor the process and scope of conservation tillage implementation and improve the efficiency of farmland quality supervision.

Currently, the estimation of maize residue cover is usually performed by traditional methods such as random sampling and pull-string methods [5–7]. However, conducting large-area MRC surveys using traditional methods is time-consuming and costly. With the development of remote sensing spatial information technology, multispectral remote sensing technology has been widely used to estimate maize residue cover in farmland on a large spatial-regional scale because of its rapid and accurate characteristics [5,7]. Some studies have shown that crop residues and soils have similar spectral characteristics. Since certain characteristics of crop residues are related to cellulose and lignin, crop residues have unique absorption features near 2100 nm [8,9]. These features serve as the basis for distinguishing crop residues from soil based on optical remote sensing images, which provides a research basis for developing a model to estimate maize residue cover. The relationships between the vegetation index (PVI) and fractional vegetation cover (P_V), and between the nonphotosynthetic vegetation index (NPVI) and fractional nonphotosynthetic vegetation cover (NP_V) were analyzed by Guo et al. [10]. Their results showed a significant linear correlation between the global environment monitoring index (GEMI) and P_V as well as a strong linear relationship between the Dead Fuel Index (DFI) and NP_V . Daughtry et al. measured the spectral reflectance of crops such as corn and soybeans as well as soil cover in the wavelength range of 400–2400 nm, confirming the feasibility of using the cellulose absorption index for residue cover identification [11,12].

Currently, the methods for extracting maize residue cover primarily involve human-computer interaction for the visual recognition of maize cover and automatic recognition using image segmentation algorithms [7]. With the integration of AI technologies, these methods have been significantly enhanced [13,14]. AI-driven image segmentation algorithms are noted for their time-saving, labor-efficient, and high-accuracy attributes, demonstrating immense potential for large-scale agricultural monitoring and precise agricultural management [13,15]. Common image segmentation methods include those based on spectral indices and thresholding methods based on spatial grayscale value distributions [16,17]. Threshold segmentation methods, such as the maximum variance between two classes (OSTU, named by Japanese Nobuyuki Otsu) and MaxEntropy algorithms, are known for their computational efficiency and simplicity and are widely used in various fields and for crop image recognition [18–20]. Unlike traditional threshold segmentation algorithms that struggle with complex grayscale distributions, AI-driven approaches such as the Yen algorithm significantly improve the ability to process images with complex distributions and uneven backgrounds. The Yen algorithm, which applies principles of inter-class variance and entropy, is particularly effective in handling multimodal distributions and challenging segmentation tasks in scenarios with significant textural differences or uneven lighting conditions. This algorithm not only enhances the accuracy and adaptability of image segmentation in diverse areas, but AI integration facilitates more precise identification and classification of agricultural features, thereby enhancing the process of monitoring and managing farmland. Moreover, these AI-driven technologies contribute significantly to sustainable farming initiatives by enabling more efficient use of land and resources [21,22].

With the rapid development of machine learning theory and technology, building estimation models for remote sensing spectral features and maize residue cover based on machine learning algorithms has become a popular research topic in precision agricultural engineering [7,23]. However, traditional machine learning models often encounter issues such as high sensitivity to sample data, increased computational complexity in the hy-

pothesis space, model overfitting owing to limited training sample data, and a mismatch between model capacity and training dataset size. Consequently, these models tend to have poor estimation accuracy and efficiency. Ensemble learning is an efficient machine learning method based on the synergy of multiple algorithms. It integrates multiple learning results using different learners for training and adopts a certain combination strategy to achieve multiple learning results rather than using a single learner [24–26]. This learning method can effectively solve the problems of traditional machine learning, such as high sensitivity to training samples, high computational complexity, and overfitting. Currently, various ensemble learning algorithms have been widely used in biology, engineering, medicine, computer vision, image processing, and other fields [27,28] and have become a research hotspot in the field of machine learning [24]. In terms of learner combination rules, the commonly used ensemble learning methods include bagging parallel and boosting serial integration learning. Various ensemble learning model algorithms have appeared in recent years, such as gradient boosting decision tree (GBDT) and extreme gradient boosting (XGBoost) [29]. Among them, CatBoost, an improvement of GBDT, shows significant potential for improving the accuracy and stability of model predictions compared with other ensemble learning models because of its efficient gradient estimation and adaptability to data distribution, as well as its ability to deal with complex patterns and nonlinear relationships [30–33].

This study aimed to construct a remote sensing estimation method for residue cover based on the Yen image segmentation algorithm and CatBoost model, addressing the current issue of wide-range, rapid, and accurate estimation of agricultural residue cover. Therefore, this study selects Sihe Town and the surrounding maize-growing areas in Jilin Province, China, as the research area. Sentinel-2A multispectral remote sensing data were used to construct the spectral index. Combined with the real-time images captured by UAVs, the maize residue cover of a single image was calculated by applying the Yen image segmentation algorithm. Subsequently, the CatBoost model was used to construct a prediction model between maize residue cover and the spectral indices of the multispectral images for estimation, and the spatial distribution map of maize residue cover in the study area was plotted. The objectives of this research included (1) proposing an adaptive threshold segmentation algorithm based on Yen, aimed at implementing a fast, convenient, and high-precision method for calculating residue cover in images; (2) constructing a residue cover estimation model by analyzing the correlation between residue cover and spectral indices; and (3) using the optimized predictive CatBoost model to create a spatial distribution map of residue cover over a large area.

2. Materials and Methods

2.1. Overview of the Study Area

The study area was located in Sihe Town and its surrounding maize-planting areas in Jilin Province, China (Figure 1). This region, situated in the northeastern part of Yushu City, lies between 126°01′ to 127°05′ east longitude and 44°30′ to 45°15′ north latitude. Yushu City is at the heart of the agricultural region of Jilin Province and is characterized by a temperate continental monsoon climate. The terrain is relatively flat, with slight undulating waves and an average elevation of 157–220 m. The annual average temperature is 4.4 °C, with an average annual precipitation of 680 mm and >2200 h of sunshine per year. The soil is fertile and well suited for crop cultivation. The predominant agricultural products in this area are corn, soybean, and rice. According to the 2022 national economic and social development statistical data for Yushu City, Jilin Province, the total grain planting area of the city is 3807.48 km², of which maize accounts for 2935.46 km², accounting for 77.1% of the total planting area. This makes it an important maize planting base in Northeast China. The geographical location of the study area is shown in Figure 1:

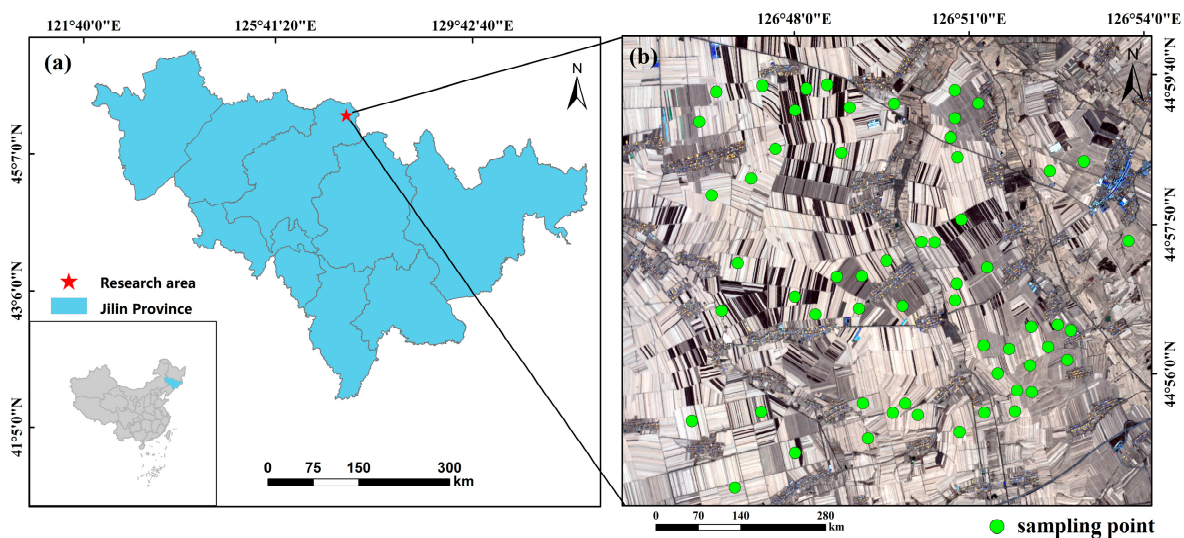
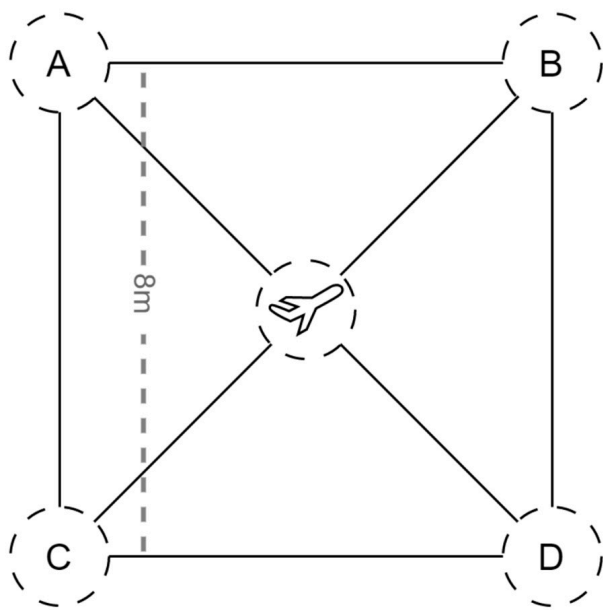


Figure 1. (a) Research zone located in Sihe Town, Jilin Province, Northeastern China; (b) geographical layout of 57 soil samples collected within the study area.

2.2. Collection and Preprocessing of Residue Cover Image Data

A DJI drone (Phantom 3) was used to collect residue mulch data in the study area, and a field collection experiment was conducted in the first half of November 2022. To ensure sufficient light conditions during the flight of the drone, sunny and windless weather conditions were selected for the image acquisition of the residue cover area. Initially, 57 sampling points were established in the study area, which were set up as $8\text{ m} \times 8\text{ m}$ square plots, as illustrated in the schematic (Figure 2a). Five markers were selected for each sampling point. The interval between sampling points was approximately 500 m. The geographic coordinates of the central points of the sampling plots were recorded using a global navigation satellite system (GNSS) to ensure precise georeferencing of each image captured. The drone was flown at an altitude of approximately 10 m above ground level to optimize the balance between area coverage and image resolution. The flight parameters included an 80% front overlap and a 70% side overlap, with a flight speed maintained at 5 m/s. The ground sampling distance (GSD) achieved was approximately 0.74 cm/pixel, which is suitable for detailed residue analysis. The UAV was equipped with a 1/2.3-inch CMOS sensor capable of capturing high-resolution imagery with 12 million effective pixels. The camera's sensitivity was adaptive to accommodate varying light conditions. Aerial photography followed a fixed flight path, capturing 10 visible-light images in a single operation, with each image having a resolution of 3000×3000 pixels. To ensure data quality, each image underwent quality screening to exclude any that were blurry, overexposed, or underexposed. The most comprehensive and clear images from each plot were selected for subsequent analyses. To avoid extensive shadows over the soil cover, all the images were captured between 11:00 AM and 2:00 PM. Figure 2b depicts the drone in the field during data collection. Finally, based on the preset ground calibration points, the images were cropped to ensure that the areas within the images corresponded exactly to the selected plots. This meticulous attention to flight and imaging parameters ensured that the data collected were of the highest quality and accurately represented the residue cover across the study area.



(a)



(b)

Figure 2. Schematic of the UAV image acquisition sample. The sampling area was established as a square, each side measuring 8 m, marked by four corner markers labeled A, B, C, and D. At the center of this square, the UAV conducted its imaging operations using a predefined method, as shown in schematic (a). Field photo (b) shows the drone ready to capture high-resolution ground images.

2.3. Remote Sensing Image Data Collection and Processing

The images selected for this study were Sentinel-2A multispectral remote sensing images. The Sentinel-2A satellite is part of the European Copernicus environmental monitoring program and is designed to provide a wealth of data and images. The multispectral camera covers 13 spectral bands, from visible to short-wave infrared, which is crucial for agricultural monitoring, inland and coastal waters, land cover classification, and other terrestrial monitoring applications. The satellite imager includes three spectral bands in the “red-edge” region for the first time, providing detailed and sensitive information about vegetation status, thus offering multidimensional, high-resolution, and high-precision data for land monitoring. Considering the synchronicity between the generation of remote sensing images and ground experiments, this study selected Sentinel-2A satellite images from 7 November 2022, as the data source, with a cloud cover of <5% in the selected image area. To eliminate geometric distortions and radiometric errors in the images, they were preprocessed using the ENVI software for geometric correction, radiometric calibration, and atmospheric correction, as shown in Figure 3. For an accurate estimation of farmland residue cover, the study used the maximum likelihood estimation algorithm to analyze and extract cultivated pixels from Sentinel-2A images. Figure 3 shows the extraction results for the cultivated pixels, with approximately 72% of the study area being cultivated land. The boundary between the cultivated and non-cultivated pixels (residential areas) was clear, and the plots were relatively intact and suitable for subsequent studies on residue cover assessment in farmlands.

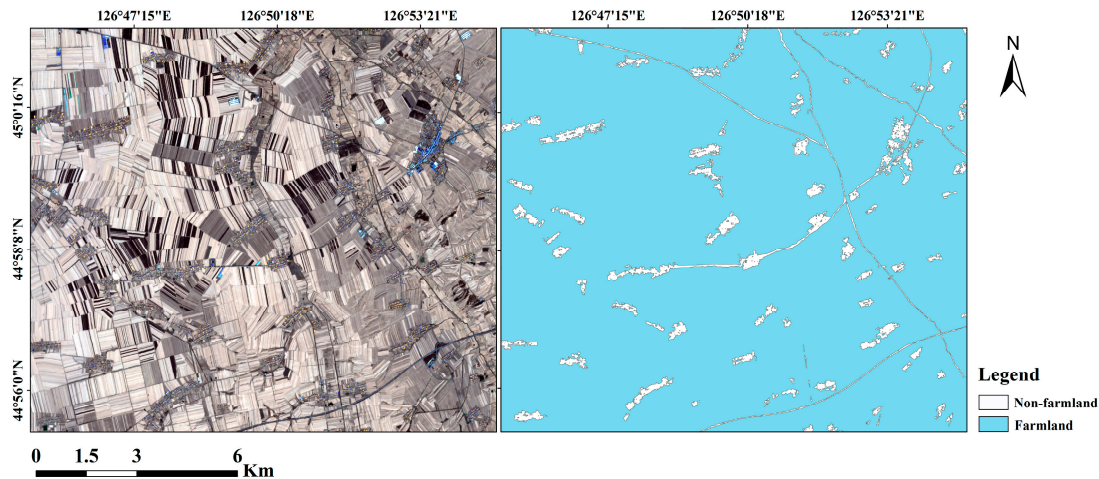


Figure 3. Results of cultivated pixel extraction in the research zone.

2.4. Methods

2.4.1. Yen Image Segmentation Algorithm

The Yen image segmentation algorithm is an adaptive threshold determination technique based on a grayscale histogram and the principle of entropy, which is widely used in the field of image processing. The core of the algorithm uses the pixel distribution information of the histogram to enhance entropy, reflecting the complexity or richness of information in the image [34]. Its primary principle considers the foreground and background of an image as independent sources of signals, each with an entropy value representing the amount of information in that region. The algorithm calculates the entropy for each potential threshold and selects the ideal threshold that best represents the overall image information based on the entropy. In addition, the Yen algorithm can suppress the interference of noise by maximizing entropy to ensure accurate segmentation when dealing with noisy and varying light images. The specific formula for implementing the Yen algorithm is as follows:

$$\begin{cases} P(i) = \frac{h(i)}{N} \\ H_{\text{fore}}(t) = - \sum_{i=t}^{L-1} P(i | T=t) \log P(i | T=t) \\ H_{\text{back}}(t) = - \sum_{i=0}^{t-1} P(i | T=t) \log P(i | T=t) \end{cases} \quad (1)$$

$$t^* = \underset{t}{\operatorname{argmax}} (H_{\text{fore}}(t) + H_{\text{back}}(t)) \quad (2)$$

Suppose that the image has a gray-level range of $[0, L - 1]$, where L is the total number of gray levels. For each possible threshold t (where $0 \leq t < L$), the image is classified into the foreground (gray value $\geq t$) and background (gray value $< t$), assuming that the gray level range of the image is $[0, L - 1]$, where L is the total number of gray levels. For each possible threshold t (where $0 \leq t < L$), the image is classified into foreground (gray values $\geq t$) and background (gray values $< t$), where the entropies of the foreground and background are $H_{\text{fore}}(t)$ and $H_{\text{back}}(t)$, respectively. $h(i)$ is the gray-level histogram of the image, i is the gray level, N is the total number of pixels in the image, and t^* is the optimal segmentation threshold.

2.4.2. SPA Feature Selection Method

In data modeling, redundant variables are a key issue that affects the accuracy of model estimation and reduces the efficiency of model computation. To reduce the model construction workload and improve the accuracy of the prediction model, SPA was used to screen the spectral indices in this study. The SPA operates as a progressive selection

mechanism, beginning with a solitary spectral index and progressively integrating novel vectors with the utmost spatial projection among the pre-selected vectors, continuing this process until it attains a predetermined count of N spectral indices, thus effectively reducing the collinearity among the spatial vectors [35,36]. Figure 4 delineates the fundamental concept and detailed methodology of SPA, where N symbolizes the maximum quantity of variables that can be chosen, and the frequency of projection maneuvers during the spectral index selection sequence can be depicted as follows:

$$(N - 1)(J - N/2) \quad (3)$$

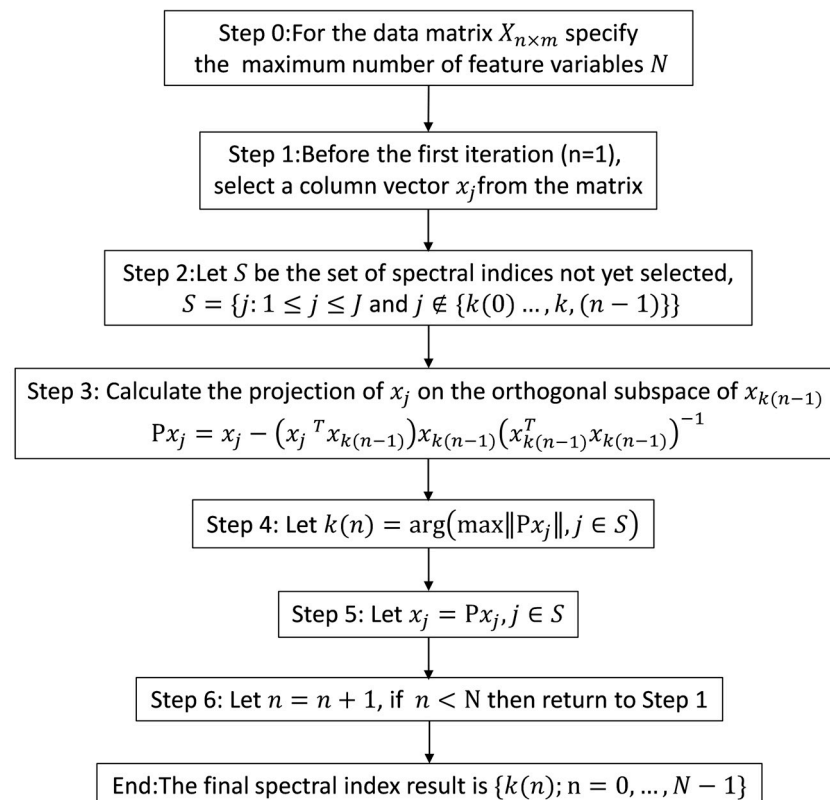


Figure 4. SPA operation flow chart.

2.4.3. CatBoost Estimation Model

The CatBoost algorithm is a machine learning approach based on gradient-boosted decision trees, central to which is the efficient and innovative handling of categorical features, along with a reduction in gradient bias and prediction deviation, thereby enhancing the algorithm's accuracy and generalization capability [30,37]. Divergent from the conventional GBDT algorithm, CatBoost employs methods of random permutation and mean label value calculation for sample processing, in addition to incorporating a prior distribution term, effectively mitigating the noise impact from low-frequency categorical data. Furthermore, CatBoost utilizes a completely symmetric tree as the base model, optimizing the processing capacity for high-dimensional sparse data. In the realm of decision tree algorithms, the mean value of labels is used as the criterion for node splitting, a method referred to as greedy target-based statistics, which is expressed formulaically as follows:

$$x_{i,k} = \frac{\sum_{j=1}^n [x_{j,k} = x_{i,k}] \cdot Y_j}{\sum_{j=1}^n [x_{j,k} = x_{i,k}]} \quad (4)$$

Improving upon the aforementioned formula by adding a prior distribution term can reduce the impact of low-frequency data and noise on the data distribution:

$$x_{i,k} = \frac{\sum_{j=1}^{p-1} [x_{\sigma_{j,k}} = x_{\sigma_{p,k}}] \cdot Y_j + a \cdot p}{\sum_{j=1}^{p-1} [x_{\sigma_{j,k}} = x_{\sigma_{p,k}}] + a} \quad (5)$$

where p is the added prior term, and a is typically a weight coefficient greater than 0.

Compared to other boosting algorithms, CatBoost's main features include: the use of combined categorical features; enriching feature dimensions through inter-feature associations; employing ordered boosting to combat noise in the training set, thereby reducing gradient estimation bias and addressing prediction shift issues; and using completely symmetric trees as the base model to enhance the model's capability in processing high-dimensional sparse data.

2.5. Evaluation Metrics

In this experiment, the Dice coefficient is used to evaluate the segmented maize residue cover images. The Dice coefficient is used to calculate the similarity between two samples [38,39]. The value of the Dice coefficient ranges from 0 to 1, indicating no similarity to complete consistency. The formula for the Dice coefficient is as follows:

$$Dice = \frac{2|A \cap B|}{|A| + |B|} \quad (6)$$

where A denotes the number of pixels of the predicted image and B denotes the number of pixels of the manually segmented image.

In the evaluation of the model, three indicators are used to assess the accuracy of the model: coefficient of determination (R^2), root mean square error ($RMSE$), and relative prediction deviation (RPD). R^2 reflects the fitting ability of the model, and $RMSE$ and RPD measure the dispersion and deviation between the predicted values and the actual values. The closer R^2 is to 1, the smaller the $RMSE$ is, the better the model prediction is, and the model has excellent prediction ability when the RPD is greater than 2 [40]. When RPD is greater than 2, the model has excellent prediction ability, and the formulas of R^2 and $RMSE$ are as follows:

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

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

$$RPD = \frac{\frac{\sum (y_i - \bar{y})^2}{n-1}}{\sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n}}} \quad (9)$$

where y_i is the predicted MRC (maize residue cover), \hat{y}_i is the measured MRC, \bar{y} is the average MRC, and n is the sample size.

3. Results

3.1. Extraction of Maize Residue Cover

To analyze the accuracy of Yen's algorithm for extracting the coverage of agricultural residues, four different segmentation algorithms and Yen's algorithm were selected for segmentation experiments on the images of the study area. Initially, the original images were converted to RGB grayscale images. Then, through segmentation algorithm analysis, the optimal threshold for each algorithm was determined. Subsequently, the RGB grayscale images were binarized according to the optimal thresholds, and the pixel values of the converted grayscale images were categorized into two groups: One representing the

background (without residue) and the foreground (with residue). The specific process of extracting the residue cover using the Yen segmentation algorithm is illustrated in Figure 5. The asterisk (*) in Figure 5 represents the optimal threshold value that maximizes the combined entropy of the foreground and background.

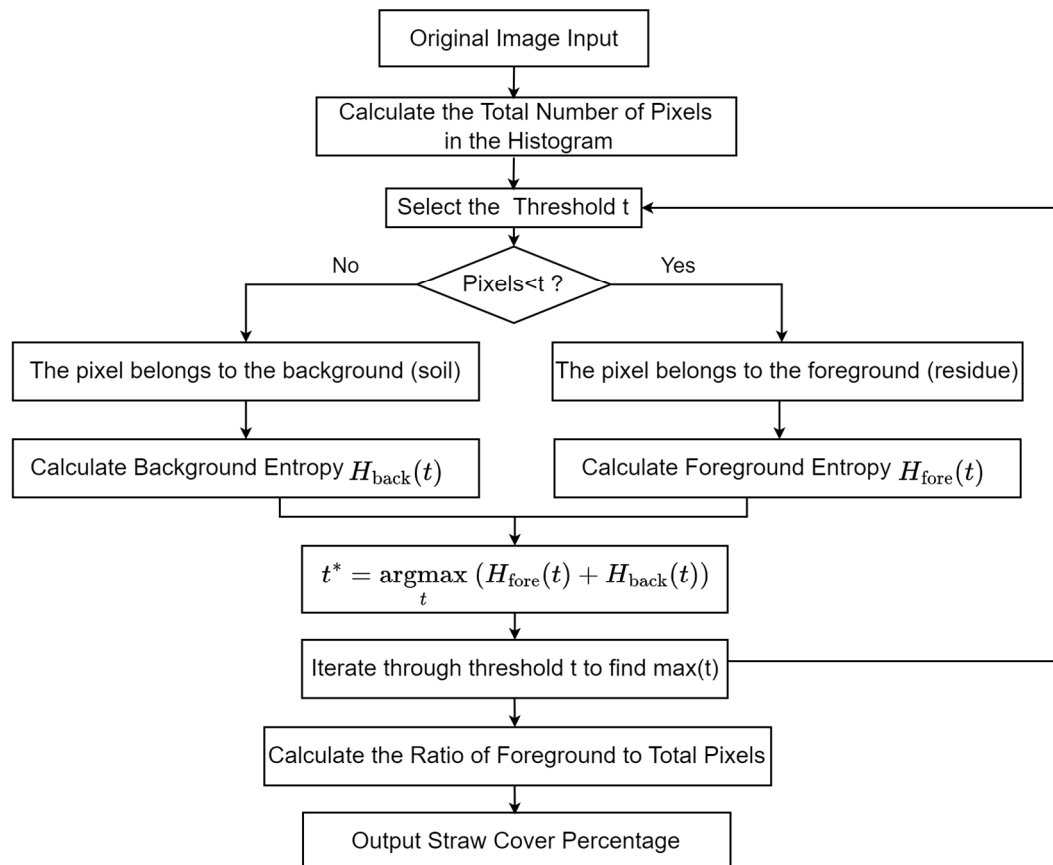


Figure 5. Principle and process of residue cover extraction using the Yen algorithm.

To compare the segmentation effects of different algorithms, this study used the results of human–computer interaction visual recognition of residue cover as the ground truth. Using the first set of field residue cover images as an example, histograms were used to compare and analyze the segmentation results of the Yen, OSTU, K-means, and MaxEntropy algorithms against the visual recognition extraction results. The results are shown in Figure 6, and the Dice coefficient values (Dice) for the classification results of the different algorithms are shown in Figure 7.

The effects of different image segmentation algorithms on maize residue cover extraction are shown in Figure 6. A comparison of Figure 6 shows that the Yen algorithm has a better segmentation effect, and the Yen segmentation algorithm has the highest clarity when compared with the true value image. The OSTU segmentation algorithm showed partial ambiguity in the separation of the residue and soil, and the boundary of the features was not sufficiently clear. K-means segmentation is the least effective method because it produces more dispersed image features and reduces the quality of the image segmentation. Figure 7 shows the results of the quantitative analysis and calculation of the four segmentation algorithms, manual interactive visual recognition, and corresponding Dice coefficients. It is evident from Figure 7 that the Yen segmentation algorithm achieved the highest Dice coefficient of 81.71%. This indicates that the segmentation effect was closest to the ground truth of human–computer interactive visual recognition. Following closely is the OSTU segmentation algorithm, whose Dice coefficient was slightly lower, still demonstrated a good segmentation performance. In contrast, the K-means segmentation algorithm performed

poorly in distinguishing between residue and soil, as indicated by its lowest Dice coefficient, which was significantly lower than the others at a minimum of 63.44%. Therefore, the Yen segmentation algorithm not only provides the clearest visual image segmentation effect but also exhibits the highest Dice coefficient in quantitative assessment, further illustrating the practicality and effectiveness of the Yen image segmentation algorithm in image processing.

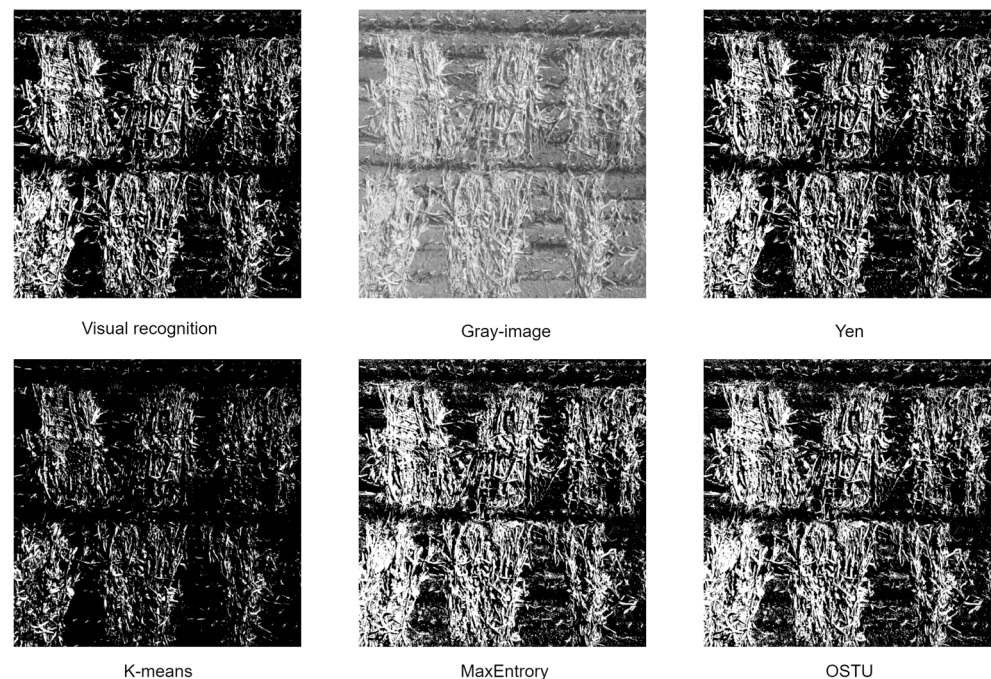


Figure 6. Grayscale image and extraction results based on five different segmentation methods.

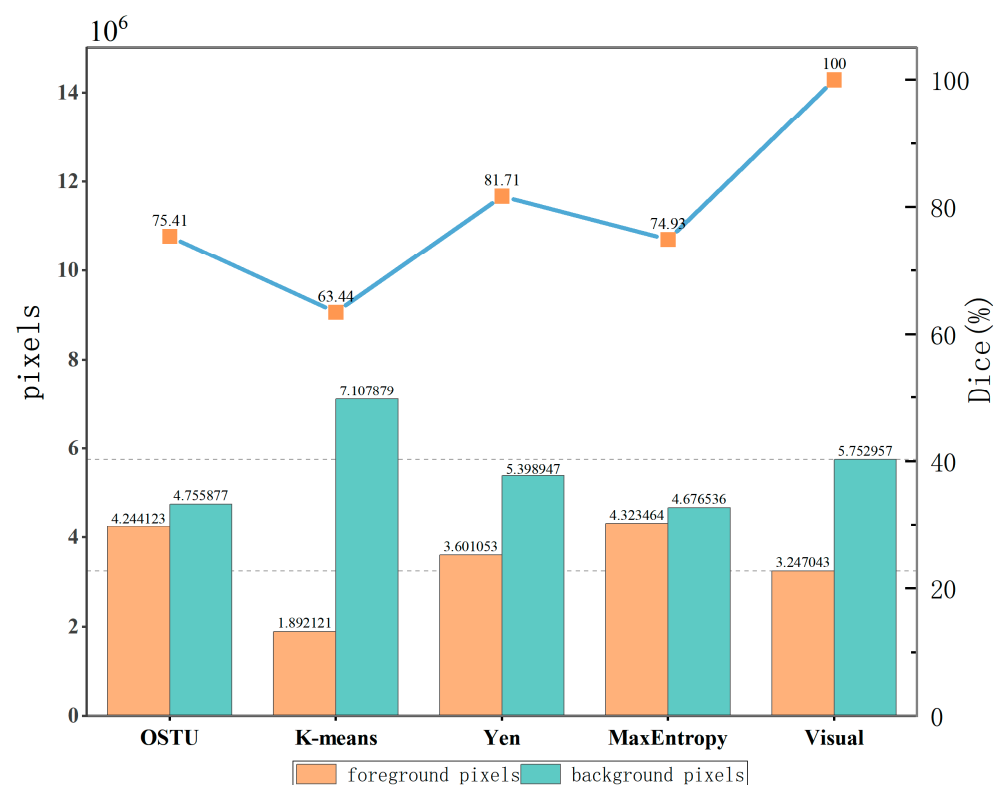


Figure 7. Line graph of pixel distribution based on extraction results from five different segmentation methods and Dice coefficient values.

3.2. Extraction of Spectral Feature Indices

Spectral features exhibit strong reflective or radiative characteristics in specific bands, whereas they may be weaker in others. Band combinations can be used to fuse information from different bands, resulting in a more comprehensive and detailed description of features and surface characteristics. To effectively amalgamate the spectral information of Sentinel-2A and enhance the estimation precision of the model, this study, after analyzing the spectral characteristics of Sentinel-2A imagery, used interpolation, ratio, and normalization techniques to create 45 spectral indices. These indices were used to develop a model for estimating the maize residue cover in the research area. To minimize information redundancy and autocorrelation among these spectral indices, the successive projections algorithm (SPA) was used for feature extraction from the constructed indices, and the extracted spectral feature indices were then used as training sample data for the ensemble learning model. The ideal iteration counts for the SPA were established using RMSECV associated with multivariate regression. As the number of iterations increased, the RMSECV decreased until it reached its minimum. Afterward, new variables were added as the iteration continued, and the multicollinearity between the variables started to increase, resulting in the slow increase and fluctuation of RMSECV. Therefore, the variable when the RMSECV reaches its minimum value is the optimal band combination containing the least amount of redundant information. From Figure 8, it can be observed that as the RMSECV reaches an average trend value of 0.3%, the number of spectral indices is reduced from 45 to 15. This effectively compresses the number of feature spectral indices, while retaining most of the information in the original dataset. The spectral indices computed by SPA feature extraction are listed in Table 1. The green dashed line in Figure 8 indicates the cutoff point where RMSECV stabilizes, signifying the selection of the optimal number of spectral indices for the model.

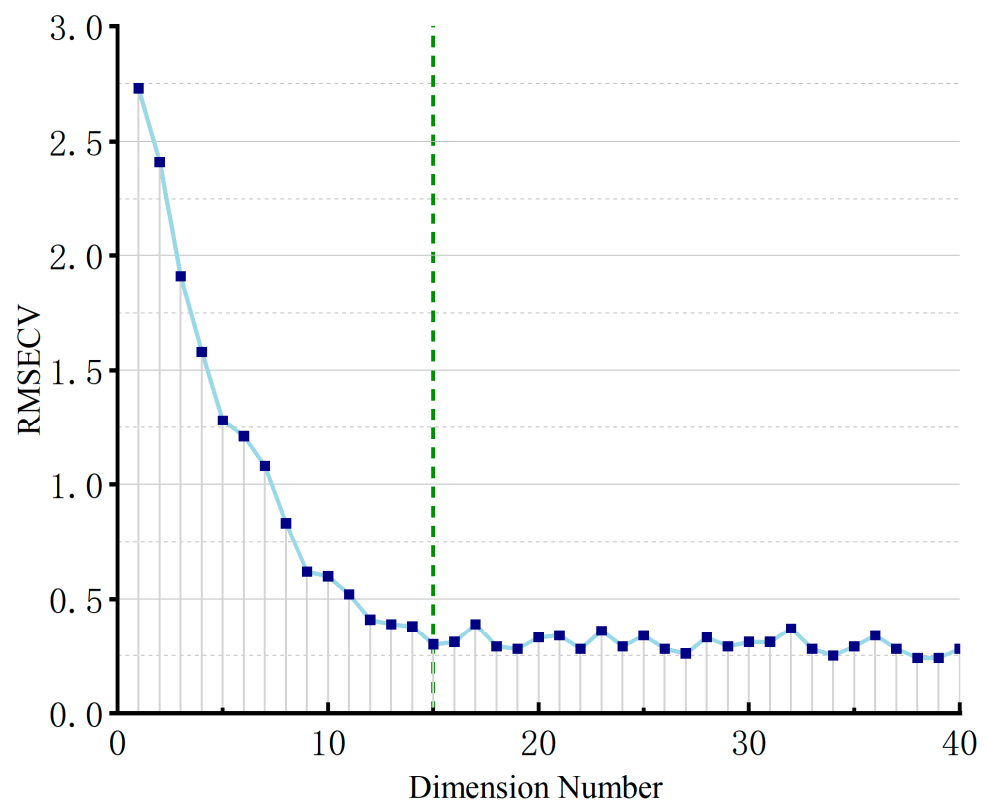


Figure 8. RMSECV (unit:%) of multiple regression with SPA iterations.

Table 1. Combined bands and corresponding tillage indices.

Sentinel-2 MSI Formula	Tillage Index	Abbreviation	Reference
$(B11 - B12)/(B11 + B12)$	Normalized Difference Tillage Index	NDTI	[41]
$(B8A - B12)/(B8A + B12)$	Normalized Difference Index 7	NDI7	[42]
$(B12 - B4)/(B12 + B4)$	Shortwave Red Normalized Difference Index	SRNDI	[7]
$B11/B12$	Simple Tillage Index	STI	[41]
$(B11 - B4)/(B11 + B4)$	Normalized Difference Senescent Vegetation Index	NDSVI	[43]
$(B1 - B2)/(B1 + B2)$	Normalized Difference Chlorophyll Index	NDCI	[44]
$(B6 - B5)/(B6 + B5)$	Normalized Red Edge Drought Index 2	NDRE1	[45]
$(B11 - B3)/(B11 + B3)$	Modified Crop Residue Cover	MCRC	[46]
$(B8A - B11)/(B8A + B11)$	Normalized Difference Index 5	NDI5	[42]
$B8/B4$	Ratio Vegetation Index	RVI	[47]
$(B8 - B7)/(B8 + B7)$	Normalized Difference Vegetation Index Red Edge 3	NDVIRE3	[48]
$(B8 - B4)/(B8 + B4)$	Normalized Difference Vegetation Index	NDVI	[49]
$(B7 - B5)/(B7 + B5)$	Normalized Difference Red Edge 2	NDRE2	[45]
$(B8 - B6)/(B8 + B6)$	Normalized Difference Vegetation Index Red Edge 2	NDVIRE2	[48]
$(B3 - B8)/(B3 + B8)$	Normalized Difference Water Index	NDWI	[50]

3.3. Construction and Evaluation of the MRC Inversion Model

In this study, the spectral indices extracted using SPA were used as independent variables X_i , with MRC as the dependent variable Y_i . Random split sampling was used to divide the sample size randomly into a 7:3 ratio for the training group (to establish the model and optimize parameters) and the validation group (to assess model accuracy and generalizability). This selection method ensured a consistent range and uniform distribution of calibration and validation samples, involving a total of 57 samples, with 40 allocated to training and 17 to validation. To analyze the estimation accuracy of the CatBoost model, the random forest (RF) model, and the multilayer perceptron (MLP) machine learning model were used for comparative analysis.

Hyperparameters play a crucial role in defining the accuracy of a model during its construction. For the CatBoost model, parameters such as the learning rate and quantity of regression trees are pivotal. In this study, R^2 was used as the primary measure to evaluate the effectiveness of the optimization objective function. To refine these models, cross-validation was used for optimization computations across the three distinct models. Table 2 presents the fine-tuned parameters for each post-optimization model. Figure 9 graphically depicts the optimization outcomes for various models.

Table 2. Optimal hyperparameter using grid search with 10-fold CV and cross-validation for different MRC models.

Method	CatBoost			RF		MLPR	
	Max Depth	Estimators	Learning Rate	Max Depth	Estimators	Max Features	Hidden Layer Size
SPA	3	12	0.03	4	150	0.3	2 ³

Combining the hyperparameter calculation results obtained from the parameter optimization calculations, the CatBoost, RF, and MLP models were used to estimate maize residue cover. Additionally, the evaluation metrics for the training and validation of the different models were calculated separately, and the results are presented in Table 3.

As shown in Table 3, considering the accuracy metrics (R^2 , RMSE, and RPD) for both training and validation datasets, as a whole, CatBoost significantly outperformed the other three algorithms, exhibiting the best predictive performance ($R^2 = 0.83$, RMSE = 1.31%, RPD = 2.1). In addition, the prediction accuracy of the validation datasets for all three models was lower than that of the training datasets. Among them, the RMSE values for both the training and validation datasets of the CatBoost model were substantially lower than those of the RF and MLP models.

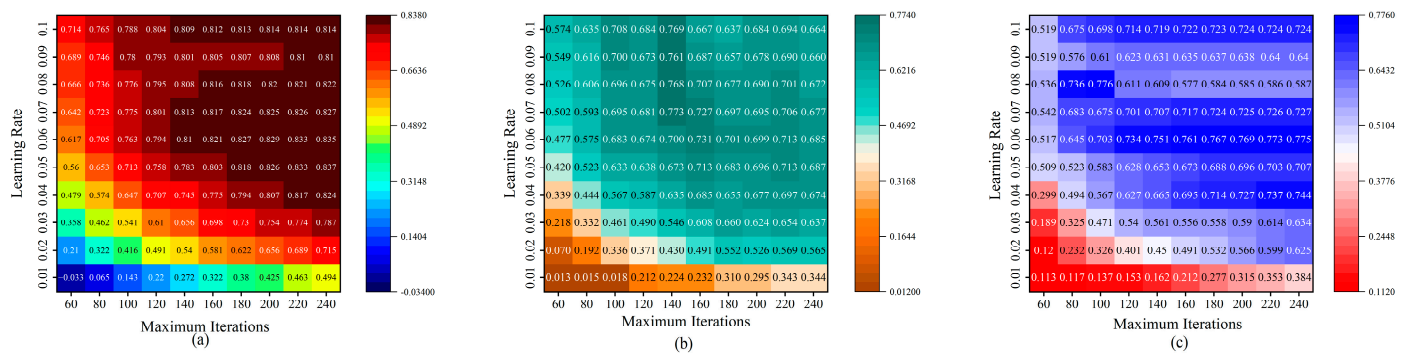


Figure 9. Optimal hyperparameters determined through a grid search combined with 10-fold cross-validation for various MRC models include (a) refined outcomes for the CatBoost parameters, (b) tuned results for the MLP parameters, and (c) adjusted findings for the RF parameters.

Table 3. Evaluation of estimation outcomes of diverse models.

Ensemble Learning Model	Training Data Set			Test Data Set		
	R^2	RMSE (%)	RPD	R^2	RMSE (%)	RPD
CBR	0.83	1.31	2.17	0.81	1.42	1.95
RF	0.62	2.54	1.64	0.58	2.43	1.52
MLP	0.55	2.27	1.32	0.51	2.98	1.11

To further examine and compare the fitting and predictive capabilities of the different models, scatter diagrams depicting the relationship between the predicted and actual values were generated (Figure 10). In these plots, the actual values of the soil samples are plotted along the horizontal axis, whereas the values predicted by the models are plotted on the vertical axis. Training samples are represented by pink dots, and validation samples are represented by blue dots. The proximity of these data points to the diagonal 1:1 line indicated the closeness of the measured values to the predicted values. As depicted in Figure 10, the CatBoost model exhibits superior performance metrics compared with the RF and MLP models, with data points from both the training and validation sets aligned closer to the 1:1 line. Among the tested algorithms, CatBoost demonstrated the highest level of accuracy, followed by the RF model; the MLP model ranked the lowest in accuracy. This suggests that the ensemble learning approach of the CatBoost model surpasses the other models in terms of fitting accuracy and stability, thereby revealing its effectiveness in estimating maize residue cover based on ensemble learning methodologies.

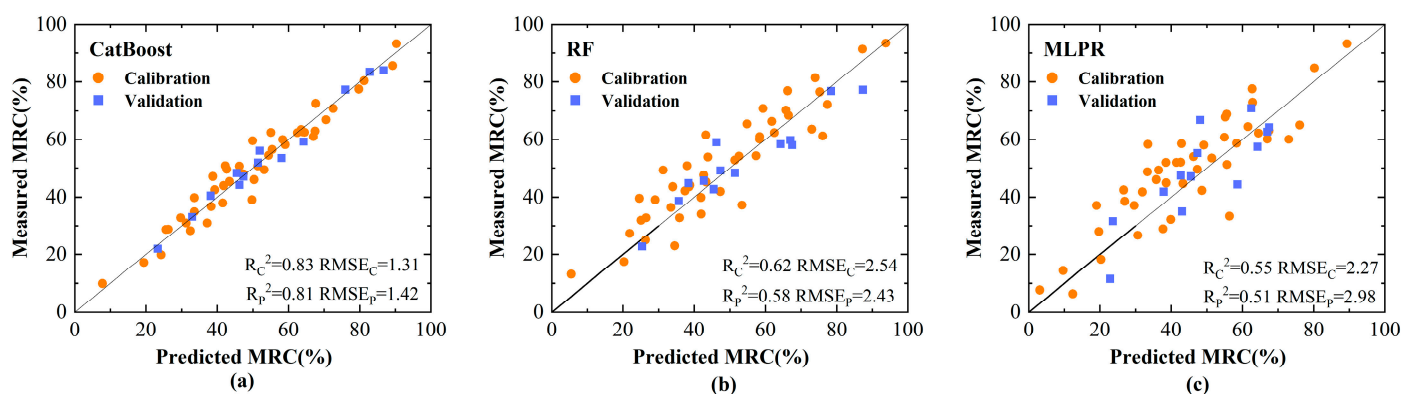


Figure 10. A comparative analysis of the actual versus forecasted values across various regression models includes: (a) the CatBoost model; (b) the RF (random forest) model; and (c) the MLP (multi-layer perceptron) model.

3.4. Spatial Distribution of Maize Residue Cover

According to the selection of the characteristic spectral index and the extraction of the cultivated land image elements, the MRC distribution map of the multispectral images in the study area obtained by the SPA_CatBoost model using the Sentinel-2A multispectral image data as input is shown in Figure 11. According to the statistics of the inversion results, the maximum and minimum values of the inversion results were 95.64% and 4.32%, respectively. In the cultivated area, 24.56% of the MRC ranged from 20% to 40%, 40.37% of the MRC ranged from 40% to 60%, and 17.54% of the MRC ranged from 60% to 80%, which is consistent with the field observations. The results showed that the MRC estimates obtained from the SPA_CatBoost model reflect a reliable MRC distribution. Further analyses showed that areas close to populated areas exhibited relatively high MRC values, which may be attributed to the concentration of residents in stacking and using residues, particularly in agricultural production and application in daily life. This tendency may reflect the dependence of residents on agricultural by-products, as well as the management and maintenance strategies for this resource. In addition, the areas close to roads showed different MRC distribution patterns, presenting moderate to low-ranked MRC values. This phenomenon may be related to road maintenance, traffic safety needs, and residue-removal activities along the roads. This variability suggests that land-use practices and human intervention in the vicinity of roads have a significant effect on maize residue cover. The MRC distribution results obtained through the SPA_CatBoost model not only align with field observation data but also reveal the impact of human activities on the spatial distribution of residue cover. This suggests that MRC distribution is influenced by a combination of land-use patterns, geographic locations, and the activities of residents and road maintenance. Therefore, future research should focus on comprehensively considering the interplay between human activities and the natural environment to promote conservation tillage practices and achieve the sustainable management of agricultural ecosystems.

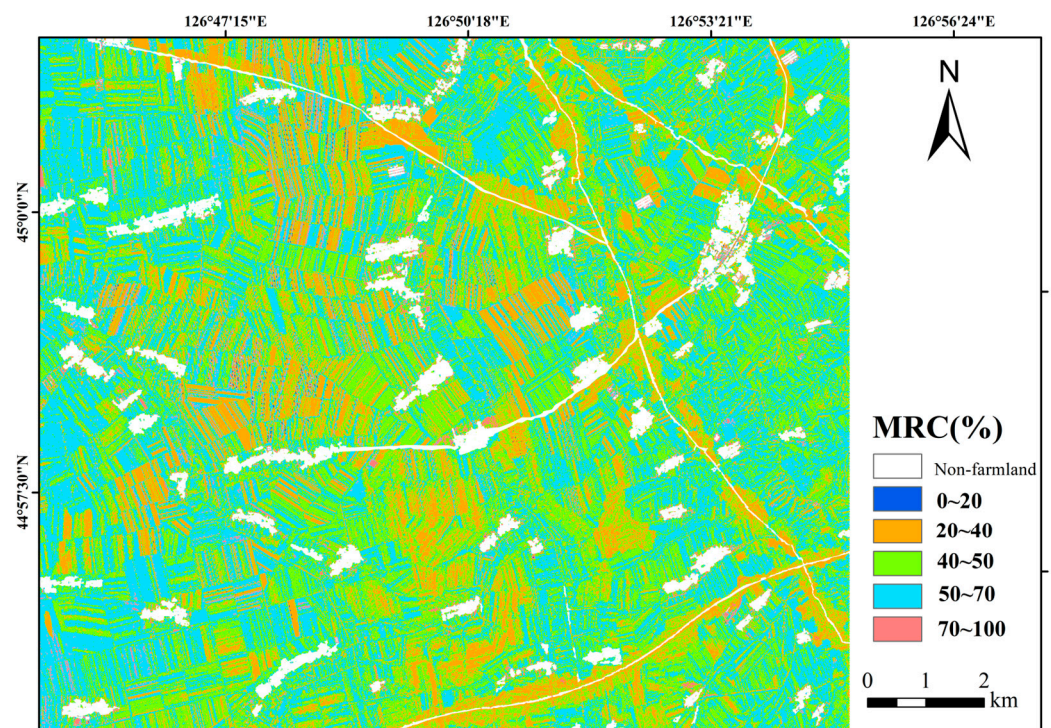


Figure 11. Distribution of MRC in the study area.

4. Discussion

4.1. Analysis of Spectral Index Characteristics for Maize Residue Cover Estimation

In this study, several Sentinel-2 MSI formulas were screened and analyzed for spectral index selection using the SPA feature extraction method and three machine learning models. Several sets of spectral indices were successfully screened, as listed in Table 2. The spectral indices selected in this study were significantly correlated with MRC, especially NDTI, NDI7, STI, and SRNDI (Table 1), which is consistent with previous research findings [1], indicating that these four indices are well-suited for estimating residue cover in the study area. Figure 12 shows the network interaction graph of the spectral indices used in this study with Sentinel-2A bands. In Figure 12, the indices shown in red represent the four significantly correlated spectral indices discussed in this section, and the orange bands indicate the corresponding band combinations for these indices. These relationships reveal that the four highly correlated indices (NDTI, NDI7, STI, and SRNDI) were all related to band B12. NDTI was calculated from bands 11 and 12 of Sentinel-2 imagery, with center wavelengths of B11 and B12 bands near 1610 nm and 2100 nm, respectively. Daughtry et al. [51] noted two moisture absorption features near 1450 nm and 1960 nm, with significant spectral reflectance differences between dry and moisture-saturated residue residues in these bands. Moreover, the SWIR bands, particularly near 2100 nm, are indicative of the presence of lignin and cellulose, essential components of crop residues that contribute to their spectral signature. This spectral response in the SWIR range is highly relevant for our analysis because it directly relates to the structural and moisture characteristics of maize residues, making NDTI and STI particularly effective for their estimation [7,52]. Additionally, Hively et al. [8] indicated that all dry crop residues have a broadband absorption feature near 2100 nm, possibly related to the content of lignin and cellulose in the residues. NDI7 is calculated based on bands 8A and 12. The results of the previous analysis exhibit a significant correlation between 8A and MRC, which can be attributed to the high sensitivity of the near-infrared band 8A to plant structure. It indicates that the reflectance of crop residue is higher than that of the bare soil. On this basis, the significant correlation between NDI7 and the MRC is validated. Ding et al. [52] developed the Soil Residue Normalized Difference Index (SRNDI) utilizing bands 4 and 12, demonstrating its significant correlation with crop residue cover (CRC) as part of comprehensive research on estimating CRC using remote sensing technologies for enhanced agricultural management and tillage intensity evaluation. Daughtry et al. [53,54] highlighted the effectiveness of the Cellulose Absorption Index (CAI) over other spectral indices (VI) for estimating crop residues. The efficacy of CAI is based on the absorption characteristics of cellulose and lignin, particularly near 2100 nm. This finding explains why the SRNDI outperformed the NDSVI in our study. Additionally, other indices, such as NDI5, SRNDR, and NDSVI, which showed slightly weaker correlations with residue cover, still indicated a degree of relevance, suggesting their potential application value in MRC estimation.

4.2. Error Analysis of Image Segmentation Algorithm Results

In this study, different segmentation algorithms were used to extract maize residue cover from the images, and Yen's algorithm was determined to yield the best results in calculating the MRC using Dice coefficients. To further understand the performance of the Yen algorithm and its behavior in specific situations, the segmentation results obtained using the Yen algorithm were compared with those obtained from manual interactive visual recognition, followed by detailed error analysis. Figure 13 shows the residue cover in three specific situations (a–c) and plots each of the 57 images (along the x-axis) along with the corresponding MRC obtained using manual interactive visual recognition and the Yen algorithm (along the y-axis). It can be observed from Figure 13 that the Yen algorithm has a small relative error, with most values within a 5% error range and only a few exceeding 10%.

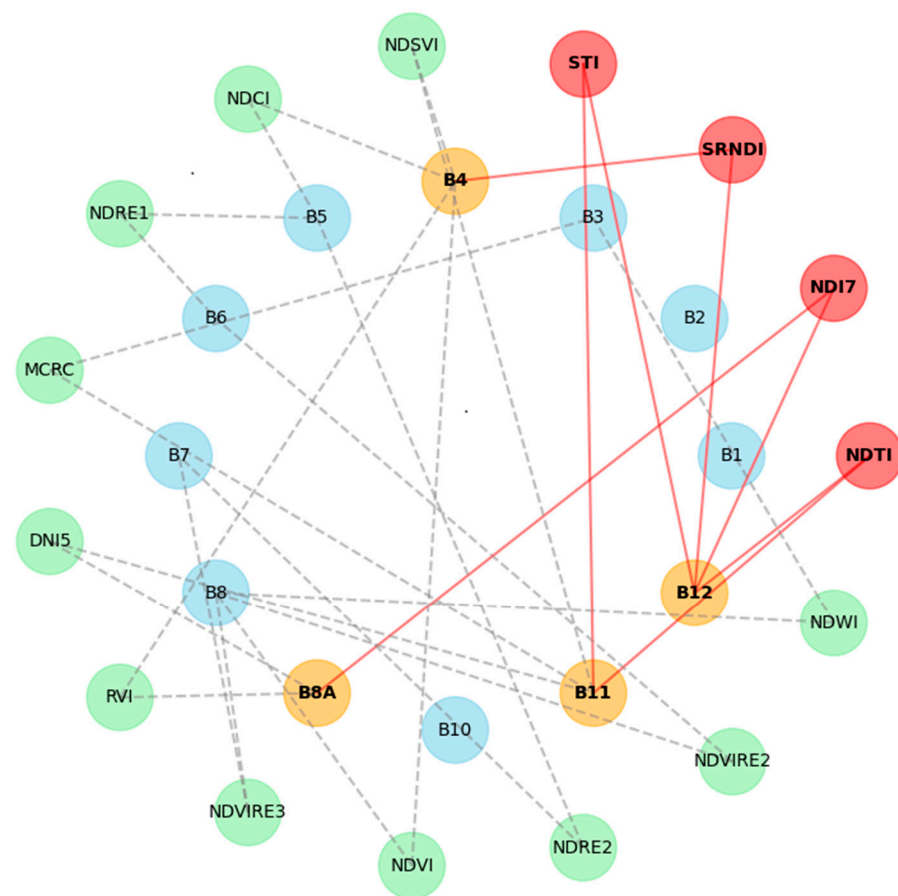


Figure 12. Multispectral index and band interaction network diagram.

In this study, a systematic analysis and evaluation of the performance of the Yen image segmentation algorithm under different residue cover conditions was conducted by comparing experimental data with field survey results. For areas with the highest residue cover (a), the algorithm showed a significant increase in the error rate when segmenting dense and similarly textured image elements, primarily because of high-density textures, making it difficult to accurately distinguish edges and revealing limitations in processing highly overlapping textures. In medium-coverage scenarios (b), the Yen algorithm demonstrated moderate accuracy, effectively handling textures with little variation, although the segmentation precision still requires improvement in situations where the contrast between the background and texture is not prominent. This suggests that, in these moderately complex contexts, the Yen algorithm is slightly inadequate for finer texture discrimination, although it maintains some stability. Furthermore, when analyzing the region with low maize residue coverage (c), the Yen algorithm performed relatively more accurately, with the lowest error rate. This may be attributed to the high contrast between the residue and the background in the low-coverage region, which provides the algorithm with visual features that are easier to distinguish, thus reducing cases of mis-segmentation. This finding suggests that the Yen algorithm is more robust when dealing with low-complexity scenes and is better at extracting textures from simple backgrounds. In these cases, image features such as texture consistency, background simplicity, and uniform illumination work together in the algorithm to achieve a high accuracy in the segmentation task. These findings not only emphasize the applicability of Yen's algorithm in dealing with different texture complexities and illumination conditions but also reveal the limitations of the algorithm in dealing with high-density textures and low-contrast scenes, providing directions for future research to optimize the algorithm for more diverse application scenarios.

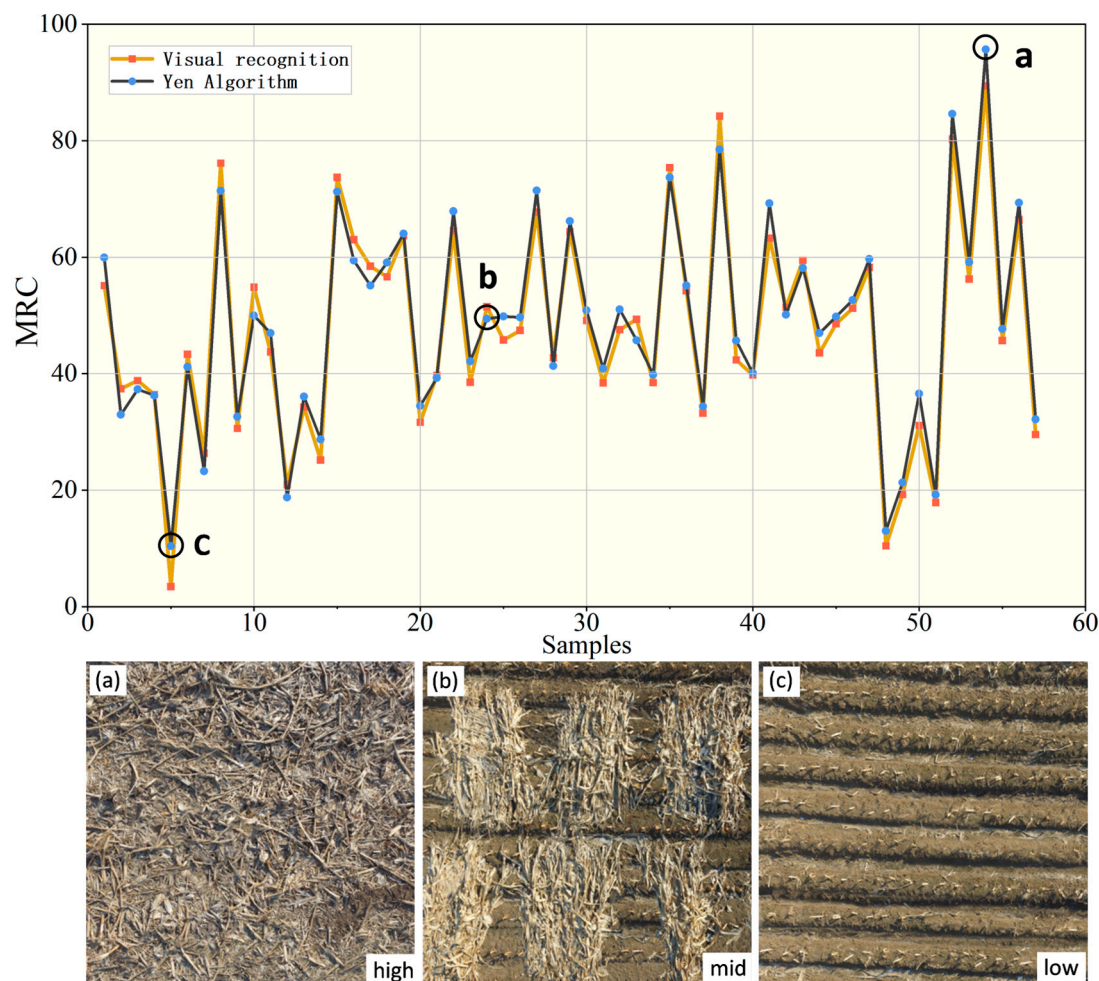


Figure 13. Graph of coverage percentages obtained from visually identified images (orange) and algorithm-segmented images (dark gray). (a) Areas with the highest maize residue cover, (b) areas with medium maize residue cover, and (c) areas with the lowest maize residue cover.

4.3. Uncertainty Analysis

In this study, selecting appropriate training samples and ensuring their high quality were key factors determining the accuracy of model training. However, some uncertainties and potential influencing factors still exist. Through scatterplot analysis of the test samples (Figure 10), maize residue cover tended to be overestimated at low values and underestimated at high values, much like other studies that have used regression models for cover mapping [55]. The overestimation of low-value areas is mainly a problem of sample data collection methods. Especially after the corn harvest, the residual stover in the farmland is usually unevenly distributed and mixed with a small amount of other vegetation on the ground, which increases the complexity of cover estimation. The underestimation of high-value areas was mainly due to the random sampling strategy. The proportion of high cover in the training samples was small, and the cover was unevenly distributed. One possible solution to this problem is to balance the distribution of sample coverage. This can be achieved by adopting a stratified sampling strategy, wherein the number of samples is evenly distributed among different coverage levels, ensuring that the model has sufficient training data across all coverage ranges. Secondly, extracting maize residue cover from sample plots using image segmentation technology is significantly affected by various environmental factors, such as residue moisture content, type, soil moisture, and soil roughness. The moisture content and type of residue directly affect the visual characteristics of images, whereas changes in soil moisture and roughness can interfere with the ability of the image segmentation algorithm to distinguish between the residue and soil boundaries. In this

study, some MRC samples exhibited overestimation at low values, reflecting the limitations of the segmentation algorithm in handling subtle differences between residues and soil. Particularly in different regions, the shadow effect of maize residue and variations in soil moisture further complicate the image segmentation process. Figure 14a shows an example of a mixture of sunlit maize residue, shadowed residue, sunlit soil, and shadowed soil in a field-captured image. In addition, Figure 14b shows an example of a maize residue mixed with high-moisture background soil. To address these issues, future research could consider increasing the monitoring and calibration mechanisms for environmental factors to reduce their impact on image analysis, thereby enhancing the accuracy and reliability of residue coverage estimation. Additionally, collecting more representative samples and performing meticulous adjustments and optimizations to the model can further improve its performance.



Figure 14. Instances of corn residue featuring shadow effects (a) and diverse moisture contents (b).

5. Conclusions

This study explores the methodology and feasibility of large-scale corn stover coverage estimation using multispectral images. A new stover coverage estimation model is constructed by integrating adaptive threshold segmentation and the CatBoost algorithm, which effectively improves the accuracy and stability of the model estimation. The experimental results of image segmentation show that the Yen algorithm has significant advantages in maintaining segmentation stability and strengthening the ability to respond to differences in image contrast. It outperforms traditional algorithms when dealing with low-complexity scenes and high-contrast texture segmentation. The screening of spectral feature indices of multispectral images using the continuous projection algorithm effectively decreases the information redundancy and high correlation between some spectral indices, significantly reducing the computational complexity. The four indices screened, namely, NDTI, STI, NDI7, and SRNDI, facilitate the improvement of the overall accuracy and efficiency of the model. Since the CatBoost model has significant advantages in processing category-type features and reducing gradient bias, it can effectively overcome the problem of insufficient model generalization ability with a small number of samples, exhibiting better accuracy and generalization in the estimation of straw coverage. In this study, the practical utility of the integrated approach is clearly demonstrated across various agricultural management practices. Precise estimation of corn stover coverage can significantly aid sustainable agricultural practices, including crop residue management, soil health monitoring, and precision farming techniques. By providing accurate, reliable data, the model supports informed decision-making for fertilizer application, irrigation planning, and crop rotation strategies, thereby enhancing yield and reducing environmental impact. Additionally, this approach greatly enhances smart and digital agriculture, providing a robust foundation for more precise and informed agricultural practices. The adaptability of this model to various

climatic and terrain conditions, although currently a limitation, underscores the potential for future refinement and application in diverse agricultural settings globally. Continued enhancements in the robustness and adaptability of the segmentation algorithm are essential to extend the practical applications of this research. Thus, ongoing investigation into improving the segmentation algorithm will focus on enhancing its utility in real-world agricultural scenarios.

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

Funding: This research was funded by the Science and Technology Development Project of Jilin Province (20210203016SF), the Natural Science Foundation of Jilin Province (20230101373JC), and the National Natural Science Foundation of China (52178042).

Institutional Review Board Statement: Not applicable.

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

Acknowledgments: The authors would like to thank the China Centre for Resources Satellite Data and Application for providing Sentinel-2A data. We are most grateful to the anonymous reviewers and editors for their valuable comments and recommendations.

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

References

1. Zhang, Y.; Li, X.; Gregorich, E.; McLaughlin, N.; Zhang, X.; Guo, Y.; Gao, Y.; Liang, A. Tillage and Cropping Effects on Soil Organic Carbon: Biodegradation and Storage in Density and Size Fractions. *Eur. J. Soil Sci.* **2020**, *71*, 1188–1199. [\[CrossRef\]](#)
2. Hao, X.; Han, X.; Wang, C.; Yan, J.; Lu, X.; Chen, X.; Zou, W. Temporal Dynamics of Density Separated Soil Organic Carbon Pools as Revealed by $\delta^{13}\text{C}$ Changes under 17 Years of Straw Return. *Agric. Ecosyst. Environ.* **2023**, *356*, 108656. [\[CrossRef\]](#)
3. Rusakova, I.V. Change in the Content of Total and Easily Degradable Organic Matter in Soddy–Podzolic Soil Associated with a Long-Term Straw Incorporation. *Mosc. Univ. Soil Sci. Bull.* **2023**, *78*, S37–S45. [\[CrossRef\]](#)
4. Saqee, F.S.; Norman, P.E.; Saffa, M.D.; Kavhiza, N.J.; Pakina, E.; Zargar, M.; Diakite, S.; Stybayev, G.; Baitelenova, A.; Kipshakbayeva, G. Impact of Different Types of Green Manure on Pests and Disease Incidence and Severity as Well as Growth and Yield Parameters of Maize. *Heliyon* **2023**, *9*, e17294. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Li, J.; Yu, W.; Du, J.; Song, K.; Xiang, X.; Liu, H.; Zhang, Y.; Zhang, W.; Zheng, Z.; Wang, Y.; et al. Mapping Maize Tillage Practices over the Songnen Plain in Northeast China Using GEE Cloud Platform. *Remote Sens.* **2023**, *15*, 1461. [\[CrossRef\]](#)
6. Dong, Y.; Xuan, F.; Li, Z.; Su, W.; Guo, H.; Huang, X.; Li, X.; Huang, J. Modeling the Corn Residue Coverage after Harvesting and before Sowing in Northeast China by Random Forest and Soil Texture Zoning. *Remote Sens.* **2023**, *15*, 2179. [\[CrossRef\]](#)
7. Jin, X.; Ma, J.; Wen, Z.; Song, K. Estimation of Maize Residue Cover Using Landsat-8 OLI Image Spectral Information and Textural Features. *Remote Sens.* **2015**, *7*, 14559–14575. [\[CrossRef\]](#)
8. Hively, W.D.; Lamb, B.T.; Daughtry, C.S.T.; Serbin, G.; Dennison, P.; Kokaly, R.F.; Wu, Z.; Masek, J.G. Evaluation of SWIR Crop Residue Bands for the Landsat Next Mission. *Remote Sens.* **2021**, *13*, 3718. [\[CrossRef\]](#)
9. Berger, K.; Hank, T.; Halabuk, A.; Rivera-Caicedo, J.P.; Woher, M.; Moyses, M.; Gerhátová, K.; Tagliabue, G.; Dolz, M.M.; Venteo, A.B.P.; et al. Assessing Non-Photosynthetic Cropland Biomass from Spaceborne Hyperspectral Imagery. *Remote Sens.* **2021**, *13*, 4711. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Guo, Z.; Kurban, A.; Ablekim, A.; Wu, S.; Van De Voorde, T.; Azadi, H.; Maeyer, P.D.; Dufatanye Umwali, E. Estimation of Photosynthetic and Non-Photosynthetic Vegetation Coverage in the Lower Reaches of Tarim River Based on Sentinel-2A Data. *Remote Sens.* **2021**, *13*, 1458. [\[CrossRef\]](#)
11. Hively, W.D.; Lamb, B.T.; Daughtry, C.S.T.; Shermeyer, J.; McCarty, G.W.; Quemada, M. Mapping Crop Residue and Tillage Intensity Using WorldView-3 Satellite Shortwave Infrared Residue Indices. *Remote Sens.* **2018**, *10*, 1657. [\[CrossRef\]](#)
12. Quemada, M.; Daughtry, C.S.T. Spectral Indices to Improve Crop Residue Cover Estimation under Varying Moisture Conditions. *Remote Sens.* **2016**, *8*, 660. [\[CrossRef\]](#)
13. Janga, B.; Asamani, G.P.; Sun, Z.; Cristea, N. A Review of Practical AI for Remote Sensing in Earth Sciences. *Remote Sens.* **2023**, *15*, 4112. [\[CrossRef\]](#)
14. Yu, Y.; Wang, C.; Fu, Q.; Kou, R.; Huang, F.; Yang, B.; Yang, T.; Gao, M. Techniques and Challenges of Image Segmentation: A Review. *Electronics* **2023**, *12*, 1199. [\[CrossRef\]](#)

15. Sishodia, R.P.; Ray, R.L.; Singh, S.K. Applications of Remote Sensing in Precision Agriculture: A Review. *Remote Sens.* **2020**, *12*, 3136. [\[CrossRef\]](#)
16. Qiao, L.; Gao, D.; Zhang, J.; Li, M.; Sun, H.; Ma, J. Dynamic Influence Elimination and Chlorophyll Content Diagnosis of Maize Using UAV Spectral Imagery. *Remote Sens.* **2020**, *12*, 2650. [\[CrossRef\]](#)
17. Hayat, M.A.; Wu, J.; Cao, Y. Unsupervised Bayesian Learning for Rice Panicle Segmentation with UAV Images. *Plant Methods* **2020**, *16*, 18. [\[CrossRef\]](#)
18. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *Automatica* **1975**, *11*, 23–27.
19. Lavania, S.; Matey, P.S. Novel Method for Weed Classification in Maize Field Using Otsu and PCA Implementation. In Proceedings of the 2015 IEEE International Conference on Computational Intelligence & Communication Technology, Ghaziabad, India, 13–14 February 2015; pp. 534–537.
20. Zhou, J.; Wu, Y.; Chen, J.; Cui, M.; Gao, Y.; Meng, K.; Wu, M.; Guo, X.; Wen, W. Maize Stem Contour Extraction and Diameter Measurement Based on Adaptive Threshold Segmentation in Field Conditions. *Agriculture* **2023**, *13*, 678. [\[CrossRef\]](#)
21. Wang, Y.; Zhuo, R.; Xu, L.; Fang, Y. A Spatial–Temporal Bayesian Deep Image Prior Model for Moderate Resolution Imaging Spectroradiometer Temporal Mixture Analysis. *Remote Sens.* **2023**, *15*, 3782. [\[CrossRef\]](#)
22. Khanal, S.; Kc, K.; Fulton, J.P.; Shearer, S.; Ozkan, E. Remote Sensing in Agriculture—Accomplishments, Limitations, and Opportunities. *Remote Sens.* **2020**, *12*, 3783. [\[CrossRef\]](#)
23. Mourtzinis, S.; Esker, P.D.; Specht, J.E.; Conley, S.P. Advancing Agricultural Research Using Machine Learning Algorithms. *Sci. Rep.* **2021**, *11*, 17879. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Sagi, O.; Rokach, L. Ensemble Learning: A Survey. *WIREs Data Min. Knowl.* **2018**, *8*, e1249. [\[CrossRef\]](#)
25. Ang, K.L.-M.; Seng, J.K.P. Big Data and Machine Learning with Hyperspectral Information in Agriculture. *IEEE Access* **2021**, *9*, 36699–36718. [\[CrossRef\]](#)
26. Nguyen, K.A.; Chen, W.; Lin, B.-S.; Seeboonruang, U. Comparison of Ensemble Machine Learning Methods for Soil Erosion Pin Measurements. *IJGI* **2021**, *10*, 42. [\[CrossRef\]](#)
27. Fu, X.; Zhou, W.; Zhou, X.; Li, F.; Hu, Y. Classifying Mountain Vegetation Types Using Object-Oriented Machine Learning Methods Based on Different Feature Combinations. *Forests* **2023**, *14*, 1624. [\[CrossRef\]](#)
28. Chi, M.; Kun, Q.; Benediktsson, J.A.; Feng, R. Ensemble Classification Algorithm for Hyperspectral Remote Sensing Data. *IEEE Geosci. Remote Sens. Lett.* **2009**, *6*, 762–766. [\[CrossRef\]](#)
29. Ahn, J.M.; Kim, J.; Kim, K. Ensemble Machine Learning of Gradient Boosting (XGBoost, LightGBM, CatBoost) and Attention-Based CNN-LSTM for Harmful Algal Blooms Forecasting. *Toxins* **2023**, *15*, 608. [\[CrossRef\]](#)
30. Hancock, J.T.; Khoshgoftaar, T.M. 4-CatBoost for Big Data: An Interdisciplinary Review. *J. Big Data* **2020**, *7*, 94. [\[CrossRef\]](#)
31. Zhang, Y.; Chang, Q.; Chen, Y.; Liu, Y.; Jiang, D.; Zhang, Z. Hyperspectral Estimation of Chlorophyll Content in Apple Tree Leaf Based on Feature Band Selection and the CatBoost Model. *Agronomy* **2023**, *13*, 2075. [\[CrossRef\]](#)
32. Lu, Q.; Si, W.; Wei, L.; Li, Z.; Xia, Z.; Ye, S.; Xia, Y. Retrieval of Water Quality from UAV-Borne Hyperspectral Imagery: A Comparative Study of Machine Learning Algorithms. *Remote Sens.* **2021**, *13*, 3928. [\[CrossRef\]](#)
33. Luo, M.; Wang, Y.; Xie, Y.; Zhou, L.; Qiao, J.; Qiu, S.; Sun, Y. Combination of Feature Selection and CatBoost for Prediction: The First Application to the Estimation of Aboveground Biomass. *Forests* **2021**, *12*, 216. [\[CrossRef\]](#)
34. Sankur, B. Survey over Image Thresholding Techniques and Quantitative Performance Evaluation. *J. Electron. Imaging* **2004**, *13*, 146. [\[CrossRef\]](#)
35. Zhang, J.; Rivard, B.; Rogge, D.M. The Successive Projection Algorithm (SPA), an Algorithm with a Spatial Constraint for the Automatic Search of Endmembers in Hyperspectral Data. *Sensors* **2008**, *8*, 1321–1342. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Chen, X.; Li, F.; Chang, Q. Combination of Continuous Wavelet Transform and Successive Projection Algorithm for the Estimation of Winter Wheat Plant Nitrogen Concentration. *Remote Sens.* **2023**, *15*, 997. [\[CrossRef\]](#)
37. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased Boosting with Categorical Features. *Adv. Neural Inf. Process. Syst.* **2018**, *31*. [\[CrossRef\]](#)
38. Jha, S.; Son, L.H.; Kumar, R.; Priyadarshini, I.; Smarandache, F.; Long, H.V. Neutrosophic Image Segmentation with Dice Coefficients. *Measurement* **2019**, *134*, 762–772. [\[CrossRef\]](#)
39. Saha, A.; Grimm, L.J.; Harowicz, M.; Ghate, S.V.; Kim, C.; Walsh, R.; Mazurowski, M.A. Interobserver Variability in Identification of Breast Tumors in MRI and Its Implications for Prognostic Biomarkers and Radiogenomics. *Med. Phys.* **2016**, *43*, 4558. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Yang, J.; Li, X.; Ma, X. Improving the Accuracy of Soil Organic Carbon Estimation: CWT-Random Frog-XGBoost as a Prerequisite Technique for In Situ Hyperspectral Analysis. *Remote Sens.* **2023**, *15*, 5294. [\[CrossRef\]](#)
41. Deventer, A.V.; Ward, A.; Gowda, P.; Lyon, J.G. Using Thematic Mapper Data to Identify Contrasting Soil Plains and Tillage Practices. *Photogramm. Eng. Remote Sens.* **1997**, *63*, 87–93.
42. McNairn, H.; Protz, R. Mapping Corn Residue Cover on Agricultural Fields in Oxford County, Ontario, Using Thematic Mapper. *Can. J. Remote Sens.* **1993**, *19*, 152–159. [\[CrossRef\]](#)
43. Qi, J.; Marsett, R.; Heilman, P.; Bieden-bender, S.; Moran, S.; Goodrich, D.; Weltz, M. RANGES Improves Satellite-Based Information and Land Cover Assessments in Southwest United States. *Eos Trans. Am. Geophys. Union* **2002**, *83*, 601–606. [\[CrossRef\]](#)

44. Li, W.; Zhou, Y.; Yang, F.; Liu, H.; Yang, X.; Fu, C.; He, B. Using C2X to Explore the Uncertainty of In Situ Chlorophyll-a and Improve the Accuracy of Inversion Models. *Sustainability* **2023**, *15*, 9516. [\[CrossRef\]](#)
45. 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. [\[CrossRef\]](#)
46. Liu, H.; Li, J.; Du, J.; Zhao, B.; Hu, Y.; Li, D.; Yu, W. Identification of Smoke from Straw Burning in Remote Sensing Images with the Improved YOLOv5s Algorithm. *Atmosphere* **2022**, *13*, 925. [\[CrossRef\]](#)
47. Jordan, C.F. Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology* **1969**, *50*, 663–666. [\[CrossRef\]](#)
48. Gitelson, A.A.; Merzlyak, M.N. Remote Estimation of Chlorophyll Content in Higher Plant Leaves. *Int. J. Remote Sens.* **1997**, *18*, 2691–2697. [\[CrossRef\]](#)
49. Li, F.; Miao, Y.; Feng, G.; Yuan, F.; Yue, S.; Gao, X.; Liu, Y.; Liu, B.; Ustin, S.L.; Chen, X. Improving Estimation of Summer Maize Nitrogen Status with Red Edge-Based Spectral Vegetation Indices. *Field Crops Res.* **2014**, *157*, 111–123. [\[CrossRef\]](#)
50. Gao, B. NDWI—A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space. *Remote Sens. Environ.* **1996**, *58*, 257–266. [\[CrossRef\]](#)
51. Daughtry, C.S.T.; Hunt, E.R.; McMurtrey, J.E. Assessing Crop Residue Cover Using Shortwave Infrared Reflectance. *Remote Sens. Environ.* **2004**, *90*, 126–134. [\[CrossRef\]](#)
52. Ding, Y.; Zhang, H.; Wang, Z.; Xie, Q.; Wang, Y.; Liu, L.; Hall, C.C. A Comparison of Estimating Crop Residue Cover from Sentinel-2 Data Using Empirical Regressions and Machine Learning Methods. *Remote Sens.* **2020**, *12*, 1470. [\[CrossRef\]](#)
53. Daughtry, C.S.T. Discriminating Crop Residues from Soil by Shortwave Infrared Reflectance. *Agron. J.* **2001**, *93*, 125–131. [\[CrossRef\]](#)
54. Daughtry, C.S.T.; Doraiswamy, P.C.; Hunt, E.R.; Stern, A.J.; McMurtrey, J.E.; Prueger, J.H. Remote Sensing of Crop Residue Cover and Soil Tillage Intensity. *Soil Tillage Res.* **2006**, *91*, 101–108. [\[CrossRef\]](#)
55. Zhang, Y.; Ma, J.; Liang, S.; Li, X.; Li, M. An Evaluation of Eight Machine Learning Regression Algorithms for Forest Aboveground Biomass Estimation from Multiple Satellite Data Products. *Remote Sens.* **2020**, *12*, 4015. [\[CrossRef\]](#)

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.